MBE Advance Access published December 20, 2005

Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins

Linda Bonen* and Sophie Calixte

Biology Department, University of Ottawa,
Ottawa, Canada K1N 6N5

Research Article

Key words: Ribosomal protein, mitochondria, introns, evolution, gene transfer, plant, human

Running head: Nuclear-encoded mitochondrial-type ribosomal proteins

* Corresponding author:
Dr. Linda Bonen, Biology Department,
University of Ottawa, 30 Marie Curie,
Ottawa, Canada K1N 6N5

lbonen@science.uottawa.ca
FAX: (613) 562-5486
Tel: (613) 562-5800 ext. 6356

ABSTRACT

Mitochondrial ribosomes contain bacterial-type proteins reflecting their endosymbiotic heritage and a subset of these genes is retained within the mitochondrion in land plants. Variation in gene location is observed however, because migration to the nucleus is still an ongoing evolutionary process in plants. To gain insight into adaptation events related to successful gene transfer, we have compiled data for bacterial-origin mitochondrial-type ribosomal protein genes from the completely-sequenced *Arabidopsis* and rice genomes. Approximately 75% of such nuclear-located genes encode amino-terminal extensions relative to their *E. coli* counterparts, and of that set, only about 30% have introns at (or near) the junction in support of an exon shuffling-type recruitment of upstream expression/targeting signals. We find that genes which were transferred to the nucleus early in eukaryotic evolution have on average about two-fold higher density of introns within the core ribosomal protein sequences than do those that moved to the nucleus more recently. About 20% of such introns are at positions identical to those in human orthologs, consistent with their ancestral presence. Plant mitochondrial-type ribosomal protein genes have dispersed chromosomal locations in the nucleus, and about 20% of them are present in multiple unlinked copies. This study provides new insights into the evolutionary history of endosymbiotic bacterial-type genes which have been transferred from the mitochondrion to the nucleus.

INTRODUCTION

Mitochondria have their own distinctive translation systems to direct the synthesis of proteins encoded by the organelle genome, and central to this process is the ribosome. It comprises two subunits (denoted as LSU and SSU) made up of ribosomal RNAs and proteins, and traces of its bacterial endosymbiotic ancestry are evident from sequence comparisons (reviewed in Lang and Gray 1999). Proteomic analysis of mammalian mitoribosomes (Koc et al. 2001a, 2001b) has established the presence of 78 different proteins (49 LSU and 29 SSU), which include homologs for all but twelve of the 54 proteins in *E. coli* ribosomes (33 LSU and 21 SSU) (reviewed in O'Brien et al. 2005). Little is known about the role of the 36 additional novel proteins acquired during eukaryotic evolution, although it has been suggested that some may have taken over certain ribosomal RNA functions because animal mitochondrial 16S LSU and 12S SSU rRNAs are much shorter than in bacteria (cf. 23S and 16S rRNAs). In contrast, plant mitochondrial rRNAs are quite long (26S/18S), but as yet the protein composition of mitoribosomes in plants has not been fully elucidated (reviewed in Bonen 2004). It does appear however, that they too contain about 80 proteins based on early electrophoretic analysis of potato and broad bean mitochondrial ribosomes (Pinel et al. 1986; Maffrey et al. 1997), and the number listed in plant mitochondrial proteomic databases continues to grow (cf. Arabidopsis mitochondrial protein database, Heazlewood and Millar 2005).

During eukaryotic evolution, many of the bacterial-type genes, including ribosomal protein ones, have been transferred from the mitochondrion to the nucleus (reviewed in Lang et al. 1999; Timmis et al. 2004). Indeed in animals, all ribosomal protein genes are nuclear-located, while in plants and certain protists, some are still in the mitochondrion and in certain cases have retained remnants of ancestral bacterial operon gene order (reviewed in Lang et al. 1999). The

mitochondrial genome of the protist *Reclinomonas americana*, which has been described as resembling a "eubacterial genome in miniature" (Lang et al. 1997) has 27 bacterial-type ribosomal protein genes, and in virtually all other eukaryotes examined to date, their mitochondrial-encoded ones are a subset of the *Reclinomonas* ones (Gray et al. 2004). Sixteen such ribosomal protein genes are present in the mitochondrion of the bryophyte, *Marchantia polymorpha* (Takemura et al. 1992) and in flowering plants, the mitochondrial-encoded subset drops to a maximum of 14, namely S1-S4, S7, S10-S14, S16, L2, L5, L16 (reviewed in Adams et al. 2002a; Adams and Palmer 2003). Such comparisons have provided insight into the timing of ribosomal protein gene transfer events, and flowering plants are particularly interesting because functional gene transfer to the nucleus is still occurring (Adams et al. 2002a). Plant mitochondrial genes (unlike those in yeast or animals) use the standard genetic code so that present-day transfer is not precluded. However because their coding regions typically undergo amino acid-altering RNA editing and some genes contain group II-type introns, successful transfer presumably involves processed RNA intermediates. The acquisition of appropriate gene expression signals and targeting information to guide the protein back into the mitochondrion has been documented to be achieved in various ways (reviewed in Adams and Palmer 2003; Timmis et al. 2004), for example by recruiting duplicated copies of amino-terminal presequences from pre-existing mitochondrial-type nuclear genes (cf. S10, Adams et al. 2000) or hitch-hiking via alternative splicing when located within the intron of a "host" nuclear gene (cf. S14, Kubo et al. 1999; Figueroa et al. 1999). After a transition stage when potentially functional copies co-exist in both the mitochondrion and the nucleus, one copy presumably becomes superfluous and is lost. Indeed remnant ribosomal protein pseudogenes (or their complete absence) in various plant mitochondrial genomes have been correlated with recently re-located functional nuclear copies.

Interestingly, there is as yet just one characterized case of a ribosomal protein gene in a transition stage, namely L5 in wheat (Sandoval et al. 2004). If a nuclear gene copy is not successfully established after a period of co-existence, the functional mitochondrial copy will be retained, as is the case for S19 in rice in contrast to other cereals where the nuclear copy has "won out" (Fallahi et al. 2005). Furthermore, ribosomal protein genes can have surprisingly complex evolutionary histories. For example, in some plant lineages (such as *Arabidopsis*) the L2 gene has been fractured into two segments, with one encoded in the nucleus and the other in the mitochondrion (Adams et al. 2001), and there are intriguing cases of the recapture of alien S2 and S11 ribosomal protein genes after earlier loss from the mitochondrion (Bergthorsson et al. 2003).

Although absence of an intact ribosomal protein gene from the mitochondrion is usually correlated with its movement to the nucleus, other possibilities include the protein no longer being needed in the ribosome or its function having been taken over by another protein. In plants, obvious candidates for the latter are chloroplast-type nuclear genes, and indeed the recruitment of a duplicated chloroplast-origin S13 homolog has been experimentally documented for *Arabidopsis* (Adams et al. 2002b; Mollier et al. 2002). In contrast, there is still a native functional S13 gene copy in the rice mitochondrion (Notsu et al. 2002). It should be noted that the chloroplast ribosome contains 58 proteins (33 LSU and 25 SSU) based on spinach proteomic analysis (Yamaguchi et al. 2000; Yamaguchi and Subramanian 2000), of which only 22 are chloroplast-encoded. Six of the nuclear-encoded chloroplast ribosomal proteins have been categorized as "plastid-specific" (Yamaguchi and Subramanian 2003) and orthologs for bacterial L25 and L30 are absent. A second group of candidates for being recruited for a role in mitoribosomes are cytosol-type ribosomal proteins, and S8, a duplicated cytosol-origin homolog

appears to have replaced the "native" mitochondrial counterpart (Adams et al. 2002b). The timing of this event is placed early in land plant evolution because a bacterial-type gene is still present in *Marchantia* mitochondria but has not been observed in the mitochondria of flowering plants.

We are interested in evolutionary events which accompany successful transfer from the mitochondrion to the nucleus, in particular how gene structure may be influenced as the formerly-mitochondrial gene adapts to its new environment. To this end, we have compiled data for the bacterial-origin mitochondrial-type ribosomal protein genes from *Arabidopsis thaliana* (a eudicot) and *Oryza sativa* (rice, a monocot), which have been separated from a common ancestor for at least 150 million years. These plants were selected because complete sequence data are available for the mitochondrion (Unseld et al. 1997; Notsu et al. 2002), chloroplast (Sato et al. 1999; Hiratsuka et al. 1989) and nucleus (AGI 2000; Yu et al. 2002; Goff et al. 2002). In our analysis, we focus on a comparison between genes that were transferred to the nucleus early in eukaryotic evolution with those that migrated more recently.

MATERIALS AND METHODS

To identify potential plant mitochondrial ribosomal protein genes, tBLASTn searches were conducted using *E. coli* ribosomal protein sequences (Blattner et al. 1997) to query the *Arabidopsis* and rice nuclear genomes. This query set of 54 proteins did not include *E. coli* S22 (AAC74553), a stationary phase-induced ribosome-associated protein which is restricted to a few bacterial lineages (not including α-proteobacteria). Note that in the chloroplast nomenclature S22 is a different protein (Yamguchi et al. 2000). We also conducted tBLASTn and BLASTp searches using mitochondrial proteins from *Marchantia* (Takemura et al. 1992) *Reclinomonas*

(Lang et al. 1997), human (Koc et al. 2001a, 2001b) and yeast (Gan et al. 2002), the latter two sets being nuclear-encoded. Proteomic data from spinach chloroplast (Yamaguchi et al. 2000; Yamaguchi and Subramanian 2000; Yamaguchi and Subramanian 2003) were also useful in distinguishing chloroplast homologs from potential mitochondrial ones. In the case of rice, certain ribosomal protein sequences lacked protein databank entries, and tBLASTx searches of EST and nr databanks using *Arabidopsis* proteins as query enabled identification of putative rice homologs. Alignments of these ribosomal proteins from *Arabidopsis*, rice, *E. coli* and human were carried out using CLUSTALW, with adjustments by visual inspection in certain cases. In this way, the junctions between core (ancestral) and non-core (acquired) coding regions were also evaluated.

Intron positions within the coding regions of the ribosomal protein genes were determined by comparisons of mRNA (or EST) data with genomic sequences. In the case of *Arabidopsis* annotated information was available at NCBI, although some genes were listed as unknown function or were mis-annotated as putative chloroplast ribosomal proteins. Similarly, using the annotated human genome entries for mitochondrial ribosomal proteins (MRP), the positions of introns were compared to those in plant homologs.

The potential mitochondrial targeting properties of the *Arabidopsis* nuclear-encoded mitochondrial-type ribosomal proteins were assessed using the algorithms Predotar v.1.03 http://www.inra.fr/predotar/ (Small et al. 2004), TargetP1.1 http://www.cbs.dtu.dk/services/TargetP/ (Emanuelsson et al. 2000) and PSort http://psort.nibb.ac.jp/ (Bannai et al. 2002), with default settings and “plant” being selected as the organism group.

RESULTS AND DISCUSSION

Identification of bacterial-origin mitochondrial-type ribosomal protein genes in Arabidopsis and rice genomes

To identify candidate genes for plant mitochondrial ribosomal proteins derived from the ancestral α-proteobacterial-type endosymbiont (cf. Esser et al. 2004), we searched the *Arabidopsis* and rice nuclear genomes for homologs to the 54 *E. coli* ribosomal proteins. This was accompanied by comparisons with plant chloroplast proteomic data to eliminate any known chloroplast homologs. Similarly, proteomic data for mitochondrial ribosomes from mammals and yeast were useful in corroborating designation of mitochondrial-specific ones. In addition, it is known from the *Arabidopsis* and rice mitochondrial genomic sequencing (Unseld et al. 1997; Notsu et al. 2002) that 7 and 11 ribosomal proteins respectively, are still encoded within their mitochondria (Figure 1). These data are summarized in Table I.

For the 54 *E. coli* ribosomal proteins, we identified 48 mitochondrial-type counterparts in rice and 46 in *Arabidopsis*, the difference being due to the presence of S1 and S13 gene orthologs in the mitochondrial genome of rice but in neither the mitochondrial nor nuclear genomes in *Arabidopsis*. However, a chloroplast-type S13 protein encoded by a divergent duplicated nuclear gene has been demonstrated to be targeted to the mitochondrion in *Arabidopsis* (Adams et al. 2002b; Mollier et al. 2002) and in an analogous fashion it is possible that one of several S1 chloroplast-origin nuclear genes (cf. Table I) may have been recruited for mitochondrial use in the *Arabidopsis* lineage. Similarly for S8, there is experimental evidence that the mitochondrial-type has been functionally replaced by a divergent cytosol-origin S8 (denoted S15a in cytosol nomenclature) in flowering plants (Adams et al. 2002b). It is possible

that the five remaining ones (S20, L25, L31, L34, L35, Figure 2, grey bars) escaped detection because of sequence divergence or that they are absent from the plant mitoribosome. An alternative possibility for S20, L31, L34 and L35 is that nuclear-encoded chloroplast-type homologs are dual targeted to both compartments. The sole exception is L25 and this protein is also absent from animal/yeast mitoribosomes as well as chloroplast ribosomes.

The "core" bacterial-origin coding regions of these plant mitochondrial-type ribosomal proteins were determined by alignment with *E. coli* homologs (Supplementary data available upon request) and are shown as black bars in Figure 2, whereas "non-core" extensions are shown in white. In several cases, the core sequences are slightly shorter than the length of the *E. coli* protein, but in only two cases do bacterial lengths greatly exceed the mitochondrial ones. The *E. coli* L6 protein has an amino-terminal extension of ~80 amino acids relative to the *Marchantia* and flowering plant mitochondrial counterparts, and the *E. coli* S1 is approximately 340 amino acids longer than in rice mitochondria, with a long carboxy-terminal extension. Several of the mitochondrial-encoded ribosomal proteins (notably S2, S3, S4 and L2) have extra domains relative to *E. coli*, and in the case of S2, the mature protein is known to lack the carboxy-terminal extension (Perotta et al. 2002).

The plant mitochondrial-type ribosomal proteins were divided into four categories designated as A-D (Figure 1 and Table I) based on their evolutionary history with respect to timing of gene transfer from the bacterial-type endosymbiont to the nucleus during eukaryotic evolution. This was inferred by absence of functional ribosomal protein genes in the mitochondrial genome. There are 26 genes placed in category A ("early transfer"), that is ones which have not been found in the mitochondrial genome of any eukaryote to date, based on the jakobid protists, *Reclinomonas ameriana* and *Malawimonas jakobiformis*, of which the former

has the largest known set of ribosomal protein genes (Gray et al. 2004). Secondly, there are 14 ribosomal proteins in category B ("mid-era transfer") based on genes still present in the *Reclinomonas* mitochondrion but not in those of land plants, as indicated by absence in *Marchantia* (although it is not excluded that there have been certain *Marchantia*-lineage or *Reclinomonas*-lineage specific losses). Among angiosperms, there are 14 ribosomal protein genes that have been identified, albeit with lineage-specific variation in their number (Adams et al. 2002a) and in our analysis of rice and *Arabidopsis*, they have been divided into category C ("late transfer") when the gene is nuclear-located in one or both of these plants or category D ("native mitochondrial") when the gene is still in the mitochondrion of both these plants. The particularly mobile nature of these genes in flowering plants is illustrated by the presence of 11 in the mitochondrion of rice compared to 7 for Arabidopsis, and for several of them, namely S10, S11 and S14 (Wischman and Schuster 1995; Kubo et al. 2000; Kadowaki et al. 1996; Figueroa et al. 1999), there is compelling support for independent transfer events having occurred in these two plant lineages (Figure 1).

Intron distribution and density in nuclear-located mitochondrial-type ribosomal protein genes

Using amino acid sequence data in conjunction with genomic and EST information from *Arabidopsis* and rice, we have determined the positions of introns within ribosomal protein coding regions (Figure 2) and this is summarized in Table I. A total of 109 introns were identified within the coding sequences of the genes for 40 of the 47 different ribosomal protein genes which are located in the nucleus in *Arabidopsis* and/or rice (Figure 2A-C, white triangles) as well as two group II-type introns within mitochondrial genes (Figure 2D, grey triangles).

Notably 7 genes (3 in category A and 4 in category B) contain no introns within the coding regions, although several of these do have an intron within non-translated sequences (cf. Table I, numbers in parentheses). Any introns located within bacterial-origin core sequences (Figure 2, black bars) presumably were acquired after transfer to the nucleus whereas ones within non-core regions might instead have been obtained at the time of integration in the nuclear genome. For category A (early transfer) genes, the possibility that the endosymbiont had pre-existing (ribozymic) introns which evolved into spliceosomal-type introns in the nucleus is not excluded (cf. Cavalier-Smith 1991).

With respect to introns within core regions, the *Arabidopsis* and rice gene homologs show virtually identical profiles for intron positions, the one exception being L9 which has an extra intron in rice (Figure 2, dotted oval). Incidentally this intron shows no sequence similarity with other (neighbouring) introns or transposable elements (cf. Coghlan and Wolfe 2004), and may reflect either recent gain in the rice lineage or loss from the *Arabidopsis* one. Overall, intron positions appear relatively stable during flowering plant evolution, although variation in intron length is seen between the two plants, with *Arabidopsis* counterparts typically being shorter (Figure 2, triangle size). For example, *Arabidopsis* has only one ribosomal protein intron longer than 1 kb (within S6) whereas rice has over 20 introns greater than that length. When considering intron positions within non-core coding sequences for categories A and B, several differences are seen between *Arabidopsis* and rice (Figure 2, dotted ovals). As expected for category C (recent transfer), there is even greater difference in gene structure between rice and *Arabidopsis* counterparts, reflecting lineage-specific independent transfer events.

To address the question as to whether transferred genes tend to accumulate introns over evolutionary time, we compared the intron density within the core bacterial-origin regions (that

is coding sequences which migrated from the mitochondrion rather than being acquired in the nucleus as is the case for non-core coding sequences) of early-transferred genes (category A) with the more recently transferred gene set (category B). Notably in category C, no introns are located within the core regions. Of the 73 core-type introns, the density was 1.7 introns/100 codons for category A compared to a value of 0.8 for category B (Figure 3). Thus, within the core, the early-transferred genes have about 2-fold higher intron density than those transferred more recently. In contrast, within the non-core coding regions (which include introns close to the junction), intron density is lower and less biased, showing values of 0.7 and 1.0 introns/100 codons, respectively, for the early and more recently transferred gene sets. Incidentally for the non-core regions, there is also a higher number of amino-terminal-located introns than carboxy-terminal ones (Figure 2).

To extend our investigation of intron gain/loss to a longer evolutionary time period, we compared the plant intron-exon organization with human mitochondrial homologs. Accurate alignment was in some cases difficult because of low amino acid similarity or indels, so we restricted our analysis to those exhibiting strong amino acid conservation immediately flanking the intron position. For this set of 73 introns (omitting genes which have no orthologs in humans), 15 are located at identical positions (Figure 2, black stars; Figure 4A) and an additional 7 are located within 6 nucleotides of each other (Figure 2, white stars; Figure 4B). The latter group show features which have been referred to as “intron sliding” and all but one of them are within the same genes as those having identical intron positions. They represent 11 different ribosomal protein genes, all except one being in category A, consistent with presence in the common ancestor of plants and animals, so dating back over one billion years or so. This set includes the L9 gene which has an extra intron in rice compared to *Arabidopsis* (Figure 2, dotted

oval), and that intron is located 12 nt away from the position of one in human MRPL9 (Figure 4C). Interestingly, in both plants and humans, the L9 gene also has an intron near the amino-terminal core/non-core junction, and the similarity in amino acid sequence (Figure 4C) raises the possibility of a shared history.

Typically the human genes have a greater number and much longer introns than their plant counterparts (cf. one example being MRPS6-intron 1 which is 51.6 kb long). In addition, there are numerous examples of alternatively-spliced transcripts for the human ribosomal protein genes, but relatively few for plant ones. Interestingly, one of the *Arabidopsis* duplicated L14 genes (designated as HLP, Skinner et al. 2001) has CAG triplet repeats at the intron 2/exon 3 junction (Figure 4D, underlined) and alternatively spliced transcripts have been reported. Moreover, a nonsense mutation in the HLL (HUELLENLOS) coding sequence (CAA to TAA, Figure 4D in black) has been found to be correlated with defective ovule growth and development (Skinner et al. 1996). None of the chloroplast-type ribosomal protein genes has introns at positions identical to the mitochondrial-origin ones, with the exception of L21 which originated from a duplicated mitochondrial-type L21 gene (Gallois et al. 2001). The chloroplast-type genes also have fewer introns, perhaps reflecting their shorter evolutionary history.

Acquisition of mitochondrial targeting signals for nuclear-encoded ribosomal proteins

Because the signals that target nuclear-encoded proteins to the mitochondrion are in many cases located at the amino-terminus, we examined their potential targeting capacity. About 75% of the ribosomal proteins possess amino-terminal extensions relative to their *E. coli* counterparts, and for the rest, presumably the “native” sequences possess the required features (either at the amino-terminus or elsewhere along the protein) to enable import into the mitochondrion. The lengths of

non-core amino-terminal extensions vary markedly among these ribosomal proteins and in several cases they are 2-3 fold longer than the core ribosomal protein region (Figure 2, white vs. black bars) so may well encode other important information in addition to targeting signals. The long amino-terminal extension of S5 had been postulated to compensate for the absence of S4 and S8 in human mitoribosomes (Koc et al. 2001a), however, that is brought into question because S4 and S8 counterparts are still present in plant mitoribosomes. In several cases, such as L7/L12, L15, and L32, homology was detected in non-core regions with mitochondrial counterparts from other eukaryotes such as animals and fungi, indicating that these sequences were acquired early in eukaryotic evolution (or perhaps were even present in the ancestral α–proteobacteria).

Since the acquisition of non-core sequences could occur through exon shuffling, we scored cases in which introns are located at the core/non-core junction. In some cases, the precise position of the breakpoint in homology between the bacterial and mitochondrial homologs was difficult to map, and we ranked as potentially coincident if located within 50 bp of the junction. By these criteria, for 14 of the 50 ribosomal protein gene transfer events (that is, approximately 30%) there is an intron at the core/non-core amino-terminal junction (Figure 2).

A number of computer algorithms have been developed for predicting the targeting locations of nuclear-encoded proteins, and for plants there is the added complexity (compared to animals or fungi) of distinguishing between routing to the chloroplast and the mitochondrion. When we subjected our data set to Predotar (Small et al. 2004), TargetP (Emanuelsson et al. 2000) and PSort (Bannai et al. 2002), we observed that while the majority of the ribosomal proteins showed strong prediction for targeting to the mitochondrion (Table I, + symbols), not all did, and the results for the three programs differed somewhat in certain cases (cf. discussion in

Heazlewood et al. 2004). Among those failing to yield high predictive values for targeting to the mitochondrion are ones for which only chloroplast-origin homologs could be found in the nuclear genome (Table I, asterisks).

When we compared the amino-terminal sequences of the *Arabidopsis* and rice ribosomal proteins in categories A and B (that is, genes transferred prior to divergence of these two plant lineages), we observed a wide variation in the degree of amino acid similarity (Supplementary data available upon request). Certain proteins exhibited high sequence identity within this region (eg. L29, L14) consistent with a common origin, whereas others (eg. L15, L10) appeared unrelated. The latter might reflect divergence due to relatively low constraint at the amino acid level or alternatively, there may have been lineage-specific shuffling events which generated non-homologous targeting sequences. Indeed this is further supported in cases where intron position differs between these two plants within the non-core regions (eg. S18, L19). As sequence data from additional plants become available, it will be of interest to pursue such issues in greater detail.

Duplicated genes for mitochondrial-type ribosomal protein genes and dispersed chromosomal locations in the *Arabidopsis* nuclear genome

The mitochondrial-type ribosomal protein genes are dispersed in the nuclear genome and their locations on the five *Arabidopsis* chromosomes are shown in Figure 5. The closest two ribosomal protein genes in *Arabidopsis* are L15 and L17 (about 13 kb apart on chromosome 5), however in rice they are on chromosomes 1 and 8. On the other hand, the duplicated L27 genes are in tandem in rice, being separated by only 4.4 kb on chromosome 8, whereas in *Arabidopsis* they are on separate chromosomes.

Among the 48 ribosomal protein genes which are nuclear-located in *Arabidopsis* and/or rice, eleven are present in multiple copies (ranging from 2 to 4) in one or both plants (Table I). For example, there are single copies of S10 and S11 in the nucleus of *Arabidopsis*, whereas both are present in duplicate copies in rice. The evolutionary history of the rice S11 duplication event which has been characterized in detail (Kadowaki et al. 1996) includes the acquisition of different targeting sequences. Some of the duplicated copies are very similar within the core (eg. 97%-100% for L17, L19, L22, L27) whereas others are quite divergent (eg. 62-74% for S16, L7/L12, L14). All have retained the same intron positions within core sequences, although several differ within untranslated regions (cf. L7/L12 copies, Table I). The duplicated genes include representatives from categories A-C (Figure 2, asterisks) in rather similar proportions (namely 6, 3 and 2 cases out of 25, 14 and 8 genes, respectively) so it does not appear that the early-transferred genes have undergone a greater number of duplication events, at least over this evolutionary time period.

CONCLUDING REMARKS

For the vast majority of the 54 proteins present in α-proteobacterial ribosomes, candidate genes for counterparts in plant mitochondrial ribosomes (as represented by *Arabidopsis* and rice) can be identified in either the nuclear or mitochondrial genome. The latter comprise a relatively small subset reflecting ancestral endosymbiont-type genes still retained in the organelle, with the others having been transferred to the nucleus over evolutionary time. For the remaining bacterial-origin type, it is possible that sequence divergence precluded their detection, however there are nuclear-located chloroplast-origin genes which might have been co-opted to substitute for mitochondrial ones, and indeed there is experimental evidence for this having occurred (Adams

et al. 2002b; Mollier et al. 2002) as well as for the recruitment of a cytosol-type homolog (Adams et al. 2002b). The sole exception is L25 which notably is absent from other organellar ribosomes, such as chloroplast or mammalian mitochondrial ones based on proteomic analysis (Koc et al. 2001b;Yamaguchi and Subramanian 2000). Thus plant mitoribosomes appear more bacterial-like than mammalian ones which lack 12 of the 54 proteins found in present-day bacterial ribosomes, including eight (namely S3, S4, S8, S13, S19, L5, L6 and L29) which have been considered "universal" among all life-forms (Lecompte et al. 2002). Interestingly, of those, all but the L6 and L29 genes are deduced to have been still present within the mitochondrion at the time of the plant-animal lineage split (Figure 1). This illustrates different evolutionary pathways that mitochondrial ribosomes have followed in various eukaryotes, and it is worth noting that as in mammals, the plant mitoribosomes are estimated to contain a total of about 80 proteins, due to the acquisition of additional novel ones during eukaryotic evolution. Even though plant mitochondrial ribosomal proteins (and ribosomal RNAs) have retained stronger bacterial-origin features than in mammals, certain relatively recent events within the land plant lineage illustrate plasticity in the make-up of the ribosome. In addition to "native" mitochondrial proteins being replaced by chloroplast (or cytosol) homologs there are reports of lateral transfer of mitochondrial ribosomal protein genes between distantly-related plant species.

About 75% of the nuclear-encoded plant mitochondrial ribosomal proteins have amino-terminal extensions compared to their bacterial orthologs, and some are much longer than typical amino-terminal targeting signals. This raises the possibility that such domains (which also include internal and carboxy-terminal ones) perform extra functions. They might have structural roles (eg. related to the longer rRNAs in plant mitochondria or ribosomal assembly) or regulatory ones (such as translational control analogous to yeast mitochondria, cf. Williams et al.

2004). Alternatively they may be involved in extraribosomal functions (Wool 1996) such as RNA level events like editing or splicing. Indeed RNA binding motifs have been identified, for example the RRM domain in *Arabidopsis* S19 (Sanchez et al. 1996). It will also be of interest to learn more about not only the function of these extra domains, but also their origin and evolutionary history.

With respect to the gene structure of these nuclear-located genes, their intron/exon organization is for the most part conserved not only among plants (as exemplified by rice and *Arabidopsis*), but also in some cases with human ribosomal protein counterparts. About 20% of the introns are at identical positions (or approximately 30% if near-identical locations are included) in the human and plant genes, suggesting that they have been present for at least 300 million years. In other studies of intron positions in eukaryotic genes, it has similarly been concluded that about 14% (Fedorov et al. 2002) and about 24% (Rogozin et al. 2003) are at identical positions in plant-animal orthologs. There are more introns within core bacterial-origin regions than in acquired coding extensions, and this appears to increase over evolutionary time based on comparisons of intron density in genes transferred early in eukaryotic evolution with those migrating to the nucleus later on. Only about 30% show support for intron-mediated acquisition of upstream sequences (that is, an exon shuffling model).

Moreover, the complete absence of amino-terminal extensions in a subset of both the early and recent transfer categories, raises questions about the nature of targeting information. For very recently-transferred genes, it has sometimes been possible to trace the origin of acquired sequences, for example through the recruitment of duplicated sequences from other mitochondrial-targeted genes (reviewed in Adams and Palmer 2003). In addition, there is growing evidence that targeting of mRNAs for translation on the surface of the mitochondrion

may be important for protein import (cf. Sylvestre et al. 2003). Proteomic analysis of plant mitochondrial ribosomes will also undoubtedly yield new insights into the mechanistic and co-evolutionary processes that enable the “native” and “acquired” ribosomal components to perform their essential roles in translation.

ACKNOWLEDGEMENTS

We thank O. Antoniuk and G. Voros for their valuable assistance in compiling data from the rice and *Arabidopsis* nuclear genomes, respectively. Financial support from the Natural Sciences and Engineering Research Council of Canada is also gratefully acknowledged.

REFERENCES

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

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., Daley, D.O., Whelan, J., Palmer, J.D. (2002b) Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. Plant Cell **14**: 931-943.

Adams, K.L., Ong, H.C., Palmer, J.D. (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., Qiu, Y.L., Stoutemyer, M. Palmer, J.D. (2002a) 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.

Arabdiopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**: 796-815.

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinform **18**:298-305

Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D. (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature **424**: 197-201.

Blattner, F.R., Plunkett, G., Bloch, C.A. et al. (1997) The complete genome sequence of Escherichia coli K-12. Science **277**: 1453-1474.

Bonen, L. (2004) Translational machinery in plant organelles. Pp.323-345 *in* H. Daniell and C. Chase, eds. Molecular Biology and Biotechnology of Plant Organelles. Springer, the Netherlands.

Cavalier-Smith, T. (1991) Intron phylogeny: a new hypothesis. Trends Genet. **7**: 145-8, 1991.

Coghlan, A., Wolfe, K.H. (2004) Origins of recently gained introns in *Caenorhabditis.* Proc. Natl. Acad. Sci. USA **101**: 11362-11367.

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

Esser, C., Ahmadinejad, N., Wiegand, C. et al. (2004) A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol. Biol. Evol. **21**: 1643-1660.

Fallahi, M., Crosthwait, J., Calixte, S., and Bonen, L. (2005) Fate of mitochondrially located S19 ribosomal protein genes after transfer of functional copy to the nucleus in cereals. Mol. Genet. Genom. **273**: 76-83.

Fedorov, A., Merican, A.F., Gilbert, W. (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. Proc. Natl. Acad. Sci. USA **99**: 16128-16133.

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.

Figueroa, P., Gomez, I., Carmona, R., Holuigue, L., Araya, A. and Jordana, X. (1999) The gene for mitochondrial ribosomal protein S14 has been transferred to the nucleus in *Arabidopsis thaliana*. Mol Gen Genet **262**: 139-144.

Gallois, J.L., Achard, P., Green, G., Mache, R. (2001) The Arabidopsis chloroplast ribosomal L21 is encoded by a nuclear gene of mitochondrial origin. Gene **274**: 179-185.

Gan, X., Kitakawa, M., Yoshino, K., Oshiro, N., Yonezawa, K., Isono, K. (2002) Tag-mediated isolation of yeast mitochondrial ribosome and mass spectrometric identification of its new components. Eur. J. Biochem. **269**: 5203-5214.

Goff, S.A., Ricke, D., Lan, T.H. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296**: 92-100.

Gray, M.W., Lang, B.F., Burger, G. (2004) Mitochondria of protists. Ann. Rev. Genet. **38**: 477-524.

Heazlewood, J.L, Tonti-Filippini, J.S., Gout, A.M., Day, D.A., Whelan, J., Millar, A.H. (2004) Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* **16**: 241-256.

Heazlewood, J.L., Millar, A.H. (2005) AMPDB: the Arabidopsis mitochondrial protein database. Nucl. Acids Res. **13**: D605-D610.

Hiratsuka, J., Shimada, H., Whittier, R. et al. (1989) The complete sequence of the rice (Oryza sativa) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217**: 185-194.

International Rice Genome Sequencing Project (2005) The map-sequence of the rice genome. Nature **436**: 793-800.

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.

Koc, E.C., Burkhardt, W., Blackburn, K., Moseley, A., Spremulli, L.L. (2001a) The small subunit of the mammalian mitochondrial ribosome: identification of the full complement of ribosomal proteins present. J. Biol. Chem **276**: 19363-19374.

Koc, E.C., Burkhart, W., Blackburn, K., Moyer, M.B., Schlatzer, D.M., Moseley, A., Spremulli, L.L. (2001b) The large subunit of the mammalian mitochondrial ribosome: analysis of the complement of ribosomal proteins present. J. Biol. Chem. **276**: 43958-43969.

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

Kubo, N., Jordana, X., Ozawa, K., Zanlungo, S., Harada, K., Sasaki, T., Kadowaki, K. (2000) Transfer of the mitochondrial *rps10* gene to the nucleus in rice: acquisition of the 5' untranslated region followed by gene duplication. Mol Gen Genet **263**: 733-739.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux C., Sankoff, D., Turmel, M., Gray, M.W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387**: 493-497.

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

Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., Poch, O. (2002) Comparative analysis of ribosomal proteins in complete genomes : an example of reductive evolution at the domain scale. Nucl Acids Res **30**: 5382-5390.

Maffey, L., Degand, H, Boutry, M. (1997) Partial purification of mitochondrial ribosomes from broad bean and identification of proteins encoded by the mitochondrial genome. Mol Gen Genet **254**: 365-371.

Mollier, P., Hoffmann, B., Debast, C., Small, I. (2002) The gene encoding *Arabidopsis thaliana* mitochondrial ribosomal protein S13 is a recent duplication of the gene encoding plastid S13. Curr. Genet **40**: 405-409.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., Kadowaki, K. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol. Genet. Genom. **268**: 434-445.

O'Brien, T.W., O'Brien, B.J., Norman, R.A. (2005) Nuclear MRP genes and mitochondrial disease Gene (epub May 19, 2005).

Perrotta, G., Grienenberger, J.M., Gualberto, J.M. (2002) Plant mitochondrial *rps2* genes code for proteins with a C-termimal extension that is processed. Plant Mol. Biol. **50**: 523-533.

Pinel, C, Douce, R., Mache, R. (1986) A study of mitochondrial ribosomes from the higher plant Solanum tuberosum L. Mol Biol Rep **11**: 93-97.

Rogozin, I.B., Wolf, Y.I., Sorokin, A.V., Mirkin, B.G., Koonin, E.V. (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr. Biol. **13**: 1512-1517.

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., Leon, G., Gomez, 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.

Sato, S., Nakamure, Y., Kaneko, T., Asamizu, E., Tabata, S. (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. **6**: 283-290.

Skinner, D.J., Baker, S.C., Meister, R.J., Broadhvest, J., Schneitz, K., Gasser, C.S. (2001) The Arabidopsis *HUELLENLOS* gene, which is essential for normal ovule development, encodes a mitochondrial ribosomal protein. Plant Cel*l* **13**: 2719-2730.

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4**:1581-1590.

Sylvestre, J., Margeot,A., Jacq, C., Dujardin, G., Corral-Debrinski, M. (2003) The role of the 3' untranslated region in mRNA sorting to the vicinity of mitochondria is conserved from yeast to human cells. Mol. Biol. Cell **14**: 3848-3856.

Takemura, M., Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Nozato, N., Akashi, K., Ohyama, K. (1992) Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. Nucl. Acids Res. **20**: 3199-3205.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Rev. Genet. **5**: 123-135.

Unseld, M., Marienfeld, J.R., Brandt, P., Brennicke, A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 gnes in 366,924 nucleotides. Nature Genet. **15**: 57-61.

Williams, E.H., Perez-Martinez, X., Fox, T.D. (2004) MrpL36p, a highly diverged L31 ribosomal protein homolog with additional functional domains in *Saccharomyces cerevisiae* mitochondria. Genet. **167**: 65-75.

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 intergration site. FEBS Lett. **374**: 152-156.

Wool, I.G. (1996) Extraribosomal functions of ribosomal proteins. Trends Biochem. Sci. **21**: 164-165.

Yamaguchi, K., Subramanian, A.R. (2000) The plastid ribosomal proteins: Identification of all the proteins in the 50S subunit of an organelle ribosome (chloroplast) J Biol Chem **275**: 28466-28482.

Yamaguchi, K., Subramanian, A.R. (2003) Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30S ribosomal subunit. Eur J Biochem **270**: 190-205.

Yamaguchi, K., von Knoblauch, K., Subramanian, A.R. (2000) The plastid ribosomal proteins: Identification of all the proteins in the 30S subunit of an organelle ribosome (chloroplast) J Biol Chem **275**: 28455-28465.

Yu, J., Hu, S., Wang, J. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). Science **296**: 79-92.

Table I (part 1)

	Category [b]	A.t. mt locus tag [c]	Introns [d]	A.t. protein	Protein aa	Targeting [e]	Rice protein [f]	Protein aa	A.t. chl tag [c]	Comments
S1 [a]	C	At5g30510	6	NP_850903	416	chl	BAC19892	209 [g]	At5g30510	h, At1g71720 [i]
S2	C	At3g03600	0	NP_187010	219	+	BAC19874	483	chl	
S3	D	mt	1	NP_085481	556		BAC19869	538	chl	h
S4	D	mt	0	NP_085497	362		BAC19883	352	chl	h
S5	A	At1g64880	7	NP_564842	515	+++	BAD08197	501	At2g33800	
S6	A	At3g18760	1	NP_566618	139	other	BAA96212	146	At1g64510	
S7	D	mt	0	P92557	148		BAC19857	148	chl	
S8 [a]	B	At4g29430	2	NP_194672	129	+++	BAD29691	129	chl	h, cytosol –type [j]
S9	A	At3g49080	3	NP_190477	430	+++	BAC81159	415	At1g74970	
S10	C	At3g22300	1	NP_188869	241	++	BAB00643	112	At3g13120	rice S10-2 (BAB00644) [k]
S11	C	At1g31817	1	NP_564385	314	+++	BAA12798	254	chl	rice S11-2 (BAB93272) [l]
S12	D	mt	0	P92532	125		BAC19872	125	chl	
S13 [a]	C	At1g77750	2	NP_177898	154	+++	BAC19882	116	At5g14320	h, chl-type in A.t. [j]
S14	C	At2g34520	(+1)	NP_850232	164	+++	BAD09056	350	chl	within rice sdh2 intron, [m]
S15-1	A	At1g15810	4	NP_563982	419	+++	BAD36035	447	chl	
S15-2	A	At1g80620	4	NP_849914	414	chl				91% vs S15-1 core
S16-1	A	At5g56940	2	NP_200504	135	+++	NP_913921	137	chl	
S16-2	A	At4g34620	2	NP_195188	113	+++				64% vs S16-1 core
S17-1	A	At1g49400	1 (+1)	NP_175365	116	other	BAD03787	107	At1g79850	
S17-2	A	At3g18880	1 (+1)	NP_188518	105					91% vs. S17-1 core
S18	A	At1g07210	4	NP_172201	261	+++	NP_922183	274	chl	
S19	C	At5g47320	3(+1)	NP_568681	212	+	BAC19887	93	chl	h, RRM domain in A.t. [n]
S20 [a]	A	At3g15190	2	NP_188137	202	chl	NP_916920	196	At3g15190	h
S21	A	At5g63300	0	NP_201135	217	chl	XP_479834	121	At3g27160	

Table I (part 2)

	Category [b]	A.t. mt locus tag [c]	Introns [d]	A.t. protein	Protein aa	Targeting [e]	O.s. protein [f]	Protein aa	A.t. chl tag [c]	Comments
L1	B	At2g42710	9	NP_850375	415	+++	(AK067243)	418	At3g63490	
L2-5'	D	mt	0	NP_085516	346		BAC19886	502	chl	A.t."half"-gene [o]
L2-3'	C	At2g44065	1 (+4)	NP_973685	214	+++	BAC19886	"	chl	A.t. "half"-gene [o]
L3	A	At3g17465	4	NP_566579	324	+++	NP_912745	319	At2g43030	
L4	A	At2g20060	7 (+1)	NP_565463	300	+++	(AK069751)	286	At1g07320	
L5	D	mt	0	NP_085491	185		BAC19898	188	At4g01310	h
L6	B	At2g18400	(+1)	NP_565438	102	other	XP_483021	103	At1g05190	h
L7/L12-1	A	At1g70190	0	NP_564986	179	+++	XP_475875	198	At3g27830	
L7/L12-2	A	At4g36420	0 (+1)	NP_195360	179	+++	AAT78788	175		63% vs L7/L2-1 core
L7/L12-3	A	At4g37660	0	NP_195481	167	+++	NP_917913	167		73% vs L7/L12-1 core
L7/L12-4	A	At3g06040	(+1)	NP_187255	186	+++	XP_464496	193		74% vs L7/L2-1 core
L9	A	At5g53070	4 (+1)	NP_200119	221	+++	AAU43965	206	At3g44890	
L10	B	At5g13510	0	NP_196855	220	chl	AAC64971	220	At3g12370	
L11	B	At4g35490	0	NP_195274	155	+++	AAP54191	155	At1g32990	At5g51610 [i]
L13	A	At3g01790	4 (+1)	NP_850494	205	other	XP_480211	229	At1g78630	
L14-1	B	At5g46160	3	NP_851140	196	+++	AAL60450	170	chl	HLP [p]
L14-2	B	At1g17560	2	NP_173200	173	+				HLL [p] 62% vs L14-1 core
L15	A	At5g64670	3	NP_201272	281	+	BAD32079	295	At3g25920	
L16	D	mt	0	NP_085480	179		BAA04793	185	chl	
L17-1	A	At5g09770	5	NP_568216	160	+++	BAD81752	161	At3g54210	
L17-2	A	At5g64650	5	NP_568992	160	+++				97% vs L17-1 core
L18	B	At5g27820	0 (+1)	NP_198134	114	++	XP_469358	114	At1g48350	
L19-1	B	At1g24240	4	NP_564213	222	++	XP_464558	222	At5g47190	
L19-2	B	At5g11750	4	NP_192900	225	+				97% vs L19-1 core
L19-3	B	At4g11630	4	NP_196736	229	other				97% vs L19-1 core
L20	B	At1g16740	1	NP_173118	126	+	NP_916444	121	chl	
L21	A	At4g30930	4	NP_567861	270	+++	XP_475904	250	At1g35680	chl=ancient mt [q]
L22-1	A	At1g52370	5 (+1)	NP_564605	269	++	BAD61282	278	chl	
L22-2	A	At4g28360	5	NP_567805	271	+++				98% vs. L22-1 core
L23	A	At4g39880	0	NP_195698	178	+	XP_507362	229	chl	
L24	A	At5g23535	4 (+1)	NP_680212	159	+	(AK121443)	190	At5g54600	
L27-1	B	At2g16930	2 (+1)	NP_565398	154	+++	XP_471831	145	At5g40950	
L27-2	B	At5g15220	2(+1)	NP_568310	154	+++	BAC99362	145		100% vs cL27-1 core
L28	A	At4g31460	1 (+1)	NP_194874	212	other	XP_478628	204	At2g33450	
L29	A	At1g07830	2 (+1)	NP_172261	144	+++	BAD44786	145	At5g65220	h
L30	A	At5g55140	1 (+1)	NP_568821	109	++	XP_479729	116	absent	
L31 [a]	B	At1g75350	1	NP_565109	144	other	NP_917838	135	At1g75350	h
L32	B	At1g26740	2	NP_564259	134	+++	AAL84312	113	chl	
L33	A	At3g06320	0 (+1)	NP_187283	58	++	(AK121541)	57	chl	
L34 [a]	B	At1g29070	1	NP_174202	157	chl	NP_915866	159	At1g29070	
L35 [a]	A	At2g24090	2	NP_850047	145	chl	BAD37618	146	At2g24090	
L36	B	At5g20180	1 (+3)	NP_850857	103	++	XP_470053	98	chl	r

TABLE I

Compilation of candidate bacterial-origin mitochondrial-type ribosomal proteins in *Arabidopsis* and rice. [a] Ribosomal proteins which lack identifiable "native mitochondrial" homologs in *Arabidopsis* (A.t.) and/or rice (O.s.), but for which chloroplast-origin or cytosol-type substitute candidates are present (see text). [b] Based on predicted time of gene transfer to nucleus (see Figure 1 and text). [c] Genes encoded within the *Arabidopsis* or rice mitochondrial genomes (NC_001284 and BA000029, respectively) are denoted by "mt" and those in *Arabidopsis* chloroplast genome (NC_001284) by "chl". [d] Introns in untranslated regions are shown for *Arabidopsis* in parentheses. [e] Prediction of subcellular location based on the algorithms Predotar (Small et al. 2004), TargetP (Emmanuelsson et al. 2000) and PSort (Bannai et al. 2002) with + symbols for mt-predicted (or by "chl" and "other" if not). [f] Accession numbers given for rice cDNA entries (in parentheses) when protein databank entries unavailable. [g] Deduced length of rice S1 protein based on our experimental data (T. Hazle and L. Bonen, unpublished). [h] Absent from human mitochondrial ribosomes (Koc et al. 2001a, 2001b). [i] Duplicated chloroplast-type candidate. [j] Adams et al. 2002b; Mollier et al. 2002. [k] Kubo et al. 2000. [l] Kadowaki et al. 1996. [m] Kubo et al. 1999. [n] Sanchez et al. 1996. [o] Adams et al. 2001. [p] Skinner et al. 2001. [q] Gallois et al. 2001. [r] NP_197518 and NP_850857 through alternative splicing of At5g20180.

FIGURE LEGENDS

Figure 1. Schematic showing proposed times of transfer (or loss) of bacterial-origin mitochondrial ribosomal protein genes from the mitochondrion to the nucleus during eukaryotic evolution. This scenario is based on gene content of present-day mitochondrial genomes such as the protist *Reclinomonas* (Lang et al. 1997) and the liverwort, *Marchantia* (Takemura et al. 1992), and shown specifically for rice and *Arabidopsis*. Parentheses indicate that no "native" mitochondrial-type ribosomal protein gene was found in the nucleus (see text). Categories A-C represent "early" to "late" transfer events. Although the L36 gene is absent from *Reclinomonas* mitochondria, it is present in that of another jakobid protist, *Malawimonas jakobiformis* (Gray et al. 2004 and http://amoebidia.bcm.umontreal.ca/pg-gobase/searches/compilations.php), so placed in category B. Note that in *Arabidopsis* the 5' half of L2 is mitochondrion-encoded and whereas L2-3' is nuclear-located (Adams et al. 2001).

Figure 2. Schematics of candidate bacterial-type mitochondrial ribosomal proteins from *Arabidopsis* (upper) and rice (lower) divided into categories A-D which are based on time of gene transfer to the nucleus. Black bars denote bacterial-origin mitochondrial-type core coding regions and white bars represent acquired non-core coding sequences. The chloroplast-origin and cytosol-type ones are shown by grey and stippled bars, respectively. Triangles in white indicate introns, with their lengths shown as increasing triangle size (<200 nt, 200-500 nt, 500-900 nt, and >900 nt) and grey triangles denote group II mitochondrial introns. Stars are shown for introns which are at identical (black) or no more than 6 bp away from (white) those located in human mitochondrial orthologs. Dotted ovals indicate non-identical positions of introns. The

symbol “mt” indicates genes which are still in the mitochondrion in either *Arabidopsis* and/or rice, and asterisks denote duplicated nuclear copies in *Arabidopsis* and/or rice.

Figure 3. Intron density within ribosomal protein genes which were transferred to the nucleus either early in eukaryotic evolution (category A, shaded bars) or more recently (categories B+C, white bars). The coding regions were divided into core (left side) or non-core (right side) and values are expressed as number of introns/100 codons. The core regions in category A (2900 codons) contain 50 introns and for categories B+C (2780 codons) there are 23 introns. For the non-core coding regions (including the junctions), in category A (2800 codons) and categories B+C (1540 codons), there are 20 and 16 introns, respectively.

Figure 4. Comparison of intron positions in *Arabidopsis* and human mitochondrial-type ribosomal protein gene orthologs for (A) introns at identical sites and (B) introns within 6 nt of each other. Intron positions are shown by black arrows, and grey shaded amino acids are identical in *Arabidopsis* and human. Note that introns are numbered based on coding sequences (and exclude any within untranslated regions). Panel C compares the positions of the rice L9 first three introns with the human counterpart. Note that the rice L9 third intron (oval) is absent in *Arabidopsis* L9 (Figure 2, dotted oval). Panel D shows the positions of the second intron (in lowercase italics) within the two *Arabidopsis* L14 genes (HLL and HLP, Skinner et al. 2001). The CAG repeats in HLP are underlined and alternative splicing for HLP is shown by #2-1 and #2-2. The nonsense mutation (CAA to TAA) in HLL in Arabidopsis HUELLENLOS mutants (Skinner et al. 2001) is shown in black. Abbreviations are: At (*Arabidopsis thaliana*), Os (*Oryza sativa*) and Hs (*Homo sapiens*).

Figure 5. Location of nuclear genes for bacterial-origin mitochondrial-type ribosomal proteins on the five Arabidopsis chromosomes (I: 30 Mbp; II:19 Mbp; III: 23 Mbp, IV: 18 Mbp, and V: 26 Mbp; from NCBI Plant Genomes site). Genes on the + strand are shown on the left side and genes encoded on the - strand at right.

Figure 1. (Bonen & Calixte)

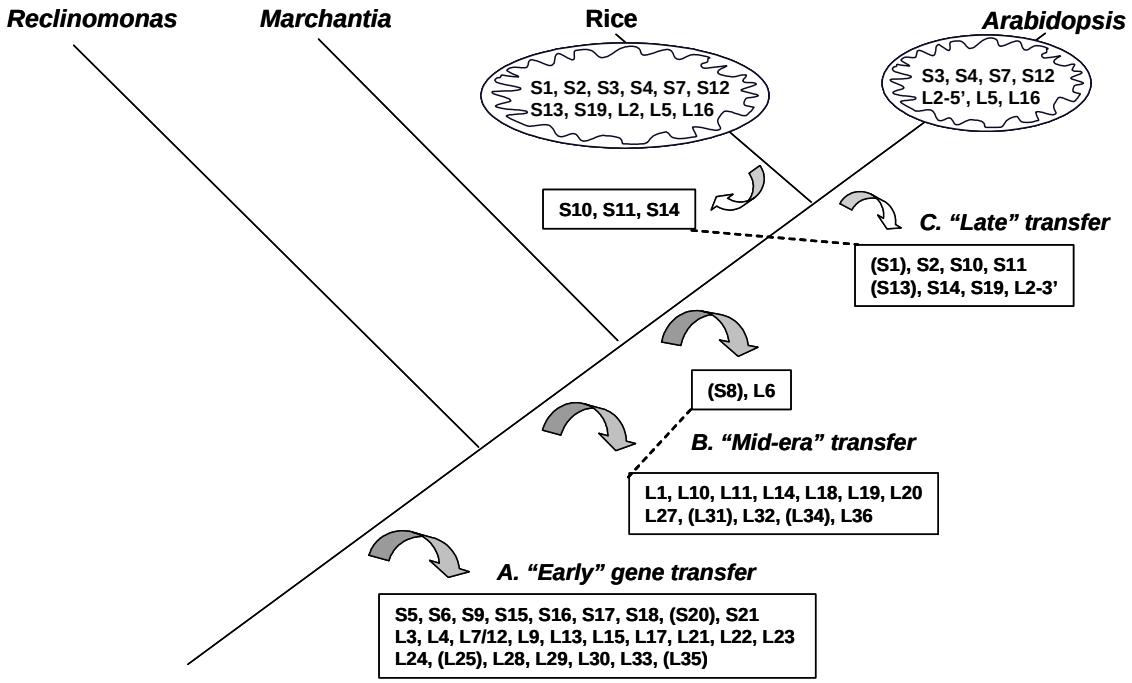

Figure 2. (Bonen & Calixte)

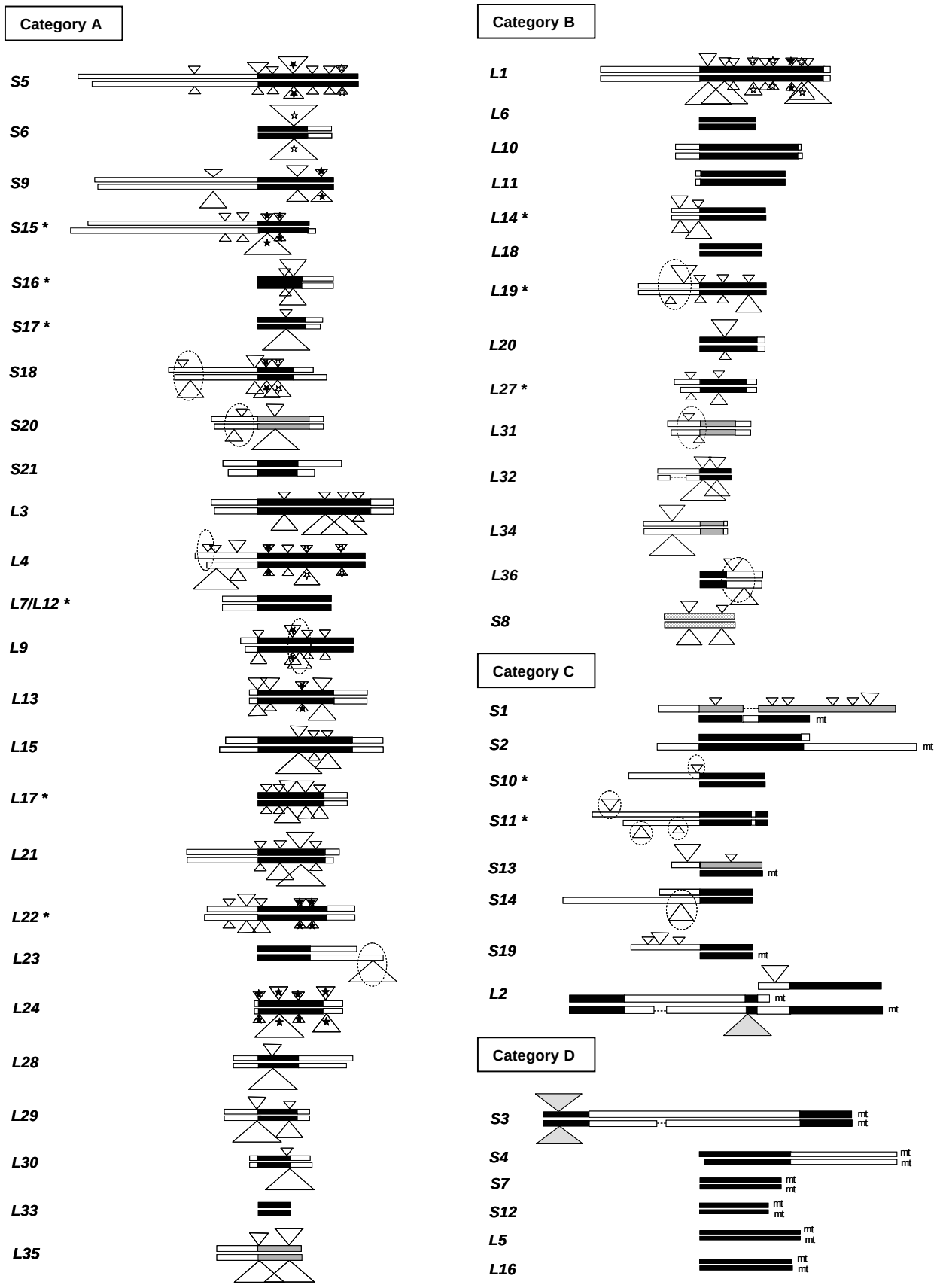

Figure 3. (Bonen & Calixte)

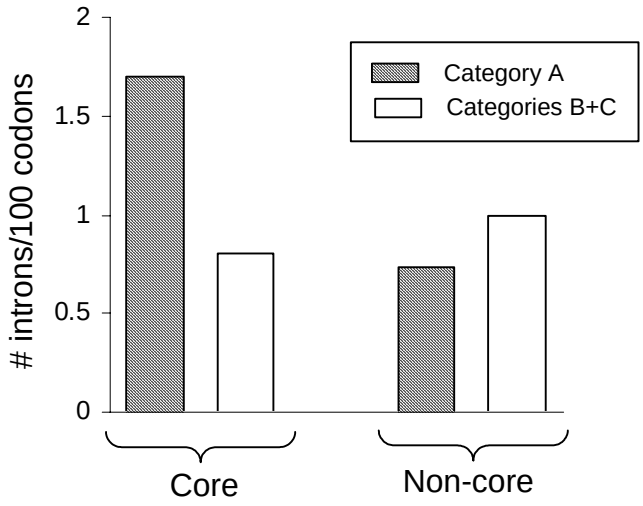

Figure 4. (Bonen & Calixte)

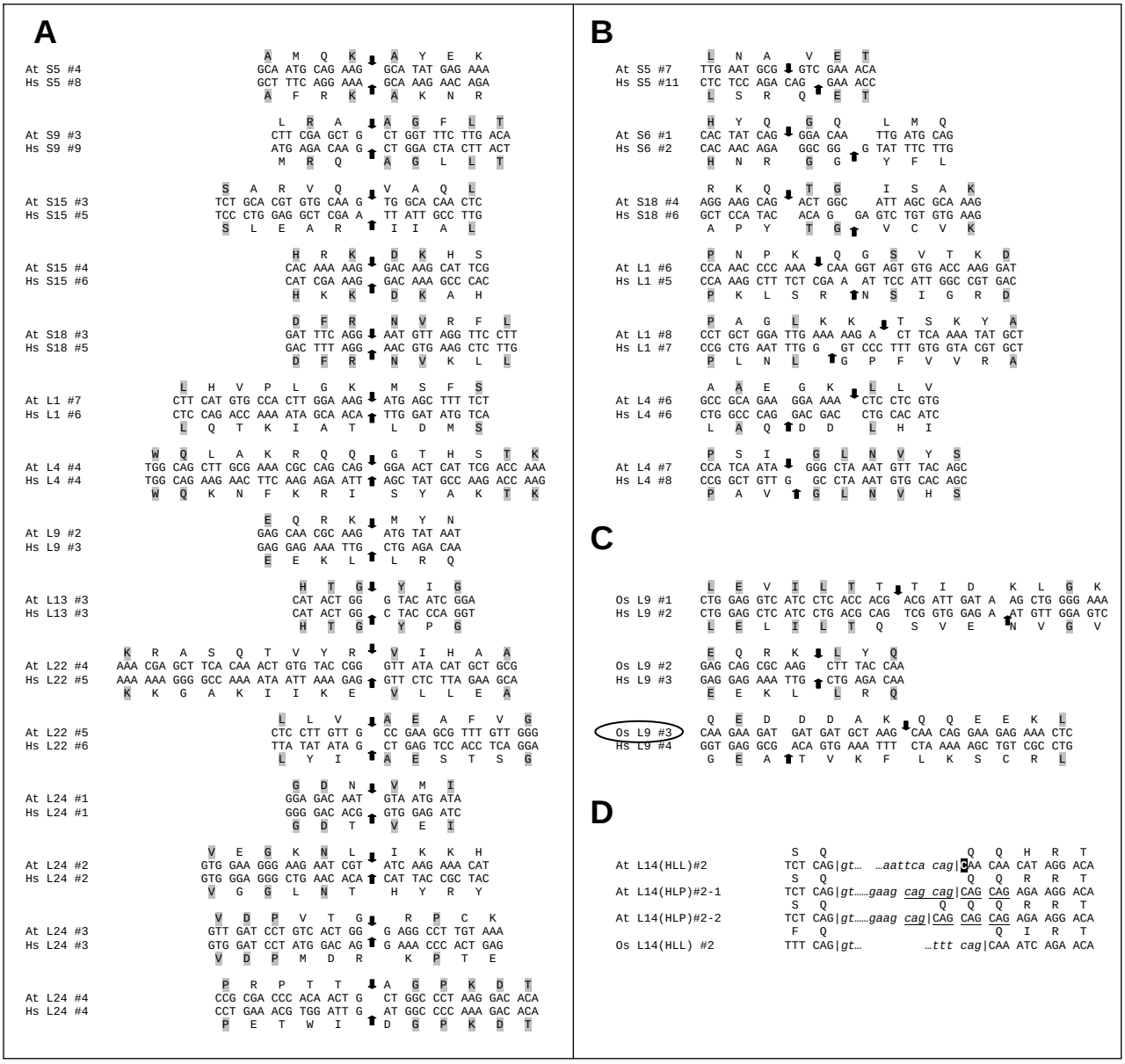

Figure 5. (Bonen & Calixte)