Plastid Sequences Contribute to Some Plant Mitochondrial Genes

Dong Wang,†,1 Mathieu Rousseau-Gueutin,†,1 and Jeremy N. Timmis*,1

†Discipline of Genetics, School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, Australia

‡These authors contributed equally to this work.

*Corresponding author: E-mail: jeremy.timmis@adelaide.edu.au.

Associate editor: Richard Thomas

Abstract

DNA of plastid (chloroplast) origin comprises between 1% and 10% of the mitochondrial genomes of higher plants, but functions are currently considered to be limited to rare instances where plastid tRNA genes have replaced their mitochondrial counterparts, where short patches of mitochondrial genes evolved using their homologous plastidic copies by gene conversion or where a new promoter region is created. Here, we show that, in some angiosperms, plastid-derived DNA in mitochondrial genomes (also called mtpt for mitochondrial plastid DNA) contributes codons to unrelated mitochondrial protein-coding sequences and may also have a role in posttranscriptional RNA processing. We determined that these transfers of plastid DNA occurred a few to 150 Ma and that mtpts can sometimes remain dormant many millions of years before contributing to the mitochondrial proteome.

Key words: chloroplast, DNA transfer, mitochondrion, mtpts.

It has been known for several decades that integrants of plastid (pt) DNA in the mitochondrial (mt) genome are abundant and widespread in seed plants (Stern and Lonsdale 1982; Stern and Palmer 1984). Recent sequencing and comparisons of plant organellar genomes have shown that mitochondrial genomes of different seed plant species contain between 1% and 10% of plastidic sequences (mitochondrial plastid DNAs: mtpts) emanating from widely various regions of the plastome (Wang et al. 2007; Lloyd et al. 2012), whereas plastid genomes contain no mtDNA (Smith 2011). Mtpts have been assumed to be “virtually entirely” nonfunctional (Hao and Palmer 2009) with only rare instances of replacement of mitochondrial tRNA genes by their plastid equivalents (Kanno et al. 1997; Miyata et al. 1998), the partial replacement by homologous recombination of two mitochondrial genes (atp1 and 18S ribosomal RNA) with their homologous plastomic copies (Hao and Palmer 2009; Sloan et al. 2010) or the creation of new promoter regions (Nakazono et al. 1996). However, we show here that mtpt sequences encode amino acid tracts in several different mitochondrial proteins in angiosperm species and are also involved in maturation of mitochondrial mRNAs.

Mtpt sequences present in the mitochondrial genome of ten species (Arabidopsis thaliana, Carica papaya, Cucumis sativus, Nicotiana tabacum, Oryza sativa subs. japonica, Sorghum bicolor, Triticum aestivum, Vigna radiata, Vitis Vinifera, and Zea mays subsp. mays), with currently available annotated plastid and mitochondrial genomes, were analyzed using BlastN (version 2.2.23) (Altschul et al. 1990). Local BlastN was carried out using the parameters previously described (Smith 2011): an expectation value of 10^-5; a word size of 11; match and mismatch scores of 2 and -3, respectively; and gap-cost values of 5 (existence) and 2 (extension).

A Perl script was written in order to determine if any mtpt insertions colocalized with a mitochondrial protein-coding gene. Each mtpt sequence that contributed to a protein-coding gene was then Blasted against the mitochondrial genomes of the species investigated in order to determine if it was present only once in the genome.

The first involvement of a mtpt sequence in a mitochondrial protein-coding sequence was found at the 3’-end of mitochondrial ccmC (cytochrome-c maturation protein) in V. vinifera (grapevine), Silene latifolia (campion), and N. tabacum (tobacco), where it encodes four additional amino acids at the carboxyl terminus of the polypeptide (fig. 1A). Transcription of this mtpt sequence was verified by two V. vinifera expressed sequence tags (ESTs) (DT037770 and DT038666) in which plastid-derived nucleotides form part of the ccmC coding sequence. Further examination of the ccmC showed that the mtpt also donated the highly conserved t-element motif (Forner et al. 2007) of the pseudouridine arm of tRNAs (5’-GGTTCRANYCC-3’) to the 3’ untranslated region, which suggests involvement in mitochondrial mRNA maturation (fig. 1B). This single copy mtpt sequence is present in almost all the sequenced mitochondrial genomes of angiosperms indicating that it was transferred about 150 Ma (Hedges et al. 2006). It contains only 74–86 bp of ptDNA in all species except N. tabacum (tobacco) where there are 1,567 bp. By comparing the short mtpt sequences and their corresponding plastidic sequences, four to ten substitutions were seen to have occurred in the mitochondrial genome of Z. mays, S. latifolia, C. lanatus, and V. vinifera whereas none occurred in N. tabacum. This result, together with phylogenetical analysis, shows that the long tobacco mtpt results from a much more recent secondary ptDNA
transfer with the same 5′ terminus. Due to the unavailability of mitochondrial genome sequences from other Asterids, we were unable to determine if this second mtpt insertion involved homology between the plastidic sequence and the preexisting mtpt or any other adjacent mitochondrial sequences. We designed primers to amplify most of this long mtpt, and we found that it was also present in other Nicotiana species and Solanum esculentum (tomato) (fig. 1C), indicating that this second transfer most probably occurred before the divergence of the Nicotiana and Solanum genera, about 24 Ma (Wu and Tanksley 2010). By comparing the long mtpt present in N. tabacum with the corresponding plastidic sequences from various species belonging to different Asterid families (data not shown), we observed that it had accumulated only three mutations—since its insertion, confirming a very recent transposition.

We found the involvement of a second mtpt sequence, again composed of most of a plastid tRNA gene (trnR), at the 3′-end of ORF25, now known to be atp4 (Heazlewood et al. 1986).
et al. 2003), but its implication has not been recognized. Using the recent complete organelle genome sequences of Zea species, we could clarify the situation which showed clearly that ptDNA encodes between 13 and 15 amino acids at the carboxy terminus of mitochondrial ATP4 of Zea species and provides the stop codon (fig. 2A). We identified a maize EST (BM332325), containing both mtpt and atp4 coding sequence, confirming contiguous transcription of the plastid integrant and its contribution to the mitochondrial protein. This 89–105 bp mtpt that is mainly comprised of the plastidic tRNA-Arg (fig. 2) contains a t-element downstream of the atp4 stop codon, again suggesting a role in 3'-mRNA maturation. Thus two different tRNA-like genes (trnR and trnI) fortuitously encode amino acids at the carboxyl termini of two different mitochondrially encoded proteins (ATP4 and ccmC) in different angiosperm species. The mtpt contributing to atp4 in Zea was present in all Zea mitochondrial genomes but absent from those of other panicoid species, including S. bicolor (sorghum) and Tripsacum dactyloides (gama grass) (fig. 2B). A refined search for this mtpt in Zea mitochondrial genomes allowed the identification of a larger mtpt that was also present only in Zea, indicating that the transfer(s) of chloroplastic sequences to the mitochondrion occurred after the divergence of Zea and Sorghum, less than 11.8 Ma (Swigonova et al. 2004). In the common ancestor of the four Zea species, this larger mtpt was at least 4.1 kb. It is reduced in size to 1.4 kb in Z. mays subsp. parviglumis and Z. mays subsp. mays, most probably due to mtDNA rearrangements. The presence of a common insertion event in F.2 . Plastid DNA inserted in the mitochondrial genome of Zea species contributes to atp4 protein-coding sequences. (A) Contribution of a mtpt sequence to the atp4 amino acid sequence in maize is demonstrated by comparisons of the amino acid sequences present at the carboxy terminus of mitochondrial atp4 in six Panicoid species. The sequences used in this alignment are: In red: Sorghum bicolor (NC_008360: 334754-334809), Tripsacum dactyloides (NC_008362: 139557-139508), Zea perennis (NC_008331: 565475-565957), Zea luxurians (NC_008333: 543464-534640), Zea mays subs. parviglumis (NC_008332: 279012-279187). Zea mays subs. mays (NC_007982: 138625-138462). In blue: The maize EST BM332325 containing both mtpt and atp4 coding sequence is also shown on this alignment. The green box indicates the amino acids contributed by the mtpt in Zea species. Identical amino acids are represented by dots (Z. perennis mtDNA used as reference). Stars correspond to stop codons. (B) Comparison of the mtpt nucleotide sequences present at the 3'-end of mitochondrial atp4 and the surrounding region in some Panicoid mitochondrial genomes. The sequences used in this alignment correspond to the mitochondrial sequences (in red) responsible for the protein product shown in (A) and to the corresponding plastome sequences: In green: Z. mays (NC_001666: 102669-102839 and 120068-120068), Saccharum hybrid cultivar NCO 310 (NC_000368: 103393-103563 and 120838-120688), Saccharum hybrid cultivar SP-80-3280 (NC_005878: 23836-24006 and 41281-41111), S. bicolor (NC_008602: 103560-103730 and 120928-120758), Coxyc lacerma-jobi (NC_013273: 103096-103266 and 120442-120272) and Panicum virgatum (NC_0015990: 101953-102123 and 119326-119156). The stop codons of the mitochondrial atp4 genes are underlined. The green box shows the plastidic fragment that increased the length of the mitochondrial atp4 coding sequence in Zea spp. Identical nucleotides are represented by dots (Z. perennis mtDNA used as reference). The black bar shows the sequence corresponding to the plastidic trnR sequence (trnR pt). Asterisks indicate the region of microhomology (8 bp) probably involved in the mtpt insertion event. The highly conserved 5'-GGTTCRANYCC-3' t-element acquired from the mtpt in Z. mays is indicated by a black rectangle.
the long and short mtpts present in all Zea mitochondrial genomes, as well as the absence of other common shared features in comparison with their corresponding plastidic sequences suggest that only one transfer occurred and that the small mtpt sequences that contribute to atp4 probably derives from a partial duplication of the large mtpt soon after its insertion in the mitochondrial genome. Comparisons of the mtpt involved in atp4 with other available panid mitochondrial genomes suggest that integration utilized micro-homology of 8 bp (TGGGGGGT) present in the preexisting mtDNA and at the 5′-end of the mtpt sequence (fig. 2B).

A third example is a mtpt at the 5′-end of nad9 (subunit 9 of NADH dehydrogenase) in T. aestivum (wheat) (fig. 3A). This 58 bp sequence, corresponding to part of plastidic ndhC, is present in the mitochondrial genome of all species belonging to the two major clades of the Poaceae with sequenced mitochondrial genomes (fig. 3B) and thus was most probably inserted before their divergence, 45–60 Ma (Massa et al. 2011). However, this mtpt sequence is capable of encoding amino acids only in T. aestivum (EST: CD924893), an acquired function facilitated by the presence of indels that remove preexisting mitochondrial stop codons when in frame in nad9 mRNA. Comparison with corresponding plastidic sequences from various species belonging to different families of Liliopsida (data not shown) revealed that these mtpts had accumulated three mutations in all species considered. Another larger mtpt encompassing the mtpt that contributes to nad9 was observed in Oryza japonica, S. bicolor, and T. dactyloides. These short and long mtpts share a deletion and similar substitutions compared with the corresponding plastidic sequence, implying that a single transfer occurred with the short mtpt originating from a partial duplication of the long version.

Thus, in contrast to previous conclusions, mtpts create novel mitochondrial genes, unrelated to their plastidic functions, in ways that are similar to their nuclear equivalent "promiscuous" DNAs (Ellis 1982)—nupts and numts (nuclear integrants of plastid and mitochondrial DNAs, respectively) (Noutsos et al. 2007). These results emphasize that mtpts are evolutionarily potent sequences. Although they may remain inactive for millions of years, they may become functional and spread through the mitochondrial gene pool. Here, we provide examples of mtpt sequences that contribute to mitochondrial protein-coding sequences and that have a likely role in mitochondrial posttranscriptional mRNA processing. With increasing mitochondrial and plastidic genome sequences becoming available, we predict the discovery of more examples that demonstrate the significance of mtpt sequences in the evolution of the cohabiting genetic compartments of eukaryotic cells.

Acknowledgments
We thank the reviewers for their helpful comments. This work was supported by the Australian Research Council’s Discovery Projects funding scheme (DP0986973). D.W. was supported by the China Scholarship Council.

References

![Fig. 3. Plastid DNA inserted in the mitochondrial genome of Poaceae species contributes to nad9 protein-coding sequences in Triticum aestivum. (A) Contribution of a mtpt sequence to the nad9 amino acid sequence in T. aestivum is demonstrated by comparisons of the N-terminal amino acid sequence of mitochondrial nad9 in five species of Poaceae: In red: Sorghum bicolor (NC_008360: 49929-49615), Tripsacum dactyloides (NC_008362: 164899-164870), Zea mays subs. mays (NC_007982: 101544-101573), Oryza sativa japonica (NC_011033: 361512-361628), and T. aestivum (NC_007579: 212573-212253). In blue: The wheat EST CD924893 containing both mtpt and nad9 coding sequence are shown on this alignment. The green box indicates the amino acid contributed by the mtpt in T. aestivum. Identical amino acids are represented with dots (T. aestivum mtDNA used as reference). The amino acid differences observed in the EST CD924893 that result from RNA editing are in bold (Lamattina et al. 1993). (B) Comparison of the mtpt nucleotide sequence present at the 5′-end of mitochondrial nad9 and the surrounding region in some Poaceae mitochondrial genomes. The sequences used in this alignment are: In red: the mitochondrial sequences of S. bicolor (NC_008360: 49929-49813), T. dactyloides (NC_008362: 165180-165064 and 476880-476764), Z. mays subs. mays (NC_007982: 101263-101379), O. sativa japonica (NC_011033: 361512-361628), and T. aestivum (NC_007579: 212568-212253). In blue: The wheat EST CD924893 containing both mtpt and nad9 coding sequence. Identical nucleotides are represented with dots (S. bicolor mtDNA used as reference). The green box indicates the mtpt sequence in the Poaceae.


