Title: Male sterility-inducing mitochondrial genomes: how do they differ?¹

Authors: Tomohiko Kubo*, Kazuyoshi Kitazaki, Muneyuki Matsunaga, Hiroyo Kagami and Tetsuo Mikami

Affiliation: Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

Referee: Dr. Françoise Budar, INRA, Station de Genetique et d’Amelioration des Plantes, 78026 Versailles, France

* Corresponding author (E-mail: gelab@abs.agr.hokudai.ac.jp)

¹This review is dedicated to the memory of Dr. Toshiro Kinoshita
TABLE OF CONTENTS

I. Introduction
II. Molecular diversity of the angiosperm mitochondrial genome
   A. Genome size and organization
   B. Genes in the genome
   C. Transferred sequences
   D. Origin unknown sequences in the intergenic region
   E. How has angiosperm mitochondrial genome expanded?
   F. How did angiosperm mitochondrial genomes become so varied?
III. The mitochondrial genome and cytoplasmic male sterility (CMS)
   A. Potential sources of CMS
   B. What do mitochondrial genomes of CMS plants tell us?
   C. CMS-associated loci
IV. How is CMS expressed?
V. Nuclear genes that suppress CMS
VI. Concluding remarks and perspectives
VII References
Abstract

Twenty-nine mitochondrial genomes from 19 angiosperm species have been completely sequenced and have been found to vary in genome size and gene content. Seven of these mitochondrial genomes are known to induce cytoplasmic male sterility (CMS), and thus can be utilized for hybrid seed production or the prevention of pollen dispersal. Genome rearrangement frequently is observed in MS-inducing mitochondria, but it also occurs as part of the normal inter- or intraspecific variation in male fertile (MF) mitochondria. Sequence analyses have revealed that the repertoire of genuine genes is indistinguishable between MS-inducing and MF mitochondria. Deleterious mutations appear to be rare in MS-inducing mitochondria, which may be consistent with the lack of systemic manifestation of CMS. On the other hand, several nucleotide substitutions remain to be investigated for their potential mild effects. Various mitochondrial ORFs are associated with CMS (CMS-ORFs). There are some common but not strict features shared by CMS-ORFs such as their uniqueness to the CMS mitochondrial genome, their association with genes for ATPase subunits, and the hydrophobic nature of their putative translation products. It should be noted that some CMS-ORFs do not satisfy all of these criteria, and ORFs that satisfy these criteria are not necessarily associated with CMS. Therefore, it is difficult to infer the capability of MS induction of mitochondrial genomes solely from their nucleotide sequences. Morphological, physiological, and molecular biological studies suggest that multiple mechanisms cause CMS. Nuclear genes that suppress CMS have been identified. Post-transcriptional suppression of CMS-ORFs mediated by a certain class of RNA binding proteins (pentatrico peptide repeat proteins) is the predominant mechanism of fertility restoration. On the other hand, CMS suppression that is not associated with post-transcriptional suppression of CMS-ORFs has also been reported, suggesting that various types of gene-products are involved in fertility restoration.
KEY WORDS: cytoplasmic male sterility, fertility restorer gene, nuclear–mitochondrial interaction, mitochondrial genome, plant mitochondria

I. Introduction

It is widely accepted that mitochondria originated from endosymbionts resembling α-proteobacteria (Gray, 1999; Scheffler, 1999). According to this Endosymbiotic Theory, the mitochondrial genome is derived from the initial endosymbiont genome. However, mitochondrial genomes have drastically diverged in terms of molecular form, genome size, gene content and arrangement, and organization of intergenic regions (Gray et al., 2004; Bullerwell and Gray, 2004). Among the known mitochondrial genomes, the angiosperm mitochondrial genome has been considered unique in terms of its organization and large size (up to ~3000 kbp, see below) ever since the initial studies of the late 1970s (e.g. Quetier and Vedel, 1977). It was not until 1997 that the first nucleotide sequence of an entire angiosperm mitochondrial genome was published (Arabidopsis thaliana; Unseld et al., 1997), whereas the human and bovine mitochondrial genomes had been sequenced entirely in the early 1980s (Anderson et al., 1981; Anderson et al., 1982).

The angiosperm mitochondrial genome attracts interest not only from pure science but also from researchers in the applied sciences including plant breeding. This is because the angiosperm mitochondrial genome can control cytoplasmic male sterility (CMS), an important breeding character for hybrid seed production. In this review, CMS is defined as male sterility (MS) caused by nuclear–mitochondrial interactions without additional defects in the female reproductive organs and vegetative organs (Budar et al., 2006). MS due to a sterilizing factor (S) in the mitochondrial genome can be eliminated by a specific, mostly dominant nuclear gene termed “restorer of fertility” (Rf). In the absence of the sterilizing factor in the mitochondrial genome, male fertility (MF) is not affected by the Rf genotype. A summary of genotypes and their corresponding phenotypes is shown in Figure 1. Because MS plants cannot self-pollinate, their use as a seed parent can yield large amounts of pure hybrid seeds.

CMS is known from over 140 plant species (Laser and Lersten, 1972), but not all of them are useful in practical applications. Therefore novel mitochondrial genotypes
(mitotypes) that induce CMS are sought for in many crops. For example, CMS that is sensitive to environmental conditions is considered unfavorable for hybrid breeding because incomplete MS allows for self-pollination which reduces the purity of hybrid seeds. In *Brassica*, MS induction by nap-type CMS mitochondria is unstable at high temperatures (Fan and Stefansson, 1986), making nap-type CMS impractical. In wheat, CMS mitochondria introduced from wild relatives does not induce MS under a short photoperiod which enables MS line maintenance by selfing. On the other hand, long photoperiods promote MS and the MS line becomes a suitable parent for hybrid seeds (Murai and Tsunewaki, 1993). Combining multiple CMS types reduces risks emerging from the reliance on a single type; the use of maize T-type CMS in hybrid seed production collapsed in the early 1970s due to the pleiotropic expression of disease sensitivity (Pring and Lonsdale, 1989).

As shown in Figure 1, the establishment of a CMS system requires at least two mitotypes, one of which can cause MS while the other has no effect. Therefore, efforts to discover novel CMS include two questions: first, the identification of new mitotypes in a gene pool, and second, an assessment of their potential to induce MS. The former task is not difficult if sufficient genetic resources are available because molecular techniques such as the detection of polymorphisms of restriction fragment lengths of mitochondrial DNA are well established. Furthermore, advances in nucleotide sequencing technology will enable us to obtain entire nucleotide sequences of all mitoypes in a gene pool. On the other hand, the second task appears formidable because the only way to test the capability of inducing MS is in field trials, including backcrossing. In other words, no general sequence marker is available for the easy identification of mitotypes with an S factor. What makes MS-inducing mitotypes different?

In this review, we first focus on recent progress in mitochondrial genome organization and variation, and then discuss CMS and *Rf*. Because of the definition of CMS in this review, we will not address fascinating mitochondrial mutants such as maize non-chromosomal stripes, tobacco CMSI, CMSII, and cucumber mosaic (Gabay-Laughnan and Newton, 2005; Bartoszewski et al., 2007; Vidal et al., 2007). Crop species are the main subject of this review, however, it should be noted that studies on mitochondrial genome diversity and CMS are also conducted in non-crop plants in population genetics and ecology (Tiffin et al., 2001; Murayama et al., 2004; Delph et al., 2007; McCauley and Olson, 2008; McCauley and Bailey, 2009). We will refrain from
discussing topics that were evaluated in the recent past; readers may refer to previous reviews (Schnable and Wise, 1998; Hanson and Bentolila, 2004; Budar et al., 2004; Knoop, 2004; Budar et al., 2006; Chase, 2007; Kubo and Mikami, 2007; Pelletier and Budar, 2007; Kubo and Newton, 2008; Fujii and Toriyama, 2008b; Woloszynska, 2010; Kitazaki and Kubo, 2010). In this review, unless indicated otherwise, mitochondrial genes are given in italicized lower case letters (e.g. \textit{atp6}) while nuclear genes appear in italicized upper case letters (e.g. \textit{AOX}). The only exception is the nuclear gene \textit{Rf}, for historical reasons and because of the necessity to distinguish dominant (\textit{Rf}) and recessive (\textit{rf}) alleles. Polypeptides are given in roman upper case letters, like ATP6, AOX and RF.

II. Molecular diversity of the angiosperm mitochondrial genome

A. Genome size and organization

The molecular form of the angiosperm mitochondrial genome is controversial and the observation that a mixture of linear, circular, and networked DNA molecules constitutes the mitochondrial genome certainly requires consideration (Andre and Walbot, 1995; Bendich, 1993; Oldenburg and Bendich, 1996; Dai et al., 2005; Manchekar et al., 2009). Nonetheless, the entire set of genetic information in the angiosperm mitochondrial genome can be described by a single circular configuration, the so-called master circle (Palmer and Shields, 1984). A total of 29 master circles from 19 angiosperm species have been completely sequenced to date (Table 1). This number will no doubt increase rapidly due to technical advances in nucleotide sequencing. For example, next-generation sequencing technologies, such as the GS-FLX (also known as ‘454’) system, have been applied recently to sequence two rice mitochondrial genomes (Fujii et al., 2010a). However, certain technical difficulties remain unresolved, such as the discrimination of mitochondrial sequences from contaminating plastid or nuclear DNA. The integrity of the obtained contigs should be verified. Moreover, the assembly of short reads also raises difficulties even though reference sequences exist, because of the variation in organization and unique sequences (F. Budar and T. Terachi, personal communication; see below).

The sizes of master circles vary from 208 to 2936 kbp (Palmer and Herbon, 1987;
Ward et al., 1981; Alverson et al., 2010); the latter value represents the largest mitochondrial genome of a eukaryote reported to date (Bartoszewski et al., 2009). As seen in Table 1, the smallest accessible mitochondrial genome sequence is 222 kbp and the largest is 983 kbp. Some master circles contain a large duplication. For example, the NA-type genome of maize has a 120-kbp duplication (Allen et al., 2007). Apart from large duplications that exceed the capacity of cloning vectors such as phages and cosmids, repeat sequences of various lengths are common in angiosperm mitochondrial genomes (Andre et al., 1992). These repeat sequences are usually classified into two classes. One class contains sequences larger than 1 kbp and active in the recombination of the repeat copies. Recombination between direct repeats in the master circle results in two subgenomic circles (loop out). On the other hand, recombination between inverted repeats results in an isomeric form of the master circle that differs in the orientation of the region between the repeats (flip flop). These recombinations may explain the heterogeneity of mitochondrial DNA molecules. Active recombination between repeat copies may be involved in DNA replication and/or repair in plant mitochondria (Marechal and Brisson, 2010).

Repeat sequences exceeding 1 kbp in length appear to be ubiquitous in angiosperm mitochondria; the only exception known for a long time was black mustard (Brassica hirta) (Palmer and Herbon, 1987). Recently, two additional exceptions have been found: grapevine and zucchini (Goremykin et al., 2009; Alverson et al., 2010). Although these two genomes constitute the two largest mitochondrial genomes to be completely sequenced to date, no repeat sequences exceeding 1 kbp have been found. It is not known whether active recombination in these genomes is mediated by sequences of less than 1 kbp, or if any site-specific recombination occurs. It is generally accepted that recombination between the short repeat sequences is infrequent but plays an important role in the evolution of the angiosperm mitochondrial genome (Andre et al., 1992; Arrieta-Montiel et al., 2009).

B. Genes in the genome

The angiosperm mitochondrial genome includes genes for subunits of Complexes I to V of the electron transport chain, ribosomal proteins, genes involved in cytochrome \( c \)
maturation, protein translocation system subunits, maturase, and genes for rRNAs and tRNAs. The arrangement of angiosperm mitochondrial genes is less conservative than in animal mitochondrial or plastid genomes, except for certain local syntenic linkages which are broken in some plant lineages (Sugiyama et al., 2004; Ogihara et al., 2005).

The number of genes in the plant mitochondrial genome varies from 50 to 69, including 30 to 37 protein-coding genes. Differences are due to the variable repertoire of genes for subunits of Complex II, ribosomal proteins and tRNAs; Table 2 shows a comparison of protein-coding genes between the smallest sequenced mitochondrial genome (rapeseed, 222 kbp) and the largest one (zucchini, 983 kbp). Differences are found for six genes, which account for only 2596 bp in total, including introns (Handa, 2003; Alverson et al., 2010). It thus can be assumed that the total number of nucleotides of genes and introns is rather constant among angiosperms at ~70 kbp, whereas the genome sizes differ more than 14-fold (208 kbp versus 2936 kbp). From these data, the variability of the size of the angiosperm mitochondrial genome can be attributed to expansion or reduction of the intergenic regions. Intergenic regions are less conservative than coding regions or introns. Stretches of such intergenic sequences that are conserved among taxa have been identified, but they are rare and their significance remains unclear (Hazle and Bonen, 2007; Alverson et al., 2010). There is no evidence that the majority of intergenic regions are under any functional constraints.

A systematic survey of protein-coding genes in angiosperm mitochondria showed that many genes for subunits of Complex II and for ribosomal proteins have been lost independently in diverse angiosperm lineages (Adams et al., 2002b). For example, loss of sdh3 occurred in 40 different angiosperm lineages during the course of angiosperm diversification. Thus, there may be undiscovered mitochondrial genes that have been overlooked so far because they were lost from those mitochondrial genomes that have been extensively studied to date, such as that of Arabidopsis. In fact, rpl10 has recently been added to the list of angiosperm mitochondrial genes (Mower and Bonen, 2009; Kubo and Arimura, 2010), but it is absent from the Arabidopsis mitochondrial genome. Genes encoding proteins other than Complex II or ribosomal proteins are generally preserved in the angiosperm mitochondrial genome, but a well-known exception is cox2 which has been lost from some legume mitochondrial genomes (Nugent and Palmer, 1991; Covello and Gray, 1992).
Some protein-coding genes lost from mitochondrial genomes migrated to the nucleus and became functional in the nuclear genome. Translation products of these migratory genes are imported from the cytoplasm into the mitochondria (Adams and Palmer, 2003; Kleine et al., 2009). In other cases, no migratory copies are found in the nuclear genome. In such cases, the function of the lost genes is taken over by homologous genes that originally functioned in other cellular compartments. For example, the loss of mitochondrial rps13 is balanced by the nuclear-encoded RPS13 of plastid origin, and the loss of rps8 is compensated for by RPS15A which originally encoded cytosolic ribosomal proteins (Adams et al., 2002a).

tRNA genes in the angiosperm mitochondrial genome are classified either as ‘native’ tRNA genes that are descendents of genes of the initial endosymbiont or as tRNA genes that have been transferred to the mitochondria from the plastid genome (‘plastid’-like tRNA genes) or from unknown organisms (Marechal-Drouard et al., 1993; Kubo et al., 2000). In general, it is difficult to determine the functionality of a tRNA gene solely on the basis of its nucleotide sequence. For example, although plastid-like trnW-CCA is encoded by several angiosperm mitochondrial genes, it is expressed in wheat, maize, Oenothera, potato, sunflower, and sugar beet but not in Arabidopsis (Marechal et al., 1987; Schuster et al., 1988; Leon et al., 1989; Marechal-Drouard et al., 1990; Kubo et al., 1995; Ceci et al., 1996; Duchene and Marechal-Drouard, 2001). In this context, angiosperm species are rare in which all mitochondrial tRNA genes as well as their activities have been characterized (Joyce and Gray, 1989; Sangare et al., 1990; Duchene and Marechal-Drouard, 2001; Glover et al., 2001; Duchene et al., 2009). In wheat and Arabidopsis, three expressed mitochondrial tRNA genes differ in their genetic origin and four have been lost from either genome (Table 3). Lack of tRNA genes in the mitochondrial genome is compensated for by the import of cognate tRNA molecules from the cytoplasm (Salinas et al., 2008; Duchene et al., 2009).

In the angiosperm mitochondrial genomes listed in Table 1, the eleven genes interrupted by an intron(s) include nad1, nad2, nad4, nad5, nad7, cox1, cox2, ccmFC, rpl2, rps3, and rps10. All these introns are of the group II type except for that in watermelon cox1, which belongs to group I. In nad1, nad2, and nad5, trans-splicing occurs in which exons are transcribed as separate precursor RNA molecules and then combined into a single mature mRNA (Bonen, 2008). No rRNA gene has been reported to be interrupted by an intron(s) to date. A tRNA gene with an introns is of plastid origin
(Ohtani et al., 2002), and splicing of this intron has not been examined. The total size of the introns in a mitochondrial genome is 28 kbp in sugar beet, 25 kbp in maize and tobacco, 23 kbp in wheat, 32 kbp in watermelon, and 31 kbp in zucchini; the total number of introns varies from 20 (sugar beet) to 25 (grapevine) (Kubo et al., 2000; Clifton et al., 2004; Sugiyama et al., 2004; Ogihara et al., 2005; Goremykin et al., 2009; Alverson et al., 2010). Variation in intron content is well exemplified in cox1 in which a group I intron was once horizontally transferred from a fungal donor, spread from plant to plant, and lost in some angiosperm lineages (Cusimano et al., 2008; Sanchez-Puerta et al., 2008). Watermelon is the only plant in Table 1 in which cox1 includes a group I intron. In contrast to the group I intron in cox1, group II introns seem to be rather stable, but some variations are known that are due to alterations in the mode of splicing (i.e., cis versus trans) or loss of the intron (Geiss et al., 1994; Pla et al., 1995; Kubo et al., 2000; Itchoda et al., 2002; Kudla, 2002; Kim and Yoon, 2010). Acquisition of group II introns during angiosperm evolution has not been reported to date.

Plant mitochondrial genes are transcribed by T7-phage type RNA polymerase (Kuhn et al., 2009). It is possible that some accessory factors are involved (Kuhn et al., 2007). In addition to promoters that share a consensus sequence motif, non-consensus promoters are also common in angiosperm mitochondria (Holec et al., 2008). Because of the absence of an efficient transcription terminator, transcription can occur anywhere in the genome (Holec et al., 2008). Accumulation of mRNA seems to be controlled by both transcriptional and post-transcriptional processes including degradation and stabilization of transcripts where nuclear-encoded factors may be involved (Holec et al., 2008).

In many angiosperm mitochondrial transcripts, a number of specific cytidine residues are converted to uridine residues by RNA editing, but the reverse editing (uridine to cytidine) is rare (Takenaka et al., 2008). The majority of RNA editing occurs in the protein-coding regions where some modifications create the initiation and/or termination codons while others convert internal codons to better conserved ones. The number of reported editing sites in the protein-coding regions ranges from 357 to 491, but this may be an underestimation of the total number as RNA editing also occurs in introns and tRNA genes, albeit less frequently. The significance of RNA editing in these regions may be an improvement of the higher-order structure of RNA, which in turn may be associated with RNA processing (Marechal-Drouard et al., 1996; Kunzmann et al., 1998; Li-Pook-Than et al., 2004; Li-Pook-Than et al., 2007). Furthermore, RNA editing may
occur in pseudogenes that share editing sites with genuine genes (Williams et al., 1998). Some editing sites exhibit tissue specificity or nuclear genotype dependency (Bentolila et al., 2008; Zehrmann et al., 2008; Picardi et al., 2010), making it difficult to determine the precise number of editing sites in an angiosperm mitochondrial genome.

The length of the 5´- and 3´-untranslated regions (UTR) is 21–645 and 10–498 bases, respectively, in most mitochondrial genes of Arabidopsis (Forner et al., 2007). However, the lack of a 5´ UTR (rps4) or 3´ UTR (ccmC, nad6 and mttB) also has been observed (Forner et al., 2007). Interestingly, ccmC translation products have been detected (Raczynska et al., 2006); the mechanisms by which such genes are expressed are unclear. The translational initiation codon is AUG but utilization of an alternative codon is possible (Bock et al., 1994; Sunkel et al., 1994; Thomson et al., 1994; Siculella et al., 1996; Sakamoto et al., 1997; Dong et al., 1998; Handa 2003).

In general, the structure of angiosperm mitochondrial genes is quite high, but several exceptions exist. One is atp6, whose translation product consists of a highly conserved C-terminal core region that corresponds to the mature ATP6 polypeptide (249–252 amino acid residues) and a variable N-terminal extension (5–389 amino acid residues) (Onodera et al., 1999). The N-terminal extension is removed from the precursor polypeptide to generate the mature ATP6 polypeptide (Krishnasamy et al., 1994; Yamamoto et al., 2005). While the roles of the N-terminal extension in angiosperm ATP6 are unclear, the N-terminal extension of yeast ATP6 is necessary for an efficient integration of mature ATP6 into the ATP9 ring of F0 (Zeng et al., 2007). Sugar beet ccmC is another example. Its coding region encodes a conserved core region (239 amino acid residues) and an N-terminal extension (279 amino acid residues) that contains duplicated partial ATP9 and a sequence of unknown origin (Kitazaki et al., 2009). As in angiosperm ATP6, the N-terminal extension of sugar beet CCMC is removed from the precursor to generate the mature CCMC polypeptide. Accumulation of the cleaved N-terminal extension is not observed. Intriguingly, no other angiosperm ccmC encoding such an N-terminal extension is known. Another structurally altered gene is rps2, which bears a C-terminal extension of unknown origin (Kubo et al., 2005). The cleaved C-terminal extension of maize rps2 accumulates as a solitary polypeptide as well as a mature RPS2 polypeptide (Perrotta et al., 2002). In carrot, a C-terminal extension is encoded in atp8 (also known as orfB) and cox1, but both genes appear to be functional copies (Robison and Wolyn, 2006a, 2006b). In the case of sugar beet ccmC, maize rps2, carrot atp8, and carrot cox1, no
association between these altered mitochondrial genes and mitochondrial impairment is evident, suggesting that angiosperm mitochondria have an ability to cope with such genetic alterations. Furthermore, a significant number of probably nonfunctional ORFs are present in angiosperm mitochondrial genomes, and these comprise sequences of unknown origin and/or partially duplicated sequences (Marienfeld et al., 1999). Expression of such ORFs is usually repressed at the post-transcriptional level (Holec et al., 2008).

C. Transferred sequences

The angiosperm mitochondrial genome contains sequences homologous to the DNA of plastids, nuclei, other organisms such as fungi, and viruses. This suggests sequence transfer from other genetic compartments.

Sequences homologous to plastid DNA have been found in all seed plant mitochondrial genomes analyzed to date. Because the organization and nucleotide sequence of the plastid genome is well conserved (Palmer, 1990), the direction of sequence transfer probably is from the plastid to the mitochondrion. The proportion of plastid-derived sequence contents in angiosperm mitochondrial genomes varies from 1.6% to 11.5% (see references in Table 1). Surprisingly, the sequences homologous to plastid DNA in zucchini mitochondria are 113 kbp in length, which corresponds to 79% of the cucumber plastid genome (Alverson et al., 2010). One of the well-known roles of plastid-derived sequences in angiosperm mitochondria is their function as a source of tRNA genes [see section II (B) Genes in the genome]. Recently, chimeric mitochondrial \( atp1 \) and \( rrn18 \) in which parts of the coding regions had been replaced with plastid \( atpA \) and \( rrn16 \), respectively, were reported (Hao and Palmer, 2009; Sloan et al., 2010). No evidence for a preferentially transferred region of the plastid genome has been found (Wang et al., 2007). Because plastid-derived sequences are conspicuous in the mitochondrial genome, their evolutionary dynamics have been examined, leading to the conclusion that acquisition and loss of plastid-derived sequences is ongoing on an evolutionary scale. For example, a sequence homologous to \( rbcL \) has been independently acquired in diverse angiosperm species (Cummings et al., 2003).

Sequences homologous to nuclear DNA constitute 0.1% to 13.4% of angiosperm
mitochondrial genomes (see the references in Table 1). The identification of nucleus-derived sequences is more difficult than that of plastid-derived ones because complete nucleotide sequences of the nuclear DNA are not available for most plants at present. Sequences homologous to nuclear DNA include retrotransposons and protein-coding genes (Marienfeld et al., 1999; Kubo et al., 2000; Notsu et al., 2002; Alverson et al., 2010). The function of nucleus-derived DNA in angiosperm mitochondria is not known, but their ongoing acquisition and loss appears likely. For example, intergenic sequences that have homology to nuclear DNA were found in some mitotypes of sugar beet. These sequences are not always preserved in the other mitotypes (Satoh et al., 2006; Kawanishi et al., 2010).

From an evolutionary viewpoint, it is interesting that plastid-derived sequences have been found in the mitochondrial genomes of the gymnosperm Cycas taitungensis and the quillwort Isoetes engelmannii (Wang et al., 2007; Grewe et al., 2009). Furthermore, a nucleus-derived sequence was found in the mitochondrial genome of Isoetes (Grewe et al., 2009). Because the angiosperm lineage diverged from the Cycas and Isoetes lineages at least 300 and 350 million years ago, respectively, the first sequence transfers from plastids and nuclei to mitochondria must have occurred before that time. A mitochondrial sequence in Cycas of 18 kbp is homologous to plastid DNA and includes tRNA genes. It is possible that some, if not all, of these tRNA genes are functional in mitochondria. Mitochondrial genomes of the liverwort Marchantia polymorpha and the moss Physcomitrella patens have no plastid- or nucleus-derived sequences (Oda et al., 1992; Terasawa et al., 2007).

Emerging evidence indicates that angiosperm mitochondrial genomes contain sequences that have been horizontally transferred from non-plant organisms as well as plants (Bergthorsson et al., 2003, 2004; Won and Renner, 2003; Davis and Wurdack, 2004; Mower, et al., 2004; Davis et al., 2005; Barkman et al., 2007; Bock, 2010). The group I intron in cox1 is one such example (Sanchez-Puerta et al., 2008) [see section II (B) Genes in the genome]. A known virus-derived sequence in Arabidopsis mitochondria is absent from sugar beet (Marienfeld et al., 1999; T. Kubo, unpublished data), suggesting that either the acquisition of this sequence occurred independently in a predecessor of Arabidopsis or that this sequence has not been maintained in all plant species. Various plasmid-derived DNAs, episomal elements that are independent of the mitochondrial genome, are sometimes found in angiosperm mitochondria; their origin may be fungi
(Handa, 2008). Remnants of plasmid-like DNA integration into the mitochondrial genome have been reported (Kubo et al., 2000; Allen et al., 2007; McDermott et al., 2008; Goremykin et al., 2009). Plant-to-plant horizontal transfer has been shown by phylogenetic analyses of mitochondrial gene sequences (Richardson and Palmer, 2007). However, the frequency of horizontal transfers in plant mitochondrial genome evolution remains controversial (Goremykin et al., 2009).

D. Origin of unknown sequences in the intergenic region

The presence of sequences of unknown origin in the intergenic regions is one of the major mysteries of the angiosperm mitochondrial genome (Kubo and Mikami, 2007; Kubo and Newton, 2008). For example, sequences of unknown origin in sugar beet mitochondria – that is, sequences without homologs in the databases – account for 55.6% of the genome (Kubo et al., 2000). Although these sequences have been considered unique to sugar beet mitochondria, it is quite possible that their apparent uniqueness is due to our limited knowledge of other mitochondrial genomes. Moreover, as research progresses, sequence data of plant nuclear DNA or DNA from other organisms will accumulate in the databases, and this new information may help to identify homologies of intergenic regions in the angiosperm mitochondrial genome. Alternatively, sequences without identifiable homologs may be truly unique because of the accumulation of unique mutations (Satoh et al., 2006).

E. How have angiosperm mitochondrial genomes expanded?

Cucurbitaceous plants lend themselves to observing genome-level variation in the mitochondrial genome because this taxonomic group includes plants with average-sized (380 kbp, watermelon) as well as the largest known genomes (2936 kbp, maskmelon) (Ward et al., 1981; Alverson et al., 2010). Introns in cucumber and zucchini generally are longer than in other angiosperm mitochondria (Bartoszewski et al., 2009; Alverson et al., 2010), but it is too early to conclude that intronic expansion is involved in genomic expansion, because longer introns are also observed in the average-sized mitochondrial
genome of watermelon. However, it is possible that the long introns are remnants of the common ancestral genome of this family that may have had large mitochondrial genome (Alverson et al., 2010). On the other hand, the number of short (<1 kbp) repeat sequences appears to exhibit a correlation with the size of the mitochondrial genome in this family: large mitochondrial genomes such as those of zucchini and cucumber are rich in short repeated sequences compared to watermelon (Lilly and Havey, 2001; Alverson et al., 2010). No correlation between the mitochondrial and nuclear genome size has been found in the family Cucurbitaceae (Arumuganathan and Earle, 1991; Ren et al., 2009; Alverson et al., 2010).

The size of the grapevine mitochondrial genome is the second largest after zucchini (Table 1). A common feature shared by grapevine and zucchini mitochondria is the relatively large number of plastid-derived sequences (Goremykin et al., 2009; Alverson et al., 2010). This suggests that transferred sequences could contribute to angiosperm mitochondrial genome expansion. Repeated sequences of the grapevine mitochondrial genome have not been investigated in detail so far.

F. How did angiosperm mitochondrial genomes become so varied?

DNA recombination plays an important role in the alteration of the organization of the angiosperm mitochondrial genome. The rate of nucleotide substitutions in the angiosperm mitochondrial genome is low compared with that in animal mitochondrial genomes (Palmer, 1990), with some exceptions (Laroche et al., 1997; Cho et al., 2004; Mower et al., 2007; Barr et al., 2007; Sloan et al., 2008, 2009). Comparative analysis of the mitochondrial genomes between closely related species and within species revealed genome rearrangements as well as deletions and/or insertions, which are based on DNA recombination mediated by short (<1 kbp) repeat sequences (Palmer, 1988; Allen et al., 2007; Nishizawa et al., 2007; Marechal and Brisson, 2010). Considering the large number of short repeat sequences (for example, 33 repeat families of 108–556 bp are found in Arabidopsis mitochondria), the integrity of the plant mitochondrial genome would be endangered if recombination occurred freely. Nuclear genes that apparently suppress such recombination have been identified (Abdelnoor et al., 2003; Zaegel et al., 2006; Shedge et al., 2007; Odahara et al., 2009).
A current model of angiosperm mitochondrial genome evolution involving DNA recombination includes the regulation of the copy number of recombinant DNA molecules: a recombinant molecule can either be maintained as a substoichiometric DNA species, become a major DNA species, or become lost (Small et al., 1989). The transition from a substoichiometric to a major species or vice versa is called substoichiometric shifting (SSS). SSS occurs during angiosperm evolution and can be induced by a specific genotype or physiological condition (Small et al., 1989; Kanazawa et al., 1994; Janska et al., 1998; Arrieta-Montiel et al., 2001; Abdelnoor et al., 2003; Kuzmin et al., 2005; Zaegel et al., 2006; Shedge et al., 2007; Bartoszewski et al., 2007; Arrieta-Montiel et al., 2009). This suggests that although DNA recombination is under strict control in angiosperm mitochondria, some DNA molecules with accumulated mutations can ‘hide’ in the angiosperm mitochondria and can be transmitted to progeny.

Intraspecific variation is common in angiosperm mitochondrial genomes. Examples of such variations include rather small alterations such as deletions (B. nigra), inversions (B. hirta), or an altered number of tandem repeats (sugar beet) (Palmer, 1988; Cheng et al. 2010). On the other hand, extensive genome rearrangements have also been documented (Fauron and Casper, 1994; Nishizawa et al., 2007). For example, organizational differences between NA- and NB-type mitochondrial genomes (both can be used as the cytoplasmic genetic component of maintainer lines) of maize are characterized by 16 rearrangements and the occurrence of 2% and 5 %, respectively, of mitotype-specific sequences not present in the other genome (Allen et al., 2007). Therefore, diversification of the angiosperm mitochondrial genome by the above-mentioned mechanisms appears to be rapid. On the other hand, coding regions are generally highly conserved; however, a 6-base insertion and a synonymous substitution were found in atp4 (also known as orf25) and nad4L, respectively (Allen et al., 2007). Similar results were obtained for rice and wheat when nucleotide sequences of the mitochondrial genomes were compared between varieties (Tian et al., 2006; Cui et al., 2009).

III. The mitochondrial genome and cytoplasmic male sterility (CMS)

A. Potential sources of CMS
Before describing the mitochondrial genomes of CMS plants, it is worth reviewing the sources of CMS. In some cases, natural CMS plants have been found in a crop variety or a breeding line. In other cases, CMS plants have been obtained from crosses between different cultivars, species, and genera (Kaul, 1988). Table 1 contains seven mitochondrial genomes that are known to induce MS on certain nuclear backgrounds. Owen-type CMS in sugar beet, T-type, S-type, and C-type CMS in maize, and nap-type CMS in rapeseed were found in crop varieties or breeding lines (Owen, 1942; Duvik, 1965; Beckett, 1971; Thompson, 1972). Genealogical analysis based on sequence polymorphism of plastid and mitochondrial DNA supported the notion that Owen-type cytoplasm is an intraspecific variant rather than introgressed alien cytoplasm (Fenart et al., 2006; Nishizawa et al., 2007). On the other hand, LD- and CW-type CMS in rice were identified after crossing different cultivar groups (japonica and indica) and different species (*Oryza sativa* and *O. rufipogon*), respectively. In both cases, the cytoplasmic donor parent is MF, but MS was expressed after recurrent backcrossing with a japonica cultivar as the pollen parent (Katsuo and Mizushima, 1958; Watanabe et al., 1968). The plants obtained from the reciprocal crosses were MF (Katsuo and Mizushima, 1958; Watanabe et al., 1968).

Another source of CMS is cell fusion, which can induce a novel type of CMS (Dubreucq et al., 1999; Carlsson et al., 2007). Recently, it was shown that RNA interference directed at a gene involved in mitochondrial genome stability induced MS with an altered mitochondrial genome (Sandhu et al., 2007). The altered mitochondrial genome was transmitted through the female parent and continued to induce MS (i.e. it never reverted to an MF genome after segregation of the transgene). No Rf for this MS has been reported to date and this is a future challenge.

B. What do mitochondrial genomes of CMS plants tell us?

Molecular studies of CMS have scrutinized mitochondrial genome organization ever since CMS-associated loci were identified in mitochondria. In addition to the seven CMS mitochondrial genomes shown in Table 1, physical maps of pet1-type CMS (sunflower), Ogura-type CMS (radish), pol-type CMS (rapeseed), G-type CMS (sea beet, a wild relative of sugar beet), common bean CMS, and petunia CMS have been provided
Comparisons of CMS and MF (i.e., maintainer line) mitochondrial genomes suggest that some CMS mitochondrial genomes are highly rearranged and include inversions, insertions and deletions, whereas others preserve colinearity and possess few rearranged regions. The former class includes two sugar beet CMS, three maize CMS, petunia CMS, and two rice CMS, while the latter class comprises sunflower CMS and pol-type CMS in rapeseed. Nucleotide sequences of CMS mitochondria suggest that DNA recombination leading to genome rearrangement is mediated by short repeat sequences (Satoh et al., 2006; Allen et al., 2007). Nucleotide sequences unique to CMS mitochondria contribute 5% to 14% of the genome, whereas those unique to MF mitochondria amount to 3% to 8% (Satoh et al., 2006; Allen et al., 2007). Genomic alterations found in CMS mitochondria reflect the mitochondrial genome diversity in angiosperms, supporting the view that CMS mitochondria have emerged through the common evolutionary mechanisms acting on angiosperm mitochondrial genomes.

It should be emphasized that the portion of the CMS genotypes listed in Table 1 that contains conserved genes of known function (compare Table 2) is indistinguishable from that of MF mitochondria of the same species. In other words, the normal repertoire of mitochondrial genes is present in CMS mitochondria, suggesting that deleterious mutations are unlikely to be responsible for CMS (Satoh et al., 2004; Allen et al., 2007; Fujii et al., 2010a) [see section III (C) CMS-associated loci for the only known exception]. Sequences of tRNA genes are identical in CMS and MF mitochondria. In rRNA- and protein-coding genes, some nucleotide substitutions and small indels exist, but the effects of these mutations, if any, and their phenotypic consequences have not been investigated to date.

C. CMS-associated loci

An apparent lesion of a genuine mitochondrial gene in CMS mitochondria has been found only in G-type CMS in sea beet (Beta vulgaris ssp. maritima) (Ducos et al., 2001). In this CMS type, a nonsense mutation truncates eight C-terminal amino acid residues of COX2, and another mutation adds 14 residues to the C-terminus of NAD9 (Fig. 2) (Ducos et al.,
2001). *Rf* for the G-type CMS exists in populations of sea beet and sugar beet (Touzet et al., 2004; Kawanishi et al., 2010), but its molecular action and the gene products remain unclear. In other cases where CMS-associated loci have been identified, the loci have included ORFs unique to CMS mitochondria. A well-known example is an ORF of 115 amino acid residues termed *urf13-T* in maize T-type CMS, which consists of duplicated fragments of the coding and flanking regions of the *rrn26* gene (Fig. 2) (Dewey et al., 1986). Like *urf13-T*, unique ORFs found in the CMS-associated loci exhibit a characteristic organization that presumably results from an accumulation of mutations; otherwise, their origin is unknown. In general, homologies between the unique ORFs are rare: CMS genomes from different sources usually include different and unique ORFs. One exception is the involvement of *orf224*-variants in some *Brassica* CMS. *orf224* was first discovered in pol-type CMS (Singh and Brown, 1991) before a related ORF, *orf222*, was found in nap-type CMS (Fig. 2) (L’Homme et al., 1997), and yet another variant exists in stem mustard CMS (Yang et al., 2009). In radish, *orf138* was first identified as a unique ORF in Ogura-type CMS (Fig. 2) (Bonhomme et al., 1992), and at least nine variants have been found subsequently in wild radish accessions and radish cultivars (Iwabuchi et al., 1999; Yamagishi and Terachi, 2001). In rice, *orf79* was initially identified as the ORF associated with BT-type CMS (Fig. 2) (Iwabuchi et al., 1993; Akagi et al., 1994). Later, homologous sequences have been found in other sources of rice CMS (Li et al., 2008; Itabashi et al., 2009). Interestingly, trans-taxonomic homology has also been observed between *orf79* and *orf107* (Fig. 2), which is associated with A3-type CMS in sorghum (Tang et al., 1996).

CMS-associated unique ORFs (hereafter CMS-ORF) occur near the 5’ or 3’ termini of genuine protein-coding genes, and their expression is generally constitutive. CMS-ORFs appear most frequently associated with genes for subunits of ATPases (Fig. 2) (Hanson and Bentolila, 2004). Intriguingly, the 5’ UTR of mitochondrial ATPase subunit genes are known to be highly variable in angiosperms (Hazle and Bonen, 2007). On the other hand, *Brassica orf222* and petunia *pcf* are linked to *nad5* (exon 3) and *nad3*, respectively, both of which encode subunits of Complex I (Young and Hanson, 1987; Rasmussen and Hanson, 1989; L’Homme et al., 1997; Fig. 2). The close linkage between CMS-ORFs and genuine mitochondrial genes may facilitate their co-transcription (‘opportunistic’ transcription; Touzet and Budar, 2004). The significance of such co-transcription for MS expression can be derived from data indicating that the molecular
action of $Rf$ sometimes involves cleavage of the CMS-ORF region from the linked genuine mitochondrial gene (Kazama et al., 2008). In addition, the linkage $orf138$-$atp8$ was broken in a fertility-reverted cybrid of *Brassica*, in which the $orf138$ mRNA had become unstable (Bellaoui et al., 1997). On the other hand, monocistronic CMS-ORFs such as sorghum $orf107$ and sugar beet $orf129$ exist (Fig. 2) (Tang et al., 1996; Yamamoto et al., 2008); it seems possible that their 5′ and 3′ regions are sufficient to ensure expression.

Approaches to identify CMS-associated loci include analyses of revertants or cybrids that have a recombinant mitochondrial genome derived from CMS and MF plants (Schnable and Wise, 1998; Hanson and Bentolila, 2004). Given that CMS-ORFs are often cotranscribed with mitochondrial genes, it also may be worthwhile attempting to find ORFs near genuine mitochondrial genes, especially those for ATPase components. Indeed, this approach has worked in some cases (Singh and Brown, 1991; Dieterich et al., 2003; Shinada et al., 2006; Kim et al., 2007; Ashutosh et al., 2008). On the other hand, the comparative genomics approach has attracted few researchers for the identification of CMS-ORF, because comparisons of mitochondrial genomes required the construction of physical maps or nucleotide sequencing of more than 600 kbp in total (see II. A. Genome size and organization), which seemed impracticable in most circumstances. Additionally, the detected alterations probably would be too numerous to be investigated in detail. Therefore, this approach was preferred only when the organizational difference between the mitochondrial genomes was estimated to be rather small. However, because of technical advances in sequencing entire mitochondrial genomes, this approach may become helpful for identifying CMS-associated loci. Comparative genomics of CMS and MF mitochondria in sugar beet helped to locate the CMS-ORF, $preSatp6$, which is an exceptional N-terminal extension because the cleaved polypeptides accumulate in CMS mitochondria (Fig. 2) (Satoh et al., 2004; Yamamoto et al., 2005; Matsunaga et al., 2010). Comparison of multiple genomic sequences could lead to the identification of a CMS-ORF in maize (Allen et al., 2007), and a similar approach was used in rice (Fujii et al., 2010a).

However, a unique ORF found in CMS mitochondria but absent in MF will not necessarily be associated with MS expression, and CMS likely is independent of the vast majority of unique ORFs. Many unique ORFs unrelated to CMS may be present in CMS mitochondria, and some of them may even be cotranscribed with genuine mitochondrial
genes (Handa, 2003; Satoh et al., 2004; Allen et al., 2007). In conclusion, nucleotide sequences can help but are insufficient for discriminating candidate CMS-ORFs. To identify a CMS-ORF, it is recommended that alterations in the expression of the candidate ORF in response to Rf should be investigated as such alterations are observed in many CMS-ORFs (Hanson and Bentolila, 2004) (see below, V. Nuclear genes that suppress CMS).

Polypeptides translated from a CMS-ORF often contain several membrane-spanning domains and/or a hydrophobic N-terminal domain (Schnable and Wise, 1998; Hanson and Bentolila, 2004). Localization studies revealed that the translation products of URF13-T in maize T-type CMS and of radish ORF138 are located in the inner mitochondrial membrane (Dewey et al., 1987; Duroc et al., 2009) where they form oligomers (Rhoads et al., 1998; Duroc et al., 2009). In this context, oligomer formation of preSATP6 may indicate its association with CMS (Yamamoto et al., 2005). On the other hand, some polypeptides including petunia PCF and sugar beet ORF129 (Fig. 2) have been detected in both the membrane and matrix fractions, suggesting that they are loosely associated with the membrane (Nivison et al., 1994; Yamamoto et al., 2008). The common bean ORF239 is unique because it can be detected outside of mitochondria (Abad et al., 1995).

It certainly is reassuring if translation products of the candidate ORF can be detected, but further evidence is needed to conclude that the candidate is the causal agent of CMS. The mitochondrial ORF that is responsible for disease sensitivity in rough lemon is translated but it is not associated with CMS (Ohtani et al., 2002; K. Akimitsu, personal communication). The potential of an ORF to induce MS has been investigated via a transgenic approach based on the introduction of a construct in which an ORF fused with the mitochondrial transit peptide is driven by a constitutive or tissue-specific promoter. The resulting transgenics expressed MS and the phenotype was cosegregated with the transgene in the progeny (He et al., 1996; Wang et al., 2006; Kim et al., 2007; Yamamoto et al., 2008; Nizampatnam et al., 2009). However, it remains puzzling that not all CMS-ORFs induced MS in similar experiments (Chaumont et al., 1995; Wintz et al., 1995; Duroc et al., 2006). Intraorganellar localization of translation products and/or the expression profile of the transgene might be different from those of endogenous CMS-ORFs.
IV. How is CMS expressed?

In sugar beet G-type CMS, which is characterized by a truncated *cox2* and a 3’-terminally extended *nad9*, reduced activity was observed in Complex IV but not in Complex I in young leaves (Ducos et al., 2001). Insufficient Complex IV activity may be associated with CMS, but it cannot be ruled out that Complex I activity is also reduced in anthers (Ducos et al., 2001).

In cases of CMS involving a CMS-ORF, the ORF’s translation product may modify mitochondrial function and thereby induce MS. CMS-ORFs have the potential to modify mitochondrial function, according to data from experiments using a heterologous system. Intensive studies using *Escherichia coli*, budding yeast, and insect cells have clarified biochemical aspects of *urf13-T*, which pleiotropically induces sensitivity to southern corn leaf blight and chemicals such as the insecticide methomyl (Dewey et al., 1988; Glab et al., 1990; Huang et al., 1990; Korth and Levings, 1993; Rhoads et al., 1998). It was concluded that *urf13-T* translation products form a tetramer in the membrane which functions as a molecular pore when it binds to the fungal toxin or the insecticide. As a result, ion leakage as well as uncoupling of oxidative phosphorylation occurs. The existence of an anther-specific substrate resembling the fungal toxin or the insecticide would resolve the question why the CMS-ORF causes MS but affects no other organ despite its constitutive expression, as proposed by Flavell (1974). However, no such substrate, the so-called factor X, has been found to date.

On the other hand, some CMS-ORFs seem to be cytotoxic without an additional substrate. Radish *orf138* and sunflower *orf522* have the potential to inhibit growth of *E. coli* (Nakai et al., 1995; Duroc et al., 2005). Furthermore, mitochondrial morphology changed when the radish ORF138 polypeptide accumulated in the mitochondria of yeast, onion, and *Arabidopsis* (Duroc et al., 2006). Cytotoxicity of rice ORF79 has also been reported; expression of transgenic ORF79 interferes with the regeneration of rice callus (Kojima et al., 2010). It should be noted that these effects only occur when the expressed CMS-ORF is intact: truncated ORF138 does not affect growth of *E. coli* (Duroc et al., 2005). Therefore, it seems likely that the amino acid sequence of CMS-ORF is functionally significant and has an unknown physiological effect, although the alternative hypothesis that accumulation of aberrant polypeptides *per se* is stressful for mitochondria
cannot be ruled out.

If the CMS-ORF is more or less cytotoxic, its effects should show in organs/tissues other than the anthers but this is rarely the case. Duvik (1965) noted that maize T-type CMS affected plant height, number of leaves, and grain yield. It is generally assumed that such phenotypic effects are so small that they are hardly detectable. Despite the lack of visible phenotypes, differential gene expression in non-anther organs has been detected in maize. Comparison of the mitochondrial proteome in ears of the NA- and T-type cytoplasm of maize (see Table 1) revealed mitotype-specific expression of several nuclear genes (Hochholdinger et al., 2004), suggesting a cytoplasmic effect.

Emerging evidence suggests that mitochondria in the vegetative organs of CMS lines are physiologically distinct from those of maintainer lines. In sunflower pet1-type CMS, Complex V (F₀F₁-ATPase) activity is reduced in seedlings (Sabar et al., 2003). A competitive effect between genuine *atp8* and *orf522* which consists of a partially duplicated *atp8* and a sequence of unknown origin (Fig. 2), has been postulated (Saber et al., 2003). Decreased Complex V activity has also been reported from rice HL-type CMS seedlings (Zhang et al., 2007) in which *orfH79* was co-transcribed with *atp6*. On the other hand, no reduction of any respiratory complex activity was observed in radish Ogura-type CMS (Duroc et al., 2009). Instead, mitochondrial oxygen consumption was altered, which led Duroc et al. (2009) to hypothesize that the ORF138 polypeptide exerts a mild uncoupling effect by forming oligomers. The terms “dominant defective mutation,” which refers to CMS-ORFs such as sunflower *orf522*, and “gain of function,” which refers to CMS-ORFs such as radish *orf138*, have been proposed for CMS-ORFs with different effects (Duroc et al., 2009).

To date, it remains obscure why mitochondrial function is changed in the entire CMS plant whereas phenotypic effects above the level of the mitochondria are restricted to anthers. The common bean is an exception to the rule because an accumulation of ORF239 occurs specifically in the mitochondria of anthers and involves mitochondrial protease (Abad et al., 1995; Sarria et al., 1998). A hypothesis currently applied to many CMS plants assumes differential requirements for mitochondrial activity between organs: the effects of CMS-ORF are so small that most organs remain phenotypically normal. Since highest mitochondrial activity is required in anthers, the small effects are sufficient to cause MS. Many studies including morphological and molecular analyses support this differential mitochondrial activity hypothesis (Warmke and Lee, 1977; Majewskasawka
et al., 1993; Huang et al., 1994; Lalanne et al., 1998).

Research into the mechanisms of CMS has focused on the anther tapetum, a nursing tissue that surrounds male meiocytes, because it is often impaired in association with CMS (Laser and Lersten, 1972; Warmke and Lee, 1977; Kaul, 1988; Majewskasawka et al., 1993). It appears likely that the mitochondrial function that is modified by CMS-ORF affects the function and/or integrity of the tapetum. Programmed cell death (PCD) of the tapetum is a key process in pollen development (Skibbe et al., 2008; Wilson and Zhang, 2009; Parish and Li, 2010). In sunflower pet1-type CMS, premature cell death is associated with MS expression, including the advanced release of cytochrome c, which triggers PCD in animal systems (Balk and Leaver, 2001). On the other hand, the tapetum morphology of some CMS lines appears indistinguishable from that of their maintainer lines. In maize S-type CMS, only pollen grains after the first or second pollen mitosis exhibited developmental abnormalities (Laughnan and Gabay-Laughnan, 1983) which suggests that the male gametophyte itself is most affected by S-type CMS.

Increased levels of reactive oxygen species (ROS), which are associated with the onset of PCD (Gamaley and Klyubin, 1999; Jabs, 1999), have been reported in rice HL-type CMS and cotton CMS anthers (Wan et al., 2007; Jiang et al., 2007). On the other hand, no apparent increase in ROS was detected in maize T-type CMS anthers (Liu and Schnable, 2002). Although ROS involvement is uncertain, anther tissues of CMS plants appear to be under stress, at least in maize, sorghum, and rice, because AOX, a nuclear gene encoding alternative oxidase, is more active in CMS plants than in maintainer strains (Pring et al., 2006; Karpova et al., 2002; Fujii and Toriyama, 2008a). AOX is a terminal oxidase in the mitochondrial electron transport chain which is a functional alternative to Complex IV (cytochrome c oxidase). AOX is known to be a marker of different types of stress. Activation of AOX is triggered by high respiratory substrate availability or high endogenous or exogenous levels of ROS (Polidoros et al., 2009). Increased AOX expression was observed in the tassel but not in the ear of maize CMS (Karpova et al., 2002). The degree of gene-induction differs among T-, S- and C-type CMS (Karpova et al., 2002), suggesting a differential cytoplasmic effect. On the other hand, no increased enzymatic activity of AOX was seen in the anthers of common bean and sorghum CMS (Johns et al., 1993; Sane et al., 1997), and reduced AOX activity was reported from immature anthers of petunia (Connett and Hanson, 1990).

Altered expression of nuclear genes in CMS anthers other than AOX have also been
reported (Pring and Tang, 2004; Geddy et al., 2005; Matsuhira et al., 2007; Yang et al., 2008; Fujii and Toriyama, 2008a; Fujii et al., 2009; Fujii et al., 2010b). The mechanism by which the expression of these genes is affected is unknown, but it is possible that some of the genes, if not all, are under the regulation of retrograde signaling, which senses the mitochondrial condition and regulates nuclear gene expression in response.

V. Nuclear genes that suppress CMS

In most cases, dominant alleles of $Rf$s restore MF, whereas fertility restoration by the recessive allele is rare. Fertility restoration by $Rf$ is either sporophytic and all pollen grains are viable in heterozygous ($Rfrf$) as well as in homozygous plants ($RfRf$), or it is gametophytic and only pollen grains that inherit the $Rf$ allele are viable.

The finding that maize $Rf1$ altered the mRNA profile of $urfl3-T$ and reduced the amount of URF13-T polypeptides (Dewey et al., 1987) raised the question whether $Rf$ generally controls the expression of CMS-ORF. In many cases, it does; RNA gel blot analyses using CMS-ORF as a probe detected different mRNA profiles in CMS and fertility-restored plants (Hanson and Bentolila, 2004). The underlying mechanism included the loss of a specific size class of RNA molecules and/or altered processing of RNA molecules. Transcriptional alteration by $Rf$ may occur in the entire plant or in a specific organ, e.g. in anthers. Simultaneously, CMS-ORF translation products are decreased. In the radish Ogura-type CMS, the mRNA profile is unchanged but accumulation of the ORF138 polypeptide is decreased by $Rfo$ (Krishnasamy and Makaroff, 1994; Bellaoui et al., 1999; Uyttewaal et al., 2008).

The effects of $Rf$s seem confined to the locus containing the CMS-ORF but some $Rf$s such as $Rfn$ in Brassica (for nap-type CMS), $Rf3$ and $rfll$ in maize (for S-type CMS), and sorghum $Rf3$ (for A3-type CMS) have been reported to pleiotropically alter the transcription profile of additional mitochondrial loci (Li et al., 1998; Tang et al., 1998; Wen and Chase, 1999; Wen et al., 2003). However, the possibility cannot be ruled out that apparent pleiotropic effects of $Rf$ in fact are due to other genes because the $Rf$ locus often exhibits a clustering of similar genes (Touzet and Budar, 2004). On the other hand, maize $rfll$ appears to have certain pleiotropic effect. Maize $rfll$ is an exceptional $rf$ as the recessive allele restores MF in a gametophytic manner (Wen et al., 2003). The recessive
allele reduces the accumulation of orf355-orf77 (Fig. 2) as well as atp1 transcripts, even though the two genes are 9 kbp apart (Wen et al., 2003; Allen et al., 2007). Since the functional dominant allele of rfl1 seems necessary for orf355-orf77 and atp1 expression (Wen et al., 2003), the homozygous recessive (rfl1rfl1) zygote is embryonic lethal because of an ATP1 deficiency. However, rfl1 pollen can germinate and participate in fertilization. It is worth noting that in pollen less mitochondrial activity is required because energy production in pollen relies on a unique pathway (Mellema et al., 2002; Gass et al., 2005). The rfl1 gene product still is unknown.

Rf5s of petunia, radish, and rice reduce the amount of CMS-ORF translation products. Details of how this is achieved differ; for example, the mRNA profiles of rice orf79 are altered whereas that of petunia pcf and radish orf138 remains unchanged (Rasmussen and Hanson, 1989; Nivison et al., 1994; Iwabuchi et al., 1993; Krishnasamy and Makaroff, 1994; Bellaoui et al., 1999; Uyttewaal et al., 2008). Molecular cloning of Rf5s from the three species revealed that all three Rf5s encode proteins containing a common degenerate motif called a pentatrico peptide repeat (PPR) (Bentolila et al., 2002; Koizuka et al., 2003; Kazama and Toriyama, 2003; Desloire et al., 2003; Brown et al., 2003; Komori et al., 2004; Akagi et al., 2004). Genes encoding PPR proteins constitute a large gene family in land plants and are involved in a range of transcriptional and post-transcriptional processes including transcription, splicing, RNA cleavage, RNA editing, and translation (Schmitz-Linneweber and Small, 2008). PPR proteins encoded by Rf5s seem to have no catalytic activity, and therefore are likely to be ‘molecular adaptors.’ (Schmitz-Linneweber and Small, 2008). Provided that RF proteins bind RNA in a sequence-specific manner and recruit catalytic proteins to the site of action, the Rf5s could affect the expression of specific genes such as CMS-ORF. Obviously, these Rf5s could affect several genes that share their binding sites. It has been shown that petunia RF, rice RF1, and radish RFO bind to transcripts containing pcf, orf79 and orf138, respectively (Gillman et al., 2007; Kazama et al., 2008; Uyttewaal et al., 2008). On the basis of these results, it is plausible to hypothesize that (some) Rf genes encode PPR proteins that control the activity of mitochondrial genes including those that function in CMS. In some plant species, a genetic linkage between Rf and PPR-protein coding genes has been found (Klein et al., 2005; Jordan et al., 2010; Barr and Fishman, 2010). It would be interesting to see whether any unique ORFs responding to the Rf can be identified in the mitochondrial genomes of such plants.
Rfs may encode several types of proteins other than PPRs. Common bean \(Fr\) (the genetic symbol for \(Rf\) in this species) is a unique \(Rf\) because it acts as if it deleted the mitochondrial (mt) DNA region termed \(pvs\) which contains the CMS-ORF, \textit{orf239} (Fig. 2) (Janska and Mackenzie, 1993; Janska et al., 1998). This phenomenon involves regulation of the copy number of specific mtDNA molecules, that is, SSS. The copy number of mtDNA molecules that uniquely contain \(pvs\) but are otherwise colinear with other molecules is decreased by an unknown mechanism, presumably during oogenesis and/or transmission to zygotes. A similar phenomenon was discovered in perl millet (Feng et al., 2009) and rapeseed (J. Imamura, personal communication). These reports clearly indicate that SSS is under genetic control and is one of the mechanisms for eliminating CMS-associated regions.

In these examples, \(Rf\) appears to reduce the amount of CMS-ORF translation products by modifying gene expression or eliminating the mtDNA molecule that encodes CMS-ORF. In these cases, the relationship between CMS-ORF and \(Rf\) is clear and consistent with a genetic model featuring specific interactions between MS-inducing mitochondria and \(Rf\). In contrast, other Rfs are more difficult to place in a direct relationship with CMS-ORF. Maize \(Rf2a\) for T-type CMS was the first cloned \(Rf\) and encodes mitochondrial aldehyde dehydrogenase (Cui et al., 1996). Although progress has been made in understanding functional aspects of the RF2A protein (Liu et al., 2001; Liu and Schnable, 2002), it remains unclear how RF2A interacts with \textit{urf13-T}, which has raised doubts about the specificity of \(Rf2a\) for T-type CMS (Touzet, 2002).

Molecular cloning of \(Rf17\) which functions to restore rice CW-type CMS in a gametophytic manner, has been successful but the precise function of its translation product (RMS) is unclear (Fujii and Toriyama, 2009). Interestingly, rice CW-type CMS plants express more RMS than fertility-restored rice. Transgenic rice over-expressing RMS exhibited MS, indicating that an excess amount of RMS is harmful for pollen production (Fujii and Toriyama, 2009). It was concluded that CW-type mitochondria induce over-expression of RMS via a retrograde signaling pathway, thereby causing MS, whereas a non-inducible allele blocks MS expression resulting in fertility restoration. If so, the role of \(Rf17\) is not the elimination of the causal agents of CMS, but the obstruction of the execution of MS. Identification of the mitochondrial locus that is associated with CW-type CMS is necessary for further insights into this process.
VI. Concluding remarks and perspectives

Because of its mode of inheritance, CMS provides a unique means of preventing pollen dispersal. For CMS utilization, it is necessary to find an MS-inducing mitotype. If multiple mitotypes are found in a gene pool and if their nucleotide sequences are available (it will be much easier to obtain complete sequences in the future), will it be possible to predict the potential of each mitotype? At present, it is difficult to answer this question but the goal is worth the challenge. For example, although the entire nucleotide sequences for mitochondrial genomes of maize C-type CMS and rice CW-type CMS are available, the causal agents of these CMS remain obscure (Allen et al., 2007; Fujii et al., 2010a). The two mitotypes are known to be capable of inducing MS, because MS is expressed under certain nuclear genotypes. One of the characteristics of CMS mitochondria is extensive genome rearrangement, as seen in maize and sugar beet (Fauron and Havlik, 1989; Kubo et al., 1999), but this is not always the case as exemplified by sunflower pet1-type CMS (Siculella and Palmer, 1988).

Functionally unassigned ORFs are ubiquitous and abundant in the angiosperm mitochondrial genome; consequently, CMS-ORFs may “hide” among a number of candidate ORFs. Plant mitochondria have a regulatory system to repress expression of such useless or potentially harmful ORFs (Holec et al., 2008). From the viewpoint of CMS-ORF evolution, cotranscription with a genuine mitochondrial gene may be one of the solutions for CMS-ORFs to overcome the mitochondrial defense that represses the expression of aberrant ORFs (Holec et al., 2008). This arrangement may also help to avoid elimination by recombination (F. Budar, personal communication). CMS-ORF translation products are likely to be subject to mitochondrial proteolytic activity, as is seen in the tissue-specific accumulation of common bean ORF239 and in the N-terminal trimming of petunia PCF (Sarria et al., 1998; Nivison et al., 1994). Genes involved in the quality control system of mitochondrial proteins have been identified (Kolodziejczak et al., 2007). However, translation of a unique ORF and accumulation of its translation products may not always cause MS; for example, maize rps2 and sugar beet ccmC are structurally aberrant functional genes that do not appear to be associated with mitochondrial dysfunction (Perrotta et al., 2002; Kitazaki et al., 2009). Perhaps MS induction requires an additional feature such as oligomer formation as in the cases of
maize URF13-T, radish ORF138 and sugar beet preSATP6 (Rhoads et al., 1998; Yamamoto et al., 2005; Duroc et al., 2009).

There is increasing evidence indicating that a CMS-ORF is necessary for CMS expression. For example, maize fertility revertants have lost urf13-T or orf355-orf77 (Zabala et al., 1997; Schnable and Wise, 1998), and their cognate Rf decreases the amount of translation products (Hanson and Bentolila, 2004). However, researchers mostly have overlooked the fact that a few non-synonymous substitutions in genuine mitochondrial genes have been found in all CMS mitochondria that have been sequenced to date (Satoh et al., 2004; Allen et al., 2007; Fujii et al., 2010a). It is likely that these genes merely represent evolutionary polymorphism and are functionally non-significant, but it also is possible that amino acid substitutions in mitochondrial proteins are involved in MS expression. In eukaryotes other than plants, the occurrence of nuclear-mitochondrial incompatibility includes amino acid substitutions or altered UTRs of genuine mitochondrial genes, some of which are associated with MS (St John et al., 2005; Harrison and Burton, 2006; Lee et al., 2008).

Although CMS-related morphological variation has not been discussed in detail in this review, it is worth noting that tapetum degeneration prior to pollen abortion is not always associated with CMS. Laser and Lersten (1972) summarized the stages of morphological deviation from microspore mother cells to mature pollen in various CMS plants. In addition, some CMS plants are known to develop homeotic conversions of stamina to other floral organs (Linke and Borner, 2005). The mechanisms underlying these phenotypic variations of CMS plants remain unknown, but transcriptomic analysis may be an approach for solving this problem.

It has been reported that multiple Rs exist for a single CMS in some cases, which suggests that a CMS is not necessarily regulated by a single mechanism. In fact, rice orf79 expression can be suppressed by either of two genes, Rfla that cleaves atp6-orf79 mRNA, or Rflb that degrades atp6-orf79 mRNA (Wang et al., 2006). Rfla and Rflb are linked in the Rf-1 locus (Wang et al., 2006). In maize T-type CMS, the function of Rfl that alters the processing of urf13-T-atp4 transcripts can be partially replaced by Rf8 or Rf* (Dill et al., 1997). In common bean CMS, a nuclear gene that alters orf239 expression exists in addition to Fr that is associated with SSS (Mackenzie, 1991). In wild radish, a novel Rf that alters the transcription pattern of orf138, unlike the previously cloned Rfo, has been found (Bett and Lydiate, 2004; Yasumoto et al., 2008; Yasumoto et al., 2009).
Furthermore, more than two \( Rf \)s are involved not only in maize T-type CMS (which requires \( Rf1 \) and \( Rf2 \) for fertility restoration), but also in other species such as sugar beet (Owen, 1942; Owen, 1945; Kaul, 1988). Molecular characterization of these \( Rf \)s will be of great help in understanding the mechanisms and evolutionary aspects of the interactions between the nucleus and mitochondria.

**Acknowledgements** The authors wish to thank Dr. Françoise Budar for reviewing the manuscript, and Dr. Sally Mackenzie, Dr. Kinya Toriyama and Dr. Tomohiko Kazama for providing valuable information. 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).

VII. References


Ceci, L.R., Veronico, P. and Gallerani, R. 1996. Identification and mapping of tRNA genes on the Helianthus annuus mitochondrial genome. DNA Seq. 6: 159-166.


Fujii, S. and Toriyama, K. 2008a. DCW11, Down-Regulated gene 11 in CW-type cytoplasmic male sterile rice, encoding mitochondrial protein phosphatase 2C is


629-635.


Marechal-Drouard, L., Kumar, R., Remacle, C. and Small, I. 1996. RNA editing of larch mitochondrial tRNA_{His} precursors is a prerequisite for processing. Nucl. Acids Res. 24: 3229-3234.


Nugent, J.M. and Palmer, J.D. 1991. RNA-mediated transfer of the gene coxII from the


Palmer, J.D. and Shields, C.R. 1984. Tripartite structure of the *Brassica campestris*


Woloszynska, M. 2010. Heteroplasmy and stoichiometric complexity of plant


Figure legends

Figure 1 Cytoplasmic male sterility (CMS) is expressed as a result of nuclear-mitochondrial interactions. When a plant possesses an MS-inducing mitochondrial genome and a non-restoring allele \( rf \), it expresses MS (MS line). A plant with the same nuclear genotype but no sterilizing factor (N) in its mitochondrial genome is male fertile (MF), and this plant is used for the propagation of the MS line (maintainer line). Because the MS line is unable to self-pollinate, a different pollen parent is required and seeds set on the MS line are hybrid. If the hybrids are required to be MF, the pollen parent should possess a restoring allele \( Rf \). If the yield of the crop is derived from a vegetative organ, the restorer allele may not be necessary.

Figure 2 Organization of CMS-associated loci mentioned in the text. Yellow parts indicate sequences of unknown origin. Codons including two nucleotide substitutions, one truncates \( Gcox2 \) and the other extends \( Gnad9-1 \) ORFs, are boxed.
Table 1 Available sequences for plant mitochondrial master circles (MC).

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>Common name</th>
<th>Cultivar/ accession</th>
<th>Size of MC</th>
<th>Accession no.</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>Thale cress</td>
<td>Columbia</td>
<td>366924 bp</td>
<td>Y08501</td>
<td>Unseld et al., 1997</td>
</tr>
<tr>
<td>Beta vulgaris subsp. vulgaris</td>
<td>Sugar beet</td>
<td>TK81-MS</td>
<td>501020 bp</td>
<td>BA000024</td>
<td>Satoh et al., 2004</td>
</tr>
<tr>
<td>Beta vulgaris subsp. vulgaris</td>
<td>Sugar beet</td>
<td>TK81-O</td>
<td>368801 bp</td>
<td>BA000009</td>
<td>Kubo et al., 2000</td>
</tr>
<tr>
<td>Silene latifolia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brassica napus</td>
<td>Rapeseed</td>
<td>Westar</td>
<td>221853 bp</td>
<td>AP006444</td>
<td>Handa, 2003</td>
</tr>
<tr>
<td>Carica papaya</td>
<td>Papaya</td>
<td>SunUp</td>
<td>476890 bp</td>
<td>EU431224</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Citrullus lanatus</td>
<td>Watermelon</td>
<td>Florida giant</td>
<td>379236 bp</td>
<td>GQ856147</td>
<td>Alverson et al., 2010</td>
</tr>
<tr>
<td>Cucurbita pepo</td>
<td>Zucchini</td>
<td>Dark green zucchini</td>
<td>982833 bp</td>
<td>GQ856148</td>
<td>Alverson et al., 2010</td>
</tr>
<tr>
<td>Solanum tabacum (syn Nicotiana tabacum)</td>
<td>Tobacco</td>
<td>Bright Yellow 4</td>
<td>430597 bp</td>
<td>BA000042</td>
<td>Sugiyama et al., 2005</td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>Grapevine</td>
<td>Pinot noir clone ENTA115</td>
<td>773279 bp</td>
<td>FM179380</td>
<td>Goremykin et al., 2009</td>
</tr>
<tr>
<td>Bambusa oldhamii</td>
<td>Green bamboo</td>
<td></td>
<td>509941 bp</td>
<td>EU365401</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Oryza rufipogon</td>
<td></td>
<td>Chinese strain W1</td>
<td>559045 bp</td>
<td>AP011076</td>
<td>Fujii et al., 2010a</td>
</tr>
<tr>
<td>Oryza sativa Indica Group</td>
<td>Rice</td>
<td>93-11</td>
<td>491515 bp</td>
<td>DQ167399</td>
<td>Tian et al., 2006</td>
</tr>
<tr>
<td>Oryza sativa Indica Group</td>
<td>Rice</td>
<td>Lead rice</td>
<td>434735 bp</td>
<td>AP011077</td>
<td>Fujii et al., 2010a</td>
</tr>
<tr>
<td>Oryza sativa Japonica Group</td>
<td>Rice</td>
<td>PA64S</td>
<td>490673 bp</td>
<td>DQ167807</td>
<td>Tian et al., 2006</td>
</tr>
<tr>
<td>Oryza sativa Japonica Group</td>
<td>Rice</td>
<td>Nipponbare</td>
<td>490520 bp</td>
<td>BA000029</td>
<td>Notsu et al., 2002</td>
</tr>
<tr>
<td>Oryza sativa Japonica Group</td>
<td>Rice</td>
<td>Nipponbare</td>
<td>490669 bp</td>
<td>DQ167400</td>
<td>Tian et al., 2006</td>
</tr>
<tr>
<td>Sorghum bicolor</td>
<td>Sorghum</td>
<td>BTx623</td>
<td>468628 bp</td>
<td>DQ984518</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Tripsacum dactyloides</td>
<td>Gama grass</td>
<td>Pete</td>
<td>704100 bp</td>
<td>DQ984517</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>Wheat</td>
<td>Chinese Spring</td>
<td>452528 bp</td>
<td>AP008982</td>
<td>Ogihara et al., 2005</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>Wheat</td>
<td>Chinese Yumai</td>
<td>452526 bp</td>
<td>EU534409</td>
<td>Cui et al., 2009</td>
</tr>
<tr>
<td>Zea mays subsp. mays</td>
<td>Maize</td>
<td>A-188, NA type mitochondria</td>
<td>701046 bp</td>
<td>DQ490952</td>
<td>Allen et al., 2007</td>
</tr>
<tr>
<td>Species</td>
<td>Type</td>
<td>Mitochondria</td>
<td>Length (bp)</td>
<td>Accession</td>
<td>Source</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Zea mays subsp. mays</td>
<td>Maize</td>
<td>B37, NB type</td>
<td>569630</td>
<td>AY506529</td>
<td>Clifton et al., 2004</td>
</tr>
<tr>
<td>Zea mays subsp. mays</td>
<td>Maize</td>
<td>B37, CMS-C</td>
<td>739719</td>
<td>DQ645536</td>
<td>Allen et al., 2007</td>
</tr>
<tr>
<td>Zea mays subsp. mays</td>
<td>Maize</td>
<td>B37, CMS-S</td>
<td>557162</td>
<td>DQ490951</td>
<td>Allen et al., 2007</td>
</tr>
<tr>
<td>Zea mays subsp. mays</td>
<td>Maize</td>
<td>B37, CMS-T</td>
<td>535825</td>
<td>DQ490953</td>
<td>Allen et al., 2007</td>
</tr>
<tr>
<td>Zea mays subsp. parviglumis</td>
<td></td>
<td></td>
<td>680603</td>
<td>DQ645539</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Zea perennis</td>
<td></td>
<td></td>
<td>570354</td>
<td>DQ645538</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
<tr>
<td>Zea luxurians</td>
<td></td>
<td></td>
<td>539368</td>
<td>DQ645537</td>
<td>Direct submission to DDBJ/EMBL/GenBank data base</td>
</tr>
</tbody>
</table>
Table 2 Comparison of gene content between the largest and smallest mitochondrial genomes in Table 1.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Name of plant</th>
<th>Genes</th>
<th>Name of plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex I</td>
<td></td>
<td>Complex I</td>
<td></td>
</tr>
<tr>
<td>nad1 *2</td>
<td>+</td>
<td>ccmb</td>
<td>+</td>
</tr>
<tr>
<td>nad2 +</td>
<td>+</td>
<td>ccmC</td>
<td>+</td>
</tr>
<tr>
<td>nad3 +</td>
<td>+</td>
<td>ccmFC</td>
<td>+</td>
</tr>
<tr>
<td>nad4 +</td>
<td>+</td>
<td>ccmFN</td>
<td>+</td>
</tr>
<tr>
<td>nad4L +</td>
<td>+</td>
<td>Protein translocation system subunit</td>
<td></td>
</tr>
<tr>
<td>nad5 +</td>
<td>+</td>
<td>mttB</td>
<td>+</td>
</tr>
<tr>
<td>nad6 +</td>
<td>+</td>
<td>matR</td>
<td>+</td>
</tr>
<tr>
<td>nad7 +</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nad9 +</td>
<td>+</td>
<td>Ribosomal protein</td>
<td></td>
</tr>
<tr>
<td>Complex II</td>
<td></td>
<td>rpl2</td>
<td>+</td>
</tr>
<tr>
<td>sdh3 +</td>
<td>−*3</td>
<td>rpl5</td>
<td>+</td>
</tr>
<tr>
<td>Complex III</td>
<td></td>
<td>rpl16</td>
<td>+</td>
</tr>
<tr>
<td>cob +</td>
<td>+</td>
<td>rps1</td>
<td>+</td>
</tr>
<tr>
<td>Complex IV</td>
<td></td>
<td>rps3</td>
<td>+</td>
</tr>
<tr>
<td>cox1 +</td>
<td>+</td>
<td>rps4</td>
<td>+</td>
</tr>
<tr>
<td>cox2 +</td>
<td>+</td>
<td>rps7</td>
<td>+</td>
</tr>
<tr>
<td>cox3 +</td>
<td>+</td>
<td>rps10</td>
<td>−</td>
</tr>
<tr>
<td>Complex V</td>
<td></td>
<td>rps12</td>
<td>+</td>
</tr>
<tr>
<td>atp1 +</td>
<td>+</td>
<td>rps13</td>
<td>+</td>
</tr>
<tr>
<td>atp4 +</td>
<td>+</td>
<td>rps14</td>
<td>+</td>
</tr>
<tr>
<td>atp6 +</td>
<td>+</td>
<td>rps19</td>
<td>−</td>
</tr>
<tr>
<td>atp8 +</td>
<td>+</td>
<td>Ribosomal RNA</td>
<td></td>
</tr>
<tr>
<td>atp9 +</td>
<td>+</td>
<td>rrn5</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rrn18</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rrn26</td>
<td>+</td>
</tr>
</tbody>
</table>

*1CMS-ORFs are not considered, *2present, *3absent, and *4pseudogene
Table 3 Origin of tRNA genes encoded in the mitochondrial genomes of *Arabidopsis* and wheat.

<table>
<thead>
<tr>
<th></th>
<th>Arabidopsis</th>
<th>Wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>trnC</em>-GCA</td>
<td>N*¹</td>
<td>P*²</td>
</tr>
<tr>
<td><em>trnD</em>-GUC</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td><em>trnE</em>-UUC</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnF</em>-GAA</td>
<td>N</td>
<td>P</td>
</tr>
<tr>
<td><em>trnG</em>-GCC</td>
<td>N</td>
<td>–*³</td>
</tr>
<tr>
<td><em>trnH</em>-GUG</td>
<td>P</td>
<td>–</td>
</tr>
<tr>
<td><em>trnI</em>-LAU</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnK</em>-UUU</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnM</em>-CAU</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnM</em>-CAU</td>
<td>–</td>
<td>P</td>
</tr>
<tr>
<td><em>trnN</em>-GUU</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td><em>trnP</em>-UGG</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnQ</em>-UUG</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnS</em>-GCU</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnS</em>-GGA</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td><em>trnS</em>-UGA</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><em>trnW</em>-CCA</td>
<td>–</td>
<td>P</td>
</tr>
<tr>
<td><em>trnY</em>-GUA</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

*¹ native class, *² plastid origin, and *³ absent from mitochondria