Review Article

Cost of Having the Largest Mitochondrial Genome: Evolutionary Mechanism of Plant Mitochondrial Genome

Kazuyoshi Kitazaki and Tomohiko Kubo

Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Kita-Ku, Sapporo 060-8589, Japan

Correspondence should be addressed to Tomohiko Kubo, tomohiko@abs.agr.hokudai.ac.jp

Received 27 January 2010; Accepted 8 March 2010

Academic Editor: Johann Greilhuber

Copyright © 2010 K. Kitazaki and T. Kubo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The angiosperm mitochondrial genome is the largest and least gene-dense among the eukaryotes, because its intergenic regions are expanded. There seems to be no functional constraint on the size of the intergenic regions; angiosperms maintain the large mitochondrial genome size by a currently unknown mechanism. After a brief description of the angiosperm mitochondrial genome, this review focuses on our current knowledge of the mechanisms that control the maintenance and alteration of the genome. In both processes, the control of homologous recombination is crucial in terms of site and frequency. The copy numbers of various types of mitochondrial DNA molecules may also be controlled, especially during transmission of the mitochondrial genome from one generation to the next. An important characteristic of angiosperm mitochondria is that they contain polypeptides that are translated from open reading frames created as byproducts of genome alteration and that are generally nonfunctional. Such polypeptides have potential to evolve into functional ones responsible for mitochondrially encoded traits such as cytoplasmic male sterility or may be remnants of the former functional polypeptides.

1. Introduction

The monophyletic origin of mitochondria, which postulates a single endosymbiotic event involving an α-proteobacteria-like organism and the common cellular ancestor of eukaryotes, remains a widely accepted concept [1]. Yet it has also become evident that the mitochondrial genomes of extant eukaryotes are more diverged than previously thought [2, 3]. The angiosperm mitochondrial genome is the largest one known to date [2, 3], but this is not the only peculiarity that makes it fascinating. According to a comprehensive review of mitochondrial biology [4], the plant mitochondrial genome “has many characteristics that probably make it one of the most interesting genomes to the molecular biologist.” These characteristics include its mode of gene expression and its organizational diversity. From the botanic point of view, we would like to add that the study of the mitochondrial genome provides us with a huge amount of information regarding evolution, nuclear-cytoplasmic interactions, and mitochondrially-encoded traits, and this information is invaluable for crop improvement.

Since the early attempts to investigate the organization and diversity of the angiosperm mitochondrial genome by electron microscopy and gel electrophoresis in the 1980s, a great deal of knowledge has been accumulated. These advances were described in our recent reviews [5, 6]. Therefore, our focus for this review is the recent activities aimed at identifying the mechanisms that maintain mitochondrial genome organization and that generate alterations, some of which associated with specific phenotypes.

2. General Features of the Angiosperm Mitochondrial Genome

The number of angiosperm species whose entire mitochondrial genome sequences are available is currently twelve, and the sequences can be found at the NCBI web site (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=33090&opt=organelle). In addition, some intraspecific variants of these genomes have been deposited in the public database. These include the NA-type cytoplasm
Two DNA species are excluded from the master circle; one includes the plasmid-like DNA molecules of either linear or circular form, which are independent of the master circle in terms of their replication and mode of inheritance [10]. The other group consists of the substoichiometric DNA molecules that exhibit different sequence from the major DNA molecules, the so-called sublimons. The sublimon is clearly distinguished from the major DNA molecule in terms of its maintenance and gene expression. The sublimons may be products of illegitimate recombination or maintained by an unknown mechanism. We will detail more about the sublimons in Section 4.

There are five general features that are characteristic of angiosperm mitochondrial genomes. First, the mitochondrial genome consists of a mixture of DNA species. This was first realized as a result of electron microscopic- and gel electrophoresis studies of mtDNA. In addition, as we mentioned above, the maps obtained by chromosome walking were very complex. To explain these observations, it was proposed that the master circle and its derivative circles coexist in the mitochondrion. The mechanism for the generation of subcircles was first proposed by Palmer and Shields [11]. The subcircles are generated via frequent homologous recombination within large (more than 1 kb) direct repeats in the master circle (i.e., the smaller circles “loop out” from the master circle). Most of the angiosperm mitochondrial genomes contain one or more sets of such direct repeats that are active in recombination. An exception is documented for black mustard (Brassica hirta) [12]. Also, details of repeated sequences were not mentioned in the first sequence report for the grapevine mitochondrial genome [13], and we failed to find repeated sequences longer than 1 kb in the DDBJ/EMBL/GenBank sequence entry FM179380 (our unpublished observation). Besides direct repeats, some mitochondrial genomes also contain repeats arranged in inverted orientations. In this case, smaller subcircles are not generated, but isomeric circles that differ from the parental circle by segmental inversion can appear. Due to the presence of these subcircles and isomeric circles, the angiosperm mitochondrial genome is considered to have a multipartite organization. It should be noted that the term multipartite organization is different from heteroplasmy because the isomeric circles and subcircles are the alternative form of the master circle. On the other hand, heteroplasmy in angiosperm mitochondria is caused by sublimons, which are present at low stoichiometry and not entirely colinear with the master circle or any other subcircles.

The molecular basis of homologous recombination in angiosperm mitochondria was mainly determined using DNA cloning and/or DNA gel blot hybridization. For example, nine different genomic products were expected if the three-member repeated sequences on the master circle recombined in petunia (3^2 = 9) [14]. When the repeated sequence was used as a probe, the expected nine signal bands appeared on a DNA gel blot containing petunia mtDNA. The relative signal intensities corresponding to the recombinant DNA molecules were comparable to those of the parental DNA molecule, indicating that the stoichiometry of the parental and recombinant molecules was similar.

(DDBJ/EMBL/GenBank accession number DQ490952), the T (Texas)-type cytoplasm (DQ490953), the S (USDA)-type cytoplasm (DQ490951), and the C (Charrua)-type cytoplasm (DQ645536) of maize (Zea mays ssp. mays), and the Owen-type cytoplasm (BA000024) of sugar beet (Beta vulgaris). There is no doubt that the number of available mitochondrial sequences will increase rapidly due to improvements in DNA sequencing technology, such as next-generation sequencing.

The “genome size” of the angiosperm mitochondrion refers to the size of the “master circle,” which is a presumptive circular molecule consisting of all the DNA sequences present at substantial stoichiometry in the mitochondrion. Master circles have usually been constructed by chromosome walking, and fragmented mitochondrial (mt) DNA sequences are linked together based on sequence homologies at the ends of the fragments. However, the resultant maps are not single circles but complex, entangled structures. To help the reader understand this, a simplified map is illustrated in Figure 1, where two closed circles form a figure 8-like structure. This is not due to a cloning artifact such as a chimeric clone. It occurs because the two mtDNA fragments are partly colinear but have diverged from one another. The maps were interpreted as equivalent to the master circles having large duplications, which sometimes reach up to 100 kb. Therefore, the net sequence complexity may be smaller than the size of the master circle. For example, the sizes of the five variant master circles of maize range from 536 to 740 kb, but after omitting the repeated sequences, the complexity ranges from 507 to 537 kb [7]. Although the entire nucleotide sequence remains unavailable, the mitochondrial genome of muskmelon is estimated to be 2400 kb [8], which is about ten times larger than that of rapeseed (222 kb). It should be noted that the presence of the master circle is supported only by molecular cloning, and controversial data have been obtained from other methods including direct visualization of mtDNA molecules and gel electrophoresis of intact mtDNAs [9].

Two closed circles share sequence at the region shown by the arrow. The arrows indicate the duplicated sequences introduced into the master circle. (a) Map configuration derived from chromosome walking. Two closed circles share sequence at the region shown by the arrow. (b) Dissolution of the figure 8-like map to create the master circle.
Such experiments indicate that active recombination occurs within mitochondrial repeated sequences, and recently, biochemical evidence of DNA recombination was obtained [15].

The second feature of the angiosperm mitochondrial genome is that it contains sequences not descended from the ancestral mitochondrial genome but apparently acquired from other genetic compartments, such as plastids, the nucleus, and sexually incompatible organisms (i.e., horizontal transfer). The acquisition of such alien sequences has not been discovered in the mitochondrial genomes of other organisms, apart from those of gymnosperms, which contain plastid-like sequences [16].

The third and fourth features are seen in the gene expression system. The third feature is that most angiosperm mitochondrial genes require the posttranscriptional conversion of specific cytidine residues to uridines, in a process called RNA editing [17]. This C-to-U editing is prevalent, while U-to-C editing is very rare. The fourth feature is the presence of group II introns in up to 10 angiosperm mitochondrial genes [18]. The only group I intron is found in the cox1 gene of diverse plant species and was horizontally transferred from fungi. Among the genes containing group II introns, nad1, nad2, and nad5 exhibit trans-splicing, where the exons are transcribed as separate transcripts and then spliced into one mature transcript.

The most striking feature of the angiosperm mitochondrial genome is that there are fewer genes than expected based on the genome size, indicating that it is the least gene-dense mitochondrial genome [19]. Despite the wide range of genome sizes (222 to 773 kb among the fully sequenced genomes), the numbers of genes in angiosperm mitochondria (about 50 to 60) do not vary greatly. Angiosperm mitochondria contain more genes than those of humans (37 genes in a 16.6 kb genome) but less than those of the liverwort Marchantia polymorpha (71 genes in 186 kb) and the moss Physcomitrella patens (67 genes in 105 kb) [20]. The mitochondrial genome of the gymnosperm Cycas taitungensis (415 kb) contains 64 genes [21], which is slightly higher than the highest number found in angiosperm mitochondria. An analysis of land plants indicates that some genes have been lost from the angiosperm mitochondria, but the genome size has not been reduced accordingly. In some cases the lost genes appear to have migrated into the nuclear genome, and their translation products are imported into the mitochondria [22]. In other cases, the gene loss is compensated by paralogues in the nuclear genome. For example, the gene rps13 is not present in the Arabidopsis mitochondrial and no migrated copy occurs in the nuclear genome. However, a nuclear rps13 copy of plastid origin appears to encode a mitochondrial RPS13 polypeptide [23]. The genes for trNA^{Leu} and trNA^{Arg} are present in Cycas mitochondria [21] but lost from those of angiosperms, and results from an analysis of potato indicate that tRNA molecules charged with Leu and Arg are imported from the cytosol into the mitochondria [24].

It should be noted that there may be additional genes in some angiosperm mitochondria that have not yet been identified. For example, a ribosomal protein gene rpl10 was recently added to the list of mitochondrial genes in tobacco, grapevine, and papaya but is not present in the mitochondria of Arabidopsis, wheat, maize, and sugar beet [25, 26].

3. What Is the Origin of the Intergenic Regions of the Angiosperm Mitochondrial Genome?

Sequence information has provided insights into the causes of genomic expansion in angiosperm mitochondria. As mentioned above, the intergenic regions are apparently expanded compared with those of other organisms. It appears that the intergenic regions originated via at least two mechanisms. One is the transfer of sequences from other cellular compartments, and even from other organisms, including sexually incompatible species. The other is the alteration of preexisting sequences. Both of the mechanisms are likely involved in the evolution of angiosperm mitochondrial genome.

Plastid-like sequences constitute 1.6% to 8.8% of seed plant mitochondrial genomes [6, 13]. Most of these sequences reside in the intergenic regions; however, it has been reported that an internal segment of a mitochondrial atp1 gene has been replaced with a plastid atpA sequence [27]. One of the roles of the plastid-like sequences is to provide tRNA genes for mitochondrial translation [28]. Plastid-like sequences can also function as promoters for mitochondrial genes, as exemplified in the rice nad9 gene [29]. Nuclear-derived sequences are also found in the intergenic regions of angiosperm mitochondria and constitute 0.1% to 13.4% of the mitochondrial genome [6]. These nuclear-like sequences are rarely conserved among angiosperm species, and no role for them has been reported to date.

It should be noted that the amounts of DNA that are known to be derived from external sources such as plastids and the nucleus are insufficient to explain the origin of the whole intergenic regions, since more than 60% of the intergenic regions show no homology with any other sequences. This does not mean that the remaining intergenic regions are unique, because the public database does not yet cover all the extant sequences on Earth. In a recent metagenomic analysis of microbial populations, large amounts of novel sequences derived from unidentified organisms were found [30]. Thus, we cannot exclude the possibility that additional sequences, homologous to those in angiosperm mitochondrial intergenic regions, might be found in other organisms. In other words, it is possible that some of the intergenic regions of angiosperm mitochondria may have been horizontally transferred from sexually incompatible species including nonplant organisms. Evidence of horizontal transfer into angiosperm mitochondria is accumulating. For example, viral sequences and a group I intron presumably transferred from fungi have been found [13, 31, 32]. Additionally, the plasmid-like DNA elements in some angiosperm mitochondria may be derived from fungi [10]. The integration of these plasmid-like DNA elements into mitochondrial genomes has often been observed [7, 33, 34].
Another source of horizontal transfer is from one plant to another. Some research groups claim that mitochondrial genes in some plants, such as the *Amborella nad5* and *Actinidia rps11* genes, seem to have been transferred between plant species horizontally, based on phylogenetic anomalies [35]. These reports remind us of transtaxonomic homologies between mitochondrial open reading frames (ORFs) in different plant species. For example, *orf79* in rice, which is associated with a mitochondrial mutation termed cytoplasmic male sterility (CMS, see below), has a sequence in the 3′ region that shows no homology with normal rice mitochondria or with those of other cereals including maize and wheat. However, in sorghum, another CMS-associated ORF termed *orf107* shares sequence homology with the 3′ half of *orf79* [36].

Three questions arise when horizontal transfer is considered. The first relates to the identity of the donor organism, and a useful tool for tackling this question is phylogenetic analysis [35]. The second question relates to the mechanisms by which DNA fragments are transferred, and the third question refers to the frequency of horizontal transfer. There is some controversy over the frequency of its occurrence in the evolution of angiosperm mitochondrial genomes [13, 32]. This is an important question to be resolved before we can estimate the amount of intergenic sequence that can be explained by horizontal transfer.

It is possible that the intergenic regions of angiosperm mitochondria have accumulated extensive sequence alterations due to duplications, inversions, insertions, and deletions, to the point that the original sources of some sequences cannot be identified by homology searches. Homologous recombination has occurred quite frequently during the evolution of angiosperm mitochondria. When the intergenic regions are investigated in detail, short (less than 2.5 kb) stretches of mitochondrial genes or other known sequences are occasionally found [37]. These might be remnants of extensive rearrangements. The mitochondrial genome of cucumber may be a good example of genome expansion, because it has the second largest mitochondrial genome identified to date (>1500 kb [38], of which 136 kb has been sequenced [39]). An analysis of the obtained sequences revealed that 15% of the region consists of repeat sequences that are either shorter or longer than 557 bp are reported as inactive [47]. The two largest repeat sequences (4.3 and 6.6 kb) are active in recombination while those that are shorter than 557 bp are reported as inactive [47].

If recombination occurred freely in mitochondria, it would result in the impairment of genomic function. Indeed, mitochondrial mutants exhibiting respiratory impairments have occurred as a result of deletions accompanied by losses of mitochondrial genes, through homologous recombination at short repeat sequences. Examples include the maize nonchromosomal stripes mutant, the cucumber mosaic mutant, tobacco CMS I and CMS II, and the *Arabidopsis* maternally distorted leaves mutant [48–51]. In order to preserve mitochondrial functioning there must be a mechanism to suppress free recombination and/or the amplification of irregularly recombined DNA molecules. Nuclear genes whose lesions result in increased amounts of irregularly recombined mitochondrial DNA molecules have been reported. These include *Msh1*, *RecA3*, and *OSB1* [52–54]. The *Arabidopsis* mutant lacking *Msh1* function has been investigated in detail [46]. In these plants there is an increase in recombination via short repeat sequences ranging from 108 to 556 bp, but shorter repeats of less than 108 bp do not respond. This suggests that another gene may suppress recombination via the very short repeat sequences. Consistent with this, the disruption of *RecA1*
in Physcomitrella results in the activation of mitochondrial genome recombinations involving repeats of less than 100 bp [55]. In the presence of functional Msh1, recombination via short repeats is apparently suppressed, but recombinant DNA molecules are not completely absent; they can be present at very low stoichiometric levels as sublimons. It is evident that Msh1 also plays a similar role in tobacco and tomato, because knockdown experiments involving the Msh1 orthologues in these plants result in novel mitochondrial genotypes [56]. Although the genes responsible have not been identified, plant lines that generate multiple mitochondrial genotypes [56]. In the presence of functional Msh1, recombinant DNA molecules involved in mitochondrial diversity and heteroplasmy. A long-standing hypothesis, that an alteration of mtDNA is first carried as a sublimon and then amplified to the level of a major DNA molecule [58, 59], is still attractive and remains to be explored. In the common bean, CMS-type mtDNA lurks as a sublimon in the ancestral, fertile strain [60]. In Arabidopsis, the major mtDNAs of the three ecotypes C24, Col-0, and Ler are organizationally different from each other; however each ecotype carries the two noncognate mtDNAs as sublimons [46, 61]. Therefore, recombinant DNA molecules involved in genomic diversity can occur as sublimons. Because sublimons are apparently transcriptionally inactive [62, 63], mutations may easily accumulate in them.

In order to accumulate genomic alterations in sublimons, the heteroplasmic state should be transmitted from one generation to the next. However, in animal mitochondria, heteroplasmy is resolved into homoplasm during oogenesis [64]. There is some controversy regarding this process. It is not clear whether the copy number of the mtDNA is decreased to exert a bottleneck effect resulting in homoplasy, or whether the copy number is not reduced but the number of effective mtDNA segregation units is small [64]. However, there has been no such study in angiosperms.

How sublimons are maintained in angiosperm mitochondria is a very important question [65]. In the common bean, sublimons are classified into two classes based on quantitative data [66]. One class is present at a ratio of approximately 10^{-3} : 1 relative to the major DNA molecule and can be detected rather easily by DNA gel blot hybridization or PCR. The other class occurs at a ratio of approximately 10^{-6} : 1, and highly sensitive methods are needed for detection of these molecules. The sublimons may be generated via occasional recombination at short repeat sequences, and/or they may replicate inefficiently, so that the copy number is reduced compared with that of the major DNA molecule. The latter possibility assumes an additional mechanism to ensure the transmission of the sublimon to the next generation; however the existence of such a mechanism remains obscure. It should be noted that paternal leakage, that is, mitochondrial transmission via pollen, has been suggested as one of the sources of sublimons [67, 68]. Most angiosperm species show exclusively maternal inheritance of the mitochondrial genome. However, paternal leakage has been documented in several cases [67] and may contribute to heteroplasmy [68].

The altered sublimon needs to be amplified to the level of the major mtDNA molecule in order to manifest as a genomic alteration. Such increases in the copy numbers of sublimons can occur as a result of mutations of nuclear genes or physiological stress. This phenomenon, together with a rapid decrease in the copy number of the mtDNA molecule to a substoichiometric level, is called substoichiometric shifting (SSS). SSS was first recognized in common bean, where the introduction of the nuclear gene Fr into a plant by crossing resulted in a decrease in a specific class of mitochondrial DNA molecules in the subsequent generation [62]. The Msh1 gene (mentioned above) was first identified as a mutant allele causing SSS in Arabidopsis [52], suggesting that the genes that suppress homologous recombination may also help to regulate the copy numbers of recombinant DNA molecules.

The SSS-like phenomenon can also be observed in the absence of any mutations. A well-known example is tissue culture, which can cause increases or decreases in the copy numbers of specific mtDNA molecules [69]. This suggests that SSS may occur as a result of developmental triggers. The tissues or organs where SSS might occur are the meristem and the egg, where the “transmitted form” of the mtDNA, containing all components of the genome including sublimons, is expected [60, 70].

### 5. Why Do Most of the Unique ORFs Not Matter in Angiosperm Mitochondria?

Two classes of mitochondrial mutants are known in angiosperms. One is associated with lesions in indispensable mitochondrial genes. These can be caused by deletions or (rarely) point mutations [71, 72]. In the other class, no deleterious mutation can be found in any of the genes; nevertheless a specific mutant polypeptide is translated in the mitochondria. When the ORF encoding the polypeptide is cloned, it is structurally distinct from the standard mitochondrial gene and consists of segments of known or unknown sequences. Therefore, in this class of mitochondrial mutants, a unique ORF could be associated with the phenotype. CMS, which shows a phenotype of pollen abortion in plants that are otherwise developmentally normal, is one representative of this class of mutations. CMS has been reported in more than 140 plant species [36], and the number of identified genes associated with the phenotype has reached 23, of which 21 involve unique ORFs [73]. An example is the maize urf13-T gene, which contains two duplicated segments of rnr26, forming an ORF that encodes 115 amino acid residues [74]. A second example is the petunia pcf gene, which contains a duplicated segment of atp9, two duplicated segments of cox2, and a sequence of unknown origin, forming an ORF that encodes 402 amino acid residues [75]. There is no sequence homology between any of the known CMS-associated ORFs except in two cases: orf224, associated with Brassica pol
CMS, and orf222, associated with *Brassica nap* CMS, exhibit 79% homology at the amino acid level [36]; and rice orf79 and sorghum orf107 show some similarity. Mitochondrially encoded disease susceptibility is another phenotype that is associated with unique ORFs: the maize orf13-T causes not only CMS but also susceptibility to a toxin produced by the pathogen *Cochliobolus heterostrophus* [36]. The rough lemon gene *Acrs* (containing a plastid-like sequence) is responsible for sensitivity to a toxin produced by the fungus *Alternaria alternata* [76]. Because the orf13-T sequence can be deleted by tissue culture to obtain male fertile and toxin-insensitive plants, the orf13-T gene is dispensable for normal mitochondrial function [36].

In addition to these unique ORFs that cause mutant phenotypes, many more unique ORFs, which cause no obvious phenotypes, are present in angiosperm mitochondrial genomes. For example, *Arabidopsis* mitochondria contain unique ORFs with duplicated segments derived from known mitochondrial genes and sequences of unknown origin [77], but no mitochondrially encoded phenotypes such as male sterility or disease susceptibility are known in this ecotype. Therefore, most unique ORFs seem to cause no problems for the plants. The large coding capacity of the angiosperm mitochondrial genome and its structural fluidity appear to have resulted in the frequent emergence and retention of unique ORFs during the course of plant evolution. Why are so many unique ORFs tolerated? To address this question, we would like to discuss the features that distinguish the CMS-associated ORFs from other unique ORFs.

As reviewed by Holec et al. [78], the multiplicity of promoter sequences and the absence of efficient transcription termination mechanisms lead to the transcription not only of mitochondrial genes but also of intergenic regions. Thus, transcription may occur anywhere in the mitochondrial genome, including regions containing unique ORFs. However, most illegitimate transcripts are degraded by a mechanism involving a 3′ to 5′ exonuclease called mitochondrial polynucleotide phosphorylase, because they are not associated with stabilizing factors that regulate the abundance of mitochondrial RNA. Cotranscription with functional mitochondrial genes is one way for a unique ORF to be transcribed and to accumulate as a stable RNA, because the stabilizing factors for the mitochondrial gene may incidentally protect the adjacent unique ORF. In fact, many CMS associated genes are cotranscribed with downstream or upstream mitochondrial genes [79]. The significance of this arrangement is evident from the observation that the restoration of fertility often accompanies the cleaving off of the unique ORF-coding sequence from the polycistronic transcript (see an example for rice in [80]). On the other hand, it should be noted that cotranscription does not always result in the translation of the unique ORF. Two ORFs in sugar beet mitochondrial, orf119 (consisting of a duplicated segment of *nad9* and 340 bp of unknown origin) and orf324 (containing a duplicated segment of *atp8* and 751 bp of unknown origin), are transcribed along with the downstream genes *atp1* and *rps13*, respectively. However, the ORF119 and ORF324 antisera failed to react with any mitochondrial proteins, and no band corresponding to such polypeptides was observed among the radiolabelled *in organello* translation products [81–83]. A possible explanation is that the unique transcripts lack cis regulatory elements required for translation. However, sugar beet orf324, like the CMS-associated orf522 of sunflower and orf222 and orf222 of *Brassica*, contains a duplicated segment of *atp8* in its 5′ region. In sugar beet, *atp8* and orf324 share 321 bp of 5′ upstream sequence and 59 bp of sequence that encodes the amino terminal part of the protein (Figure 2(a)). It is possible that orf324 is translated, but the translation products are not detected due to protein degradation.

The translation of a unique ORF is inevitable if it fused in frame with a genuine mitochondrial gene. This situation has occasionally occurred during the course of evolution. Some examples of such unique ORFs, whose translation products have been analyzed, are listed in Table 1. In each case, the origin of the fused sequence is not completely clear. It is intriguing that both the N-terminal and C-terminal regions of mitochondrial genes can be targets of genomic alteration, but not all result in mitochondrial impairment. Of the nine genes listed in Table 1, only two are associated with CMS. In the cases where the fused polypeptide is cleaved (7 genes), the peptide associated with the unique ORF could be detected in two cases. The mechanism by which the core region is cleaved off is not fully understood, because we currently lack sufficient knowledge of posttranslational process in angiosperm mitochondria.

A well-known example of this situation is *atp6*, which consists of a conserved C-terminal region of 249 to 252 amino acid residues (core region) and a diverged N-terminal region of 5 to 389 residues, depending on the mitotype (the N-terminal extension) (Figure 2(b)). Because the standard translation initiation codon (Met) is at the amino-terminus of the extension and the core region starts with a Ser residue, translation of the core region must follow the translation of the N-terminal extension. From the fused polypeptide, the core region is yielded as a mature *ATP6* polypeptide after cleavage of the N-terminal extension at a site just before the consensus motif Ser-Pro-Leu. This motif is the target site of an endopeptidase, presumably the orthologue of the yeast ATP23 [91, 92] (see below). There is no evidence that the cleaved N-terminal extension is accumulated as a solitary polypeptide, apart from one exception: sugar beet *atp6*, whose N-terminal extension is associated with CMS and known as preSatp6 [83]. In cereal mitochondria, the gene *rps2* encodes a fused C-terminal extension [93] that is not encoded in the *Marchantia, Physcomitrella*, or *Cycas* mitochondria. The fused *rps2* of maize is translated, then the C-terminal extension is cleaved off, but it can be detected as a solitary polypeptide [86]. Sugar beet has a unique *ccmC* gene in which the conserved translational initiation codon is lost, but instead, the chimeric N-terminal extension consists of a duplicated segment of *atp9* and a sequence of unknown origin [90] (Figure 2(b)). The mature CCMC polypeptide is yielded from the fused polypeptide by cleavage of the N-terminal extension [90]. It has been suggested that the putative cleavage site may precede Ser-Pro-Leu, by analogy with *atp6*. The solitary N-terminal extension has not been detected. To date, no phenotypes...
Figure 2: Organizational relationships between genuine mitochondrial genes and unique ORFs at the DNA (a) and polypeptide (b) levels. (a) Comparison of sugar beet atp8 and orf324. Open reading frames are boxed. Dashed lines indicate the boundaries of shared sequences. Two in-frame initiation codons (atg) are shown, of which the second is conserved among angiosperm atp8 genes. Sugar beet orf324 contains a duplicated segment of atp8 and a sequence of unknown origin. (b) Comparison of polypeptides deduced from the sugar beet atp6 and the sugar beet ccmC. The N-terminal extensions and the core regions are indicated. The putative cleavage motif is Ser-Pro-Leu. The N-terminal extension of sugar beet ccmC consists of a duplicated segment of atp9 and a sequence of unknown origin.

have been associated with the maize rps2 or the sugar beet ccmC.

It is not known whether there are any roles for these extensions in angiosperm mitochondria, or whether the fusion proteins have acquired novel functions, apart from the exception of the sugar beet atp6 [83]. No extensions have been found in mammalian atp6 genes; however an N-terminal extension (10 amino acid residues) of atp6 is required for the efficient assembly of the mature ATP6 polypeptide into the ATP9 ring in Saccharomyces cerevisiae [94]. In this process, a dual-function metalloprotease, ATP23, cleaves the N-terminal extension off and, together with the putative molecular chaperon ATP10, assembles the mature ATP6 into the F0 ATPase [91, 92]. It should be noted that potential homologues of atp23 and atp10 are also encoded in the Arabidopsis nuclear genome as at3g03420 and at1g08220, respectively (our unpublished observation).

It is possible that the translational products of unique ORFs, if any, are subjected to degradation by mitochondrial quality control systems such as AAA- and Lon proteases. These proteases are indeed encoded in Arabidopsis and have been shown to participate in protein degradation [95–97]. The mitochondrial protein degradation system is also associated with the anther-specific accumulation of the CMS-associated ORF239 polypeptide in the common bean [98]. The petunia CMS-associated PCF protein is trimmed into a smaller polypeptide from a precursor [75]. Therefore, it seems likely that most unique polypeptides are degraded. However, some of the cleaved polypeptides derived from unique ORFs (e.g., CMS-associated ORFs and the maize rps2A extension) might be tough targets for the protein degradation system due to their higher-order structures. For example, if a polypeptide were embedded in the mitochondrial membrane and its extracellular domain were inaccessible to molecular chaperones, this would prevent it from being pulled out of the membrane for degradation [99]. In rice, ORF79 is a stable membrane-integrated polypeptide expressed in all tissues. In a line of CMS rice carrying a variant of orf79, termed L-orf79, the L-ORF79 polypeptide is not detected [100]. Itabashi et al. [100] suggested that this might be due to a sequence alteration in the 5′ untranslated region that prevents translation. However, it is also possible that a Glu to Ala substitution in the coding region may alter the higher-order structure, making L-ORF79 accessible to the degradation system. A variant of atp6 in sugar beet termed preS-3atp6, whose encoded amino acid sequence shows 88% homology with that of preSatp6, is translated and the corresponding polypeptide is detected in leaves and roots but not anthers. This suggests that the polypeptide may be accessible to a tissue-specific degradation system, unlike preSATP6, which is quite stable [83, 101].

Overall, unique ORFs do not create problems for plants as long as their translation products are controlled by posttranscriptional and posttranslational processes. In addition, it appears that some nontoxic or nonfunctional polypeptides derived from unique ORFs (e.g., see Table 1) can exist in angiosperm mitochondria. This suggests that the accumulation of unique polypeptides per se is insufficient to cause mitochondrially encoded phenotypes. In other words, the CMS- or disease susceptibility-associated polypeptides may have particular characteristics that enable them to modulate mitochondrial function, leading to specific phenotypes.
Table 1: Unique ORFs that are fused with genuine mitochondrial genes.

<table>
<thead>
<tr>
<th>Name of gene</th>
<th>Position of fused unique ORF</th>
<th>Detection of fused polypeptide</th>
<th>Detection of cleaved core region</th>
<th>Detection of cleaved extension</th>
<th>Note</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrot atp8</td>
<td>C(57)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>[84]</td>
</tr>
<tr>
<td>Carrot cox1</td>
<td>C(224, 284)</td>
<td>No</td>
<td>Yes</td>
<td>ND</td>
<td></td>
<td>[85]</td>
</tr>
<tr>
<td>Maize rps2A</td>
<td>C(330)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td>[86]</td>
</tr>
<tr>
<td>Sugar beet Gnd9-1</td>
<td>C(14)</td>
<td>Yes</td>
<td>No</td>
<td>ND</td>
<td>Complex I exhibits normal activity.</td>
<td>[71]</td>
</tr>
<tr>
<td>Sorghum cox1</td>
<td>C(101)</td>
<td>Yes</td>
<td>No</td>
<td>ND</td>
<td>Presumably associated with CMS.</td>
<td>[87]</td>
</tr>
<tr>
<td>Petunia atp6</td>
<td>N(145)</td>
<td>No</td>
<td>Yes</td>
<td>ND</td>
<td></td>
<td>[88]</td>
</tr>
<tr>
<td>Radish Ogura atp6</td>
<td>N(174)</td>
<td>No</td>
<td>Yes</td>
<td>ND</td>
<td></td>
<td>[89]</td>
</tr>
<tr>
<td>Sugar beet atp6</td>
<td>N(387)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Cleaved extension forms oligomer. Presumably associated with CMS.</td>
<td>[83]</td>
</tr>
<tr>
<td>Sugar beet ccmC</td>
<td>N(277)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>[90]</td>
</tr>
</tbody>
</table>

1C: carboxyl terminal; N: amino terminal; numbers in parenthesis indicate the amino acid residues encoded by the fused ORF.
2Polypeptides corresponding to the genuine mitochondrial gene.
3Polypeptides corresponding to the unique ORF.
4The extension may not be removed completely.
5ND: not determined.
6Polypeptides containing both the N-terminal extension and the core region of ATP6 have not been detected in any plant, perhaps due to rapid processing, or for unknown reasons (see references listed in the table).

The identification of such “elite” unique polypeptides is of practical importance. A transgenic approach sometimes (but not always) works well, wherein plants expressing specific unique ORFs fused with mitochondrial import signal peptides express male sterility or toxin sensitivity [79, 101–105]. Microorganisms such as *E. coli* can also be used as hosts to examine the toxin sensitivities conferred by specific ORFs [36, 76]. It is known that the polypeptides encoded by maize *urf13-T*, radish *orf138* (associated with CMS), and sugar beet *preSartp6* form oligomers in the inner membrane [36, 83, 106], which may function as ionophores [107]. However, sugar beet ORF129, which also causes CMS when targeted to the mitochondria, shows a different localization pattern [101]. Thus, there appear to be multiple ways by which unique polypeptides can modulate mitochondrial function.

6. Concluding Remarks and Perspectives

The angiosperm mitochondrial genome is the largest one known to date and is highly diverged in terms of genome size, gene arrangement, and sequences in the intergenic regions. The mechanisms by which angiosperm mitochondrial genomes have diverged may include horizontal transfer from other organisms, although the extent of this contribution remains to be evaluated. The recombination of mtDNA should be under the strict control to maintain genomic integrity. In leaves and other vegetative tissues that have been used for mtDNA analyses, two contradictory activities have been found to coexist. One seems to promote recombination within repeats of over 1 kb, and the other suppresses recombination between repeats of less than 1 kb. The existence of heteroplasmy in angiosperm mitochondria provides us with a unique opportunity for exploration in the field of cytoplasmic genetics, because the situation in plants appears to be different from that in mammals. One of the major questions to be resolved is the maintenance and transmission of sublimons. It might be necessary to examine the mitochondrial genome organization in the meristem and the egg, both of which are difficult to isolate and examine. The extent of paternal leakage of the mitochondrial genome might be underestimated, since this is a likely cause of heteroplasmy and horizontal transfer. Because the functional constraints of the intergenic regions are low, and because the angiosperm mitochondrial genome undergoes frequent rearrangements on an evolutionary scale, unique ORFs are common. Fused mitochondrial genes provide appropriate examples to examine how the expression of unique ORFs is controlled. Posttranslational control mechanisms, including the quality control system, are important for the proper expression of fused mitochondrial genes. It is apparent that useless polypeptides derived from unique ORFs can exist in mitochondria, presumably because they can escape the quality control system. We would like to hypothesize that there exists an evolutionary association between these polypeptides and CMS-associated polypeptides. The apparently nonfunctional polypeptides may have the potential to evolve into causal agents of CMS, or they may be remnants of ancient CMS-causing genes.

Abbreviations

CMS: Cytoplasmic male sterility
ORF: Open reading frame
SSS: Substoichiometric shifting.
Acknowledgments

This work was supported in part by Grants in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology and the Program for Promotion of Basic and Applied Research for Innovation in Bio-oriented Industry (BRAIN).

References


(ccmC) is expressed as a precursor with a long NH$_2$-terminal extension in sugar beet, "Journal of Plant Physiology, vol. 166, no. 7, pp. 775–780, 2009.


Submit your manuscripts at
http://www.hindawi.com