Proceedings of ANAS (Biological Sciences), 65(5-6): 184-194 (2010)

Mitochondrial DNA Insertions in *Arabidopsis* Genome: is Organelle-to-Nucleus Gene Transfer Continued?

Amina U. Abdulazimova[1], Yagut Yu. Akbarova[1,2], Nurmemmed Sh. Mustafayev[1], Victor V. Solovyev[3], Ilham A. Shahmuradov[1,4]*

[1]*Bioinformatics Laboratory, Institute of Botany, Azerbaijan National Academy of Sciences, 40 Badamdar Shosse, Baku AZ 1073, Azerbaijan*
[2]*College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman*
[3]*Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20, UK*
[4]*Department of Biosciences, COMSATS Institute of Information Technology, Islamabad 44000, Pakistan*

Our previous studies revealed that nuclear genome of *Arabidopsis* contains multiple mitochondrial DNA splinters with intact copies of many organellar genes. In this paper, using the current annotations of nuclear genomes of these species, we show that nuclear copies of twelve known mitochondrial genes from *Arabidopsis* respectively, have known cDNA support to be transcribed in nucleus. We also found that nuclear copies of eleven mitochondrial ORFs, annotated as hypothetical nuclear genes in *Arabidopsis*, are covered by long transcripts. So, nuclear copies of many mitochondrial genes seem to be expressed in both genomes. Our findings suggest that, at least, some of these mitochondrial genes represent candidates to be evolutionary transferred to nucleus and present an intermediate stage of mitochondria-to-nucleus gene transfer process that continues at present time.

Keywords: *mitochondria, nuclear genome, organellar DNA splinters, gene transfer, plant genome evolution*

INTRODUCTION

Plant genetic system consists of nuclear genome and semiautonomous genomes of mitochondria and plastids. Mitochondrial function is conserved in almost all eukaryotic cells, and it is universally accepted that this DNA-containing organelle had a single, endosymbiotic α–proteobacterial origin (Burger et al., 2003).

Compared to animals, mitochondrial genomes of higher plants are larger in size and contain multiple structural rearrangements due to intra- and/or inter-molecular recombination events. Moreover, even among higher plants, sizes of mitochondrial genomes vary widely (Handa, 2003; see also references therein).

Most mitochondrial proteins are encoded by nuclear genes, synthesized in cytosol and imported into organelles. Thus, in yeast as many as 75% of all nuclear genes could be derived from protomitochondria (Esser et al., 2004). In plants mitochondrial genome encode less than 10% of proteins required for the organelle function (Adams et al., 2001; Adams et al., 2002; Dunkley et al., 2006).

It is suggested that most of genes originally present in ancient endosymbionts, ancestors of the organelles, have gradually been transferred to the nucleus, and their products are transferred to corresponding organelles after translation. Although evolution of plastid and mitochondria involved dramatic reduction in the complexity of their genomes, this process has not gone to completion: a subset of genes remains in all organelles studied to date (Brennicke et al., 1993; Sanchez et al., 1996; Doolittle, 1998; Martin and Herrmann, 1998; Figueroa et al., 1999; Adams et al., 2000; Zerges, 2002; Barbrook et al., 2006). Thus, for example, pea mitochondrial genome contains a truncated *rps7* gene lacking 40 codons at its 5' terminus. The presence of *rps7*-homologous sequences in nuclear genomes of pea and soybean is consistent with recent transfer of functional mitochondrial *rps7* gene to the nucleus in certain plant lineages (Zhuo et al., 1999). It was reported that an insertion of more than 620-kb fragment of mitochondrial DNA (mtDNA) exists in *Arabidopsis* chromosome 2 (Lin et al., 1999; Stupar et al., 2001). Our studies have revealed that nuclear genome of *O. sativa japonica* includes multiple splinters insertions of mtDNA, including large insertions on chromosomes 3 and 12, which cover significant part of mitochondrial genome and include intact copies of many mitochondrial genes (unpublished data).

mtDNA insertions represent a significant source of nuclear chromosomal variation (Lough et al., 2008). Recent studies of Noutsos et al. (2007) suggest that nuclear transfer of organellar DNA

*E-mail: ilham@pgenomics.org

(orgDNA) results mainly in nonfunctional nuclear sequences or in genetic disfunction: most of the protein sequences do not correspond to preexisting organellar ones or they represent markedly reshaped protein domains, due to seemingly adaptation of encoded proteins to new functions.

Transfer of genetic material from mitochondrion to nucleus seems to be an ongoing process. For example, mtDNA insertions represent a significant source of nuclear chromosomal variation (Lough et al., 2008). Stegemann and Bock (2006) reported the successful experimental reconstruction of functional gene transfer between an organelle and the nucleus. However, mechanisms of organelle-to-nucleus gene transfer remains mostly unknown. A particular organellar protein is likely encoded by nuclear or plastid genome, but not by both genomes. At the same time, until the transfer is completed, both organelle and nuclear genomes could contain a functional gene copy. For example, functional mitochondrial and nuclear *rpl5* genes seem to coexist (Sandoval et al., 2004).

In light of the facts described above and using last versions of annotated chromosomes of *Arabidopsis*, we performed a genome-wide search for splinters of mtDNA in nucleus, with a genome-wide analysis of such insertions for their functional and evolutionary consequences. Here we report the results of these studies.

MATERIALS AND METHODS

Following sets of publicly annotated five chromosomes of *Arabidopsis* (NC_003070.6, NC_003071.4, NC_003074.5, NC_003075.4 and NC_003076.5; July 2008) and mitochondria genome of *Arabidopsis* (NC_001284.2) were used for this study.

To search for cDNA/mRNA support for transcription of nuclear insertions of organelle DNA, we used six sets of *Arabidopsis* cDNA sequences, total 138,017 (ATH1.cdna.gz from ftp://ftp.tigr.org/pub/data/a_thaliana/ath1/SEQUENCES; cDNA_full_reading_020509.txt.gz, cDNA_full_reading_030312.txt.gz, cDNA_full_reading_050524.txt.gz and cDNA_full_reading_051102.txt.gz from http://rarge.gsc.riken.jp/archives/rafl/sequence; ATcdna155.bz2 from http://gremlin3dev.gdcb. iastate.edu/AtGDB/download.php).

Analysis of genome annotations of *Arabidopsis* was performed using computer programs specially developed by us for these studies. Pairwise comparison of amino acid sequences has been carried out by BLAST program (Altschul et al., 1997). Search for statistically significant open reading frames (ORFs) and putative target compartments of proteins were done by BESTORF and ProtComp programs, respectively (http://www.softberry.com).

RESULTS

General features of nuclear insertions of mtDNA in Arabidopsis

To identify all significant nuclear "harbours" of mtDNA, we preformed BLAST comparison of complete mitochondrial genome of *Arabidopsis* with the nuclear genome, and selected organelle-nucleus sequence homologues with length ≥300 bp and with ≥80% similarity for further studies. Results of this analysis are summarized in Table 1.

Five *Arabidopsis* chromosomes contain 21 harbours of mtDNA of total length 276,487 bp. Some nuclear chromosomes are richer in sequences from organelle genomes. Thus, ~1.5% (~271 kb) of chromosome 2 have mitochondrial origin and it represents actually a single large insertion of mtDNA of ~270 kb reported previously (Lin et al., 1999; Stupar et al., 2001). It should be also noted that we did not find ~620 kb insertion in chromosome 2, reported by Stupar et al. (2001), and, therefore, we believe that this is in fact a 270-kb insertion.

Our analysis shows existence of one or more full-length copies (without mutations truncating their coding sequences, hereinafter referred to as "intact copies") of 15 known and 34 ORF mitochondrial genes in the nuclear genome.

To understand possible functional and evolutionary consequences of mtDNA insertions in nucleus, we analyzed the distribution of mtDNA insertions in coding and non-coding regions of the nuclear genome (Table 2).

In *Arabidopsis*, mtDNA insertions of 100 bp or longer affected 60 genes. 74 insertions of mtDNA affected one or more coding exons of 59 genes; 25 insertions of mtDA were found in introns of 12 genes. Finally, 80 insertions of mtDNA are located in intergenic spacers. If only larger (≥150 bp) insertions were taken into account, results were similar. So, mtDNA insertions are mostly found in intergenic spacers.

Intact nuclear copies of mitochondrial genes seem to be transcribed in Arabidopsis

Although most of intact nuclear copies of organellar genes seem to reflect their recent migration into nucleus, some of them might represent intermediate variants of organellar genes being at a stage of transfer to nuclear genome. If such transfer takes place, migrating genes can be transcribed if they acquire nuclear promoters. Moreover, insertion of orgDNA into already existing genes might produce novel proteins that include functional domain from organellar genes.

To study these phenomena in *Arabidopsis*, we performed BLAST comparison of (i) mtDNA insertions in nucleus and intact nuclear copies of mitochondrial genes with sequenced cDNA sets for the corresponding species, and (ii) mitochondrion-encoded proteins with nuclear proteins annotated.

Taking into account that cDNA sets should be incomplete, in *Arabidopsis* we found cDNA evidence of transcription for about 26% (94700 out of 366924 bp) of mitochondrial genome presented in nucleus. These putative transcripts include intact and partial copies of 28 and 6 mitochondrial genes, respectively. Moreover, 66 nuclear transcripts of mitochondria-derived sequences do not contain any annotated organellar genes (Table 3).

Intact copies of twelve known *Arabidopsis* mitochondrial genes (*atp1, atp6-1, atp6-2, atp9, ccb256, cob, cox3, orfB/atp8, rpl5, rps4, rps7* and *rps12*) were found to have cDNA support and annotated as nuclear genes (Table 4; see also Table 1). Moreover, 16 mitochondria-encoded ORF genes have cDNA support (Table 1) and 11 of them are annotated as unknown nuclear genes (Table 5). Interestingly, all mitochondria-derived 23 (12 known and 11 ORF) nuclear genes are located in large (~270 kb) insertion of mtDNA in *Arabidopsis* chromosome 2, which is consistent with probable transcription of ~35% of this 270 kb mitochondria-derived sequence, judging by cDNA support. Mutual locations of known genes on mtDNA and their intact copies on chromosome 2 is presented in Figure 1. Nuclear intact copy of only one mitochondrial hypothetical gene (*orf224*) seems to be transcribed.

Results of these studies suggest that, at least, 12 *Arabidopsis* genes, known to be expressed in mitochondria, have nuclear copies that seem to be transcribed, taking into account that transcription

Table 1. Characteristics of nuclear insertions of mtDNA in *Arabidopsis*

Mitochondrial genome size	366,924 bp
Nuclear "harbours" of mtDNA*	21 (276487)
mtDNA portion of chromosomes	Chr2: 271000 bp, ~1.5%
Insertions of mtDNA ≥20,000 bp	Chr2: 3239035-3509762, ~270000 bp
Portion of mtDNA in nucleus	~268000/366924, 73%
Mitochondrial genes with intact nuclear copies	15 known, 34 ORF genes: ***atp1, atp6-1, atp6-2, atp9, ccb256***, *ccb382*, ***cob, cox3,*** *matR, nad1,* ***orfB/atp8, rpl5, rps4, rps7, rps12,*** *orf100a/b, orf101a,* ***orf102b***, *orf105a, orf114,* ***orf106a-e, orf107b-g, orf109b,*** *orf111a/c,* ***orf118***, *orf119,* ***orf122a***, *orf127,* ***orf136a/b, orf139a, orf141***, *orf145a/b,* ***orf152a,*** *orf153b, orf154, orf157, orf158,* ***orf159***, ***orf161***, *orf170, orf184, orf199,* ***orf262***, ***orf215b,*** *orf240b, orf275, orf291,* ***orf292***, ***orf313***

*Nuclear "harbours" of mtDNA are "contigs" of nuclear genomic sequences consisting of 300 bp or longer orgDNA homologues (with ≥80% similarity), where only ≤5 bp "gaps" with nuclear DNA (nuDNA) are allowed. Total length of *Arabidopsis* chromosomes annotated is 119,186,497 bp.
Nuclear copies of mitochondrial genes with cDNA support are given in bold.

Table 2. Distribution of mtDNA insertions in the nuclear coding and non-coding regions of *Arabidopsis*

	L ≥ 150[1]	L ≥ 100[2]
Genes with mtDNA insertion(s) in coding and/or non-coding part	60[4]	60[4]
Genes with mtDNA insertion(s), at least, in 1 coding exon	58 (74)	59 (74)
Genes with mtDNA insertion(s), at least, in 1 intron	10 (23)	12 (25)
Intergenic spacers with mtDNA insertion(s)	70	80

[1]Length of ≥80% similarity region is ≥150 bp or the length of complete exon (intron).
[2] Length of ≥80% similarity region is ≥100 bp or the length of complete exon (intron).
[3]Some nuclear genes contain mtDNA insertions in both exon(s) and introns(s).

Table 3. Nuclear insertions of mtDNA that appear to be transcribed in *Arabidopsis*

Portion of mtDNA included in nuclear transcripts*	~94700/366924 (26%); 131 transcripts
Mitochondrial genes included in nuclear transcripts completely	28 (48), 48 transcripts
Mitochondrial genes included in nuclear transcripts partially	6 (out of 48) 24 transcripts
Nuclear transcripts without the whole or partial copy of mitochondrial gene	66

*Portion of mtDNA inserted in nuclear genome and having cDNA evidence of nuclear transcription.

and following translation of organellar gene and its identical nuclear copy does not result in identical proteins, as nuclear transcript does not undergo RNA editing, corresponding organellar transcript does (Table 4, Figure 2).

In *Arabidopsis*, 7 out of 12 nuclear genes of mitochondrial origin encode proteins with predicted mitochondria sub-cellular localization, with one of them (*atp1*) is the most real candidate for the mitochondria-targeted functional gene (Figure 3). Its nuclear homologue, annotated as AT2G07734 gene, has cDNA support and seems to have acquired one additional exon of 827 bp of unknown origin, encoding 276 aa, where the first 54 aa are predicted to be a strong mitochondrial transit peptid. 5 C to U editing transitions results in four amino acid changes in the mitochondrial protein. As to nuclear splinters of the remaining organellar genes, we did not find unambiguous targets of the corresponding proteins.

Possible functional and evolutionary effects of such insertions are discussed below.

DISCUSSION

Patterns of mtDNA insertions in Arabidopsis nucleus

With the exception of one long (~270 kb) insertion in chromosome 2, *Arabidopsis* genome seems be "not hospitable to shelter" sequences of organellar origin. Moreover, in current genome release we did not find ~620 kb insertion in chromosome 2, reported by Stupar et al. (2001) earlier, finding a 270-kb insertion instead.

Does organelle-to-nucleus gene transfer continue in plants?

Our analysis revealed intact nuclear copies of 72 known and 20 ORF genes annotated in mitochondrial genome (Table 2). This finding can be explained mostly by recent nuclear insertions history. Nevertheless, some of the genes might represent intermediate variants of organellar genes being currently at a stage of transfer to nucleus. This suggestion is supported by our results described above.

Significant portion of mtDNA, harbored by nucleus, seems to be transcribed (Table 3). Although most of these transcripts lack any, intact or partial, copies of organellar genes, we found cDNAs that include complete sequences of 28 mitochondrial genes (Table 1). In particular, intact nuclear copies of 12 *Arabidopsis* known mitochondrial genes (*atp1, atp6-1, atp6-2, atp9, ccb256, ccb382, cob, cox3, orfB/atp8, rpl5, rps4, rps7* and *rps12*) are annotated as nuclear genes and have known cDNAs (Table 4; see also Table 1). 16 mitochondria-encoded ORF genes have cDNA support (Table 1) with annotation of 11 out of 16 genes as unknown nuclear genes (Table 5). These findings suggest that, at least, 12 *Arabidopsis* genes, expressed in mitochondria, have also transcribed nuclear copies. However, as mentioned above, an organellar gene and its identical nuclear copy do not encode identical proteins if transcript of an organellar gene undergoes post-transcriptional RNA editing: edited mRNAs produces different polypeptides than those predicted by translating DNA open reading frames. Whether only edited mRNAs or both edited and "original" mRNAs are translated, is not known. It was proposed that RNA editing adds necessary functionality to proteins. In transgenic tobacco plants, delivering the "unedited" mitochondrial protein (translated from unedited mRNA) to the mitochondria leads to mitochondrial dysfunction generating the male sterile phenotype (Araya et al., 1998). RNA editing was described in both chloroplast and mitochondrial genes of plants; editing of mitochondrial transcripts appears to occur in all higher plants (Schuster et al., 1991; Araya et al., 1998). In particular, in *Arabidopsis* RNA editing effects 441 C to U changes in mitochondrial coding sequences (Giege and Brennicke, 1999). Nine out of 12 known mitochondrial genes with transcribed intact nuclear copies, undergo RNA editing resulting in amino acid changes (Table 4). Therefore, although using universal genetic code in chloroplasts and in mitochondria of green plants (Jukes and Osawa, 1990) "allows" transfer of organelle genes into nucleus, nuclear splinters of organellar genes with post-transcriptional RNA editing in organelle already posses "non-synonymous mutations" in their coding sequences (see Figure 2), except in two cases: (1) transcripts of nuclear copies also undergo RNA editing at the same positions; (2) RNA-mediated gene transfer takes place, as with *cox2* transfer to nucleus (Covello and Gray, 1992).

Although the loss of organellar genes due to transfer to nucleus was reported for only small portion of mitochondrial genes (Table 6), organelle-to-nucleus gene transfer is likely continuing process. How genes are transferred from prokaryotic organellar genomes in eukaryotic nuclear genomes and how the genes become functional in their new genetic environment is largely unknown. Results of Stegemann and Bock (2006) suggest that DNA-mediated gene transfer can give rise to functional nuclear genes if followed by suitable rearrangements in nuclear genome including: (1) transcriptional activation by capturing a promoter from an upstream nuclear gene; (2) utilization of AT-rich noncoding sequences downstream from plastid gene as RNA cleavage and polyadenylation sites; and (3) targeting information to direct protein product back to the same organelle or another compartment.

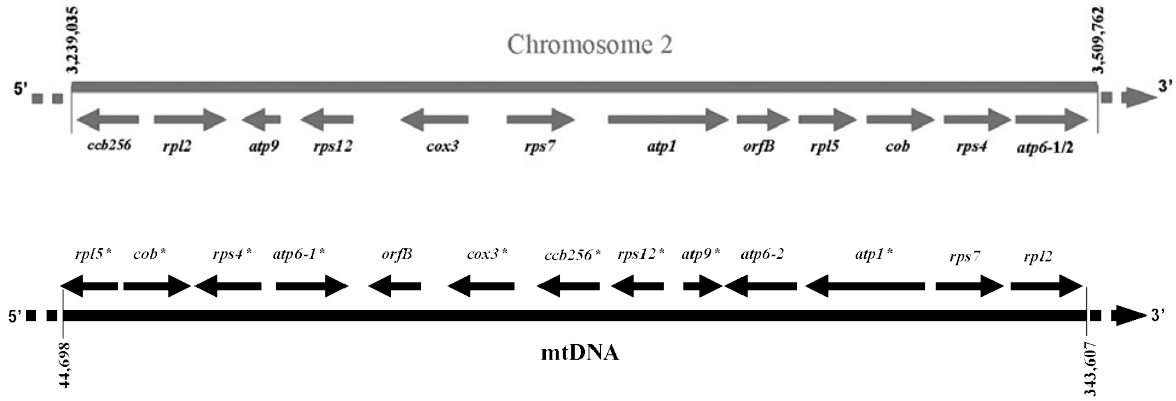

Figure 1. Schematic presentation of annotated full-length nuclear genes (*rpl5*/AT2G07725, *cob*/AT2G07727, *rps4*/AT2G07734, *atp6-1*/AT2G07741, *orfB*/AT2G07707, *cox3*/AT2G07687, *ccb256*/AT2G07771, *rps12*/AT2G07675, *atp9*/AT2G07671, *rps7*/AT2G07696 and *rpl2*/AT2G07715) or partial (*atp1*/AT2G07698 and *atp6-2*/AT2G07741) copies of mitochondria-encoded genes. Asterisks denote genes where mRNA editing results in amino acid changes in the protein sequences. All nuclear genes shown, except *rpl2*/AT2G07715, have cDNA supports.

Table 4. Nuclear copies of *Arabidopsis* mitochondrial genes of known function with cDNA support

Organelle gene	mRNA editing	Protein encoded by organelle gene	Annotated nuclear gene or chromosome location of organelle gene copy	cDNA support for nuclear gene	Similarity of organellar and nuclear proteins	Nuclear protein target[8]
cob	7/7[1]	apocytochrome B, 393 aa	AT2G07727, 393 aa	gb EF488902, gb EF488903	386/393, 98%[5]	M
atp1	4/5[2]	ATPase subunit 1, 507 aa	AT2G07698, 777 aa	gb EF488886, gb EF488887	6:507 – 276:777, 99%[6]	M
atp6-1	1/1[1]	ATPase subunit 6, 385 aa	AT2G07741, 385 aa	gb EF488888, gbEF488889	384/385, 99.8%[7]	EC
atp6-2	?	ATPase subunit 6, 349 aa	AT2G07741, 385 aa	gb EF488888, gb EF488889	252/252, 100%[7] [98:349 – 134:385]	M
atp9	4/4[1]	ATPase subunit 9, 85 aa	AT2G07671, 85 aa	gb EF488892, gb EF488893	81/85, 95%[7]	M
orfB	0	ATPase subunit 8, 158 aa	AT2G07707, 158 aa	gb BT026104, gb EF488891	158/158, 100%[7]	PM
cox3	6/7[2]	cytochrome c oxidase subunit 3, 265 aa	AT2G07687, 265 aa	gb EF488906, gb EF488907	259/265, 98%[7]	EC
ccb256	25/28[2]	cytochrome c iogenesis protein, 256 aa	AT2G07771, 256 aa	gb EF488897, gb BT014995	231/256, 90%[7]	PM
rpl5	9/10[3]	ribosomal protein L5, 185 aa	AT2G07725, 185 aa	gb BT011585, gb EF488938	176/185, 95%[7]	M
rps4	14/15[3]	ribosomal protein S4, 362 aa	AT2G07734, 362 aa	gb EF488944, gb EF488945	348/362, 96%[7]	N
rps7	0	ribosomal protein S7, 148 aa	AT2G07696, 148 aa	gb BT011309, gb EF488946	148/148, 100%[7]	M
rps12	6/8[4]	ribosomal protein S12, 125 aa	AT2G07675, 125 aa	gb EF488948, gb EF488949	119/125, 95%[7]	M

[1]Every edited [C>U] nucleotide affect different codons (i.e. 7, 1 and 4 amino acid changes, respectively). [2]Some edited nucleotides produce synonymous codons (1, 3 and 1 cases, respectively), others produce non-synonymous codons. [3]10 (15) edited nucleotides result in 9 (14) amino acid changes (in 1 case 2 edited nucleotides affect the same codon, "double editing"). [4]8 edited nucleotides produce 6 non-synonymous codons (1 double editing, 1 synonymous). [5]There is only 1 mismatch between CDSs of mitochondrial and nuclear genes. [6]There are 2 mismatches between CDSs of mitochondrial and nuclear genes. [7]CDSs of mitochondrial and nuclear genes are identical. [8]Predicted by ProtComp (http://www.softberry.com). M – mitochondria, N – nucleus, PM – plasma membrane, EC – extracelluar. "?" – unknown.

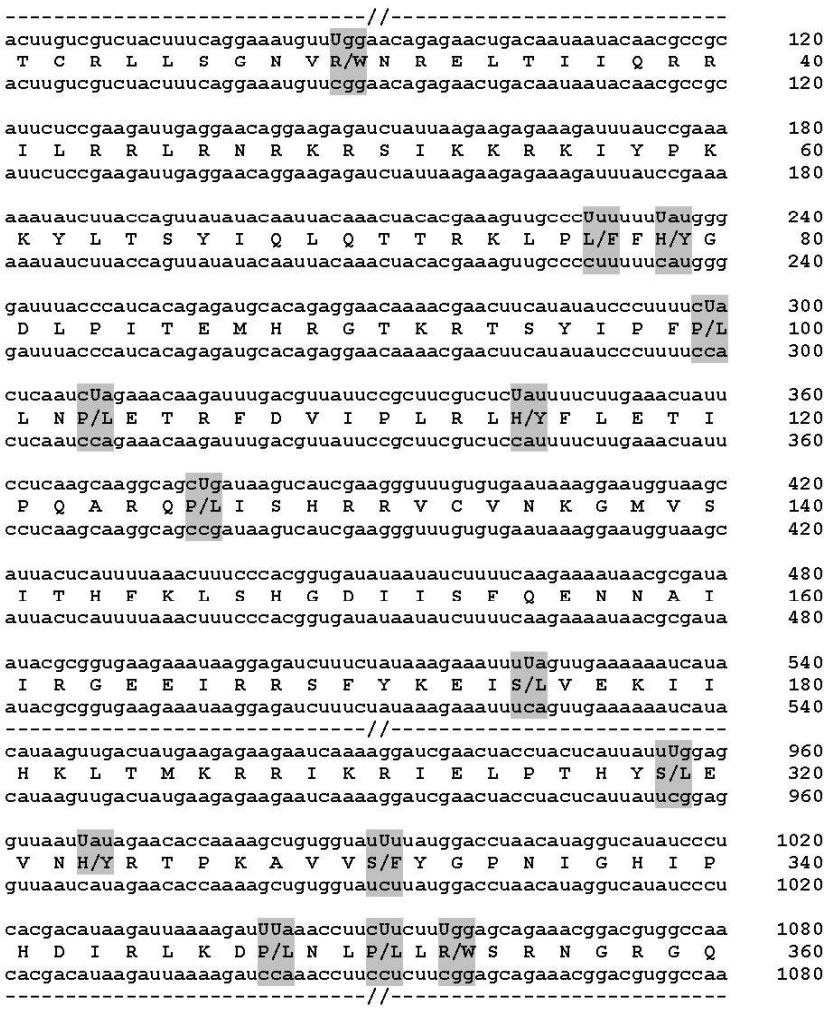

Figure 2. Comparison of edited mRNA for mitochondrial *rps4* gene and predicted mRNA for its nuclear homologue annotated as AT2G07734 gene of *Arabidopsis* (supported by cDNAs, gb EF488944 and gb EF488945). Only regions with edited nucleotides in the organellar mRNA are shown. Editing of 15 nucleotides (C to U) results in 14 amino acid changes in the mitochondrial protein (in one case, two edited nucleotides are located in the same codon).

It was also shown that *rpl5* gene has been independently transferred to the nucleus in the maize lineage and has acquired regulatory elements for its expression and a mitochondrial targeting peptide from an unknown source (Sandoval et al., 2004). Recently, it was found that genes for ribosomal proteins L2 and S4 in *Arabidopsis* mitochondrial genome contain information for protein targeting into the mitochondria. Similarly, genes for ribosomal proteins L2 and S19 in rice mitochondrial genome contain information for protein targeting into mitochondria. These last findings suggest that, before transfer to the nucleus, targeting information already existed in these genes (Ueda et al., 2008).

A particular organellar protein function is likely encoded by nuclear or organelle genome, but not by both genomes at once. However, until the organelle-to-nucleus gene transfer is completed, both organelle and nuclear genomes could contain a functional gene copy. In wheat, mitochondrial and nuclear functional *rpl5* genes appear to be maintained (Sandoval et al., 2004). Mitochondrial gene *cox2* has been transferred to the nucleus during legume plant evolution. This gene represent an intermediate stage of gene transfer process: 9 legumes contain intact copies of both nuclear and mitochondrial *cox2*, where both *cox2* genes are transcribed in seven legumes. Inactivation of *cox2* in each genome has taken place multiple times via a variety of ways, including the loss of detectable transcripts or transcript editing and the partial or complete gene loss (Adams et al., 1999).

Table 5. *Arabidopsis* mitochondrial ORF genes with intact nuclear copies annotated as hypothetical genes and covered by long transcripts

Gene	Support cDNA	Similarity area and identities	Annotated as nuclear gene[1]
orf152a: 459 nt	gb AY131999.1: 2128 nt	1:459 – 840:1298, 457/459	AT2G07702, hypothetical protein
orf122a: 369 nt	gb AY234409.1: 1627 nt	1:369 – 1098:1466, 368/369	AT2g07706, hypothetical protein
orf141: 426 nt	gb AY231406.1: 1212 nt	1:426 – 501: 926, 426/426	AT2G07708, hypothetical protein
orf102b: 369 nt	gb AY132000.1: 1135 nt	1:369 – 200:568, 369/369	AT2G07713, hypothetical protein
orf313: 942 nt	gb AK230169.1: 5275 nt	1:942 – 924:1865, 942/942	AT2G07718, hypothetical protein[2]
orf159: 480 nt	gb DQ069801.1: 862 nt	1: 480 – 36:515, 478/480	AT2g07672, hypothetical protein
orf106e: 321 nt	gb DQ069838.1: 861 nt	1:321 – 655:335, 320/321	AT2G07673, hypothetical protein
orf118: 357 nt	gb BT010728.1: 1162 nt	1:357 – 716:1072, 357/357	AT2G07674, hypothetical protein
orf215b: 648 nt	gb AK230107.1: 3524 nt	1:642 – 1385:2026, 641/642	AT2g07678, hypothetical protein
orf109: 330 nt	gb AK229938.1: 883 nt	1:330 – 536: 865, 329/330	AT2G07776, hypothetical protein
orf262: 789 nt	gb AY648320.1: 1844 nt	1:789 – 756:1544, 789/789	AT2g07777, hypothetical protein

[1] In *Arabidopsis* genome annotation these genes are given without cDNA support.
[2] AT2G07718 is annotated as putative cytochrome *b* gene.

Table 6. The higher plant genes transferred from organelle genomes to nucleus

	Gene	Reference
Mitochondria	*rpl2*	Adams et al., 2001
	rpl5	Sandoval et al., 2004
	rpl16, rps1, rps2, rps3, rps4	Adams et al., 2002
	rps7	Zhuo et al., 1999
	rps10	Adams et al., 2000; Kubo et al., 2000; Wischmann and Schuster, 1995
	rps11	Kadowaki, 1996
	rps12	Grohmann et al., 1992
	rps13	Sanchez et al., 1996
	rps14	Figueroa et al., 1999; Sandova et al., 2004; Kubo et al., 1999; Brochier et al., 2000
	rps19	Subramanian et al., 2001; Watanabe et al., 1994; Kanno et al., 1997
	sdh3, sdh4	Adams et al., 2001
	atp8	Sandoval et al., 2004
	cox2	Pérez-Martínez et al., 2001; Adams et al., 1999; Covello and Gray, 1992
Chloroplast	*infA*	Millen et al., 2001
	ndhF	Wakasugi et al., 1994
	rpl22	Gantt et al., 1991
	rpl23	Bubunenko et al., 1994
	rps16, ycf4	Millen et al., 2001
	accD	Katayama and Ogihara, 1996

In light of these facts, 12 *Arabidopsis* mitochondrial genes (*atp1, atp6-1, atp6-2, atp9, ccb256, cob, cox3, orfB/atp8, rpl5, rps4, rps7* and *rps12*) encode known proteins that are expressed in organelle and, at least, transcribed in nucleus, and may be at the intermediate stages of organelle-to-nucleus transfer. Moreover, 7 out of 12 mitochondrial genes (*atp1, atp6-1, atp6-2, atp9, ccb256, cob* and *cox3*) were not described earlier as genes transferred to nucleus in any species (see Table 6). Nuclear copies of 7 out of 12 mitochondrial genes in *Arabidopsis* might encode proteins targeted to mitochondria. One of them (*atp1*) is the most plausible candidate for the mitochondria-targeted nuclear functional gene (Figure 3). Its nuclear homologue, annotated as AT2G07734 gene with cDNA support, seems to have acquired one additional exon of 827 bp of unknown origin, encoding 276 aa, of which the first 54 aa were predicted as a strong mitochondrial transit peptide. Interestingly, 5 C to U editing transitions results in 4 amino acid changes in the mitochondrial protein. As to nuclear splinters of the remaining organellar genes, we could not find unambiguously a sub-cellular target of the corresponding proteins.

At the same time, we can't exclude that these transcribed nuclear inserts of organellar genes might represent novel genes with modified function and/or destination.

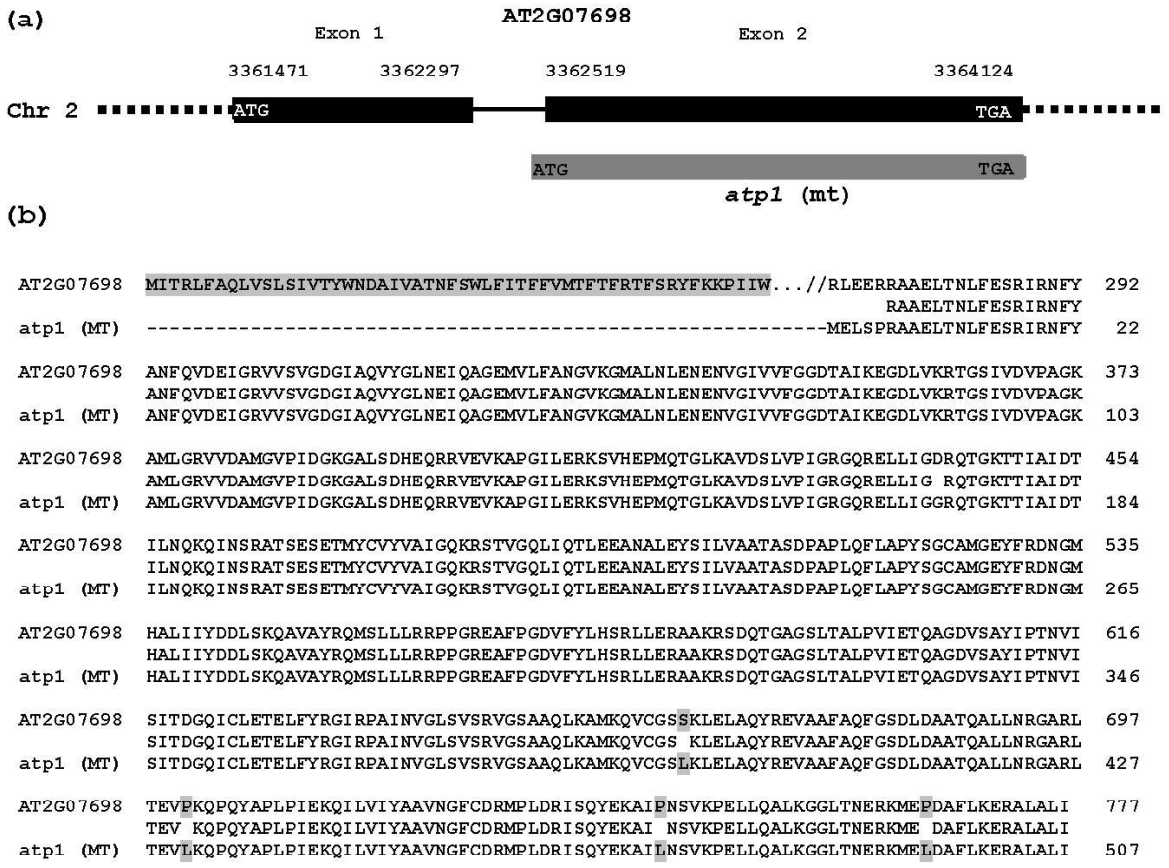

Figure 3. Comparison of mitochondrial *atp1* gene, encoding ATP synthase subunit alpha, and nuclear gene AT2G07698, consisting of two coding exons (Exon 2: insertion of *atp1*, Exon 1: genomic sequence of unknown origin which encodes 276 aa sequence, where the first 54 aa (highlighted) were predicted to be mitochondrial transit peptide (ProtComp program; http://www.softberry.com). Nuclear gene AT2G07698 has cDNA support (gb EF488886, gb EF488887). Amino acid changes caused by C→U RNA editing are highlighted.

Are there restrictions to organelle-to-nucleus transfer?

The evolution of plastid and mitochondria from endosymbionts of cyanobacterial α-proteobacterial origin, seems to have involved a dramatic reduction of complexity of organellar genomes, with many genes either discarded or transferred to the nucleus, and with only a small subset of genes remaining in an organelle. It was found that some organellar genes (chloroplast *infA*, mitochondrial *rpl2, rpl16, rpsl, rps7, rps10, rps13, rps19, sdh3, cox2* and *sdh4*) were more frequently transferred to nucleus (see Table 1 and references therein), with some genes "preferring to stay in organelle". This observation raises questions like these: Is there some selectivity for organellar genes to be transferred to nucleus? If yes, what kinds of restrictions exist? If no, should we expect an existence, in other species yet not studied, of a different spectrum of organellar genes transferred into a nuclear genome and, moreover, the further gradual reduction organellar genomes and, finally, is it possible that after some evolutionary time all organellar functions will be encoded exclusively by nuclear genome?

To answer these questions, several hypotheses have been put forward: (1) it is a challenge to import hydrophobic membrane proteins into organelle (von Heijne, 1987; Popot and de Vitry, 1990); (2) mitochondrial hydrophobic proteins, being synthesized in endoplasmatic reticulum, might be directed to "wrong addresses" (von Heijne, 1987); (3) presence of some organellar proteins in cytoplasm might result in undesirable effects (Herrmann, 1997; Martin and Schnarrenberger, 1997); and (4) expression of some genes is directly and quickly regulated by redox status of organelle and therefore these genes are retained in organelle (Allen, 1993; Pfannschmidt, 2003; Galganska et al., 2008; Liu et al., 2008). To date, there is direct experimental evidence that supports hypothesis 1 in mitochondria (Claros et al., 1995; Daley et al., 2002), and hypo-

thesis 4 in chloroplasts (Allen, 1993; Liu et al., 2008) and mitochondria (Galganska et al., 2008). Moreover, protein length also might play some role, with facilitating or hampering protein transport (Martin and Herrmann, 1998; Lang et al., 1999; Race et al., 1999).

At the same time, it should also be mentioned that the formed gene content of organelle probably might mostly reflect an optimal balance of genetic control of plastid and mitochondrial functions. Based on this point of view, a scenario of extinction of organelle genome(s) looks improbable. In any case, we are still far away from complete understanding the evolutionary past and future, development and coexistence of nuclear and organellar genomes.

ACKNOWLEDGEMENTS

This work, supported by the Higher Education Commission of Pakistan through Foreign-Faculty Start-up grant, is a collaborative research performed by the Department of Biosciences (COMSATS Institute of Information Technology, Islamabad, Pakistan), the Department of Fundamental Problems of Biological Productivity (Institute of Botany, Azerbaijan National Academy of Sciences, Baku, Azerbaijan) and Department of Computer Science (Royal Holloway, University of London, UK).

REFERENCES

Adams K., Ong H., Palmer J. (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 K.L., Daley D.O., Qiu Y.-L., Whelan J., Palmer J.D. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature **408:** 354-357.

Adams K.L., Qiu Y.-L., Stoutemyer M., Palmer J.D. (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 K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L., Doyle J.J., Palmer J.D. (1999) Intracellular gene transfer in action: dual transcription and multiple silencing of nuclear and mitochondrial *cox2* genes in legumes. Proc. Natl. Acad. Sci. USA **96:** 13863-13868.

Allen J. (1993) Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. J. Theor. Biol. **165:** 609-631.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389-3402.

Araya A., Zabaleta E., Blanc V., Begu D., Hernould M., Mouras A., Litvak S. (1998) RNA editing in plant mitochondria, cytoplasmic male sterility and plant breeding. EJB – Electr. J. Biotech. **1:** 1-9.

Barbrook A.C., Howe C.J., Purton S. (2006) Why are plastid genomes retained in non-photosynthetic organisms? Trends Plant Sci. **11:** 101-108.

Brennicke A., Grohmann L., Hiesel R., Knoop V., Schuster W. (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Lett. **325:** 140-145.

Brochier C., Philippe H., Moreira D. (2000) The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. Trends Genet. **12:** 529-533.

Bubunenko M., Schmidt J., Subramanian A. (1994) Protein substitution in chloroplast ribosome evolution: A eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. J. Mol. Biol. **240:** 28-41.

Burger G., Lang B.F. (2003) Parallels in genome evolution in mitochondria and bacterial symbionts. IUBMB Life **55:** 205-212.

Claros M.G., Perea J., Shu Y., Samatey F.A., Popot J.-L., Jacq C. (1995) Limitations to *in vivo* import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome b. Eur. J. Biochem. **228:** 762-771.

Covello P., Gray M. (1992). Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. EMBO J. **11:** 3815-3820.

Daley D.O., Clifton R., Whelan J. (2002) Intracellular gene transfer: Reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome *c* oxidase. Proc. Natl. Acad. Sci. USA **99:** 10510-10515.

Doolittle W.F. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. **14:** 307-310.

Dunkley T.P.J., Hester S., Shadforth I.P., Runions J., Weimar T. et al. (2006) Mapping the *Arabidopsis* organelle proteome. Proc. Natl. Acad. Sci. USA **103:** 6518-6523.

Esser C., Ahmadinejad N., Wiegand C., Rotte

C., Sebastiani F., Gelius-Dietrich G. et al. (2004) A genome phylogeny for mitochondria among a-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol. Biol. Evol. **21:** 1643-1660.

Figueroa P., Gomez 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.

Galganska H., Budzinska M., Wojtkowska M., Kmita H. (2008) Redox regulation of protein expression in *Saccharomyces cerevisiae* mitochondria: possible role of VDAC. Arch. Biochem. Biophys. **479:** 39-45.

Gantt J.S., Baldauf S.L., Calie P.J., Weeden N.F., Palmer J.D. (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved in the gain of an intron. EMBO J. **10:** 3073-3078.

Giege P., Brennicke A. (1999) RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. Proc. Natl. Acad. Sci. USA **21:** 15324-15329.

Grohmann L., Brennicke A., Schuster W. (1992) The mitochondrial gene encoding ribosomal protein S12 has been transferred to the nuclear genome in *Oenothera*. Nucl. Acids Res. **20:** 5641-5646.

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

Herrmann R. (1997) Eukaryotism, towards a new interpretation. In: Eukaryotism and Symbiosis. (Schenk H., Herrmann R., Jeon K., Muller N., Schwemmler W., eds.), Springer, Vienna: 73-118.

Jukes T.H., Osawa S. (1990) The genetic code in mitochondria and chloroplasts. Experientia **46:** 1117-1126.

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.

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

Katayama H., Ogihara Y. (1996) Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. Curr. Genet. **29:** 572-581.

Kubo N., Harada K., Hirai A., Kadowaki K. (1999) A single nuclear transcript encoding mitochondrial RPS 14 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. et al. (2000) Transfer of the mitochondrial *rps10* gene to the nucleus in rice: acquisition of the 59 untranslated regions followed by gene duplication. Mol. Gen. Genet. **263:** 733-739.

Lang B., Gray M., Burger G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. Annual Rev. Genet. **33:** 351-397.

Lin X., Kaul S., Rounsley S., Shea T.P., Benito M.I., Town C.D., Fujii C.Y. et al. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature **402:** 761-768.

Liu K.L., Shen L., Wang J.Q., Sheng J.P. (2008) Rapid inactivation of chloroplastic ascorbate peroxidase is responsible for oxidative modification to Rubisco in tomato (*Lycopersicon esculentum*) under cadmium stress. J. Integr. Plant Biol. **50:** 415-426.

Lough A.N., Roark L.M., Kato A., Ream T.S., Lamb J.C., Birchler J.A., Newton K.J. (2008) Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. Genetics **178:** 47-55.

Martin W., Herrmann R.G. (1998) Gene transfer from organelles to the nucleus: How much, What happens, and Why? Plant Physiol. **118:** 9-17.

Martin W., Schnarrenberger C. (1997) The evolution of the Calvin cycle from prokarytic to eukaryotic chromosomes: A case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32:** 1-18.

Millen R.S., Olmstead R.G., Adams K.L., Palmer J.D., Lao N.T., Heggie L. et al. (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell **13:** 645-658.

Noutsos C., Kleine T., Armbruster U., DalCorso G., Leister D. (2007) Nuclear insertions of organellar DNA can create novel patches of functional exon sequences. Trends Genet. **123:** 597-601.

Pyrez-Martmez X., Vbzquez-Acevedo M., Tolkunova E. et al. (2000) Unusual location of a mitochondrial gene: Subunit III of cytochrome c oxidase is encoded in the nucleus of *Chlamydomonad algae*. J. Biol. Chem. **275:** 30144–30152.

Pfannschmidt T. (2003) Chloroplast redox signals: how photosynthesis controls its own genes. Trends Plant Sci. **8:** 33-41.

Popot J.-L., de Vitry C. (1990) On the microassembly of integral membrane proteins. Annual

Rev. Biophys. Biophys. Chim. **19:** 369-403.

Race H., Herrmann R., Martin W. (1999) Why have organelles retained genomes? Trends Genet. **15:** 364–370.

Sanchez H., Fester T., Kloska S., Schroder W., Schuster W. (1996) Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. EMBO J. **15:** 2138-2149.

Sandoval P., Leun G., Gumez 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.

Schuster W., Wissinger B., Hiesel R., Unseid M., Gerold E., Knoop V. et al. (1991) Between DNA and protein – RNA editing in plant mitochondria. Physiol. Plant. **81:** 437-445.

Stegemann S., Bock R. (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. Plant Cell **18:** 2869–2878.

Stupar R.M., Lilly J.W., Town C.D., Cheng Z., Kaul S., Buell C.R. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc. Natl. Acad. Sci. USA **98:** 5099-5103.

Subramanian S., Fallahi M., Bonen L. (2001) Truncated and dispersed *rpl2* and *rps19* pseudogenes are cotranscribed with neighbouring downstream genes in wheat mitochondria. Curr. Genet. **39:** 264-272.

Ueda M., Fujimoto M., Arimura S.-I., Tsutsumi N., Kadowaki K.-I. (2008) Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. Mol. Biol. Evol. **25:** 1791-1793.

von Heijne G. (1987) Why mitochondria need a genome. FEBS Lett. **198:** 1-4.

Wakasugi T., Tsudzuki J., Ito S. et al. (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA **91:** 9794-9798.

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

Wischmann C., Schuster W. (1995) Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. FEBS Lett. **374:** 152-156.

Zerges W. (2002) Does complexity constrains organelle evolution? Trends Plant Sci. **7:** 175-180.

Zhuo D., Nguyen-Lowe H.T., Subramanian S., Bonen L. (1999) The S7 ribosomal protein gene is truncated and overlaps a cytochrome c biogenesis gene in pea mitochondria. Plant Mol. Biol. **40:** 91-97.