Cytoplasmic–Nuclear Genomic Barriers in Rice Pollen Development Revealed by Comparison of Global Gene Expression Profiles among Five Independent Cytoplasmic Male Sterile Lines

Sota Fujii1,4,5, Mari Yamada1,4, Masahiro Fujita2, Etsuko Itabashi1, Kazuki Hamada3, Kentaro Yano3, Nori Kurata2 and Kinya Toriyama1,

1Laboratory of Environmental Biotechnology, Graduate School of Agricultural Science, Tohoku University, Sendai, 981-8555 Japan
2Plant Genetics Laboratory, National Institute of Genetics, Mishima, 411-8540 Japan
3School of Agriculture, Meiji University, Kawasaki, 214-8571 Japan
4These authors contributed equally to this work
5Present address: ARC Centre of Excellence in Computer Systems Biology, University of Western Australia, 35 Stirling Highway, Crawley 6009, WA, Australia

*Corresponding author: E-mail, torikin@bios.tohoku.ac.jp; Fax, +81-22-717-8830
(Received February 9, 2010; Accepted February 28, 2010)

Keywords: Cytoplasmic male sterility • Epistacy • Microarray • Mitochondria • Oryza sativa L. • Pollen development.
Abbreviations: ABI4, abscisic acid-insensitive 4; ANOVA, analysis of variance; AOX1a, alternative oxidase 1a; CMS, cytoplasmic male sterility; GUN1, genomes uncoupled 1; HSF, heat shock factor; HSP, heat shock protein; ORF, open reading frame; PCD, programmed cell death; RFLP, restriction fragment length polymorphism; ROS, reactive oxygen species.

Introduction

To date, the wealth of genetic analysis of mitochondria and nuclear interactions supports the idea that mitochondrial genotypes are part of the genomic barrier between species, and mitochondrial–nuclear interactions should be coordinated at the genetic and/or metabolite level for sound eukaryote development. For instance, in marine copepods, hybrid breakdown occurs when mitochondrially encoded cytochrome c oxidase and nuclear-encoded cytochrome c are incompatible in their genotypes (Rawson and Burton 2002, Harrison and Burton 2006). Interpopulation crosses of copepods result in growth weaknesses, and this hybrid breakdown phenomenon was restored by the cross-introduction of the nucleus against its own mitochondria (Ellison and Burton 2008). It is obvious that mitochondria adaptively co-evolved with the nucleus, and numerous biological features are determined by mitochondria.

Cytoplasmic male sterility (CMS) is one of the most ideal phenomena known in angiosperms to describe the incompatibilities between mitochondrial–nuclear genomic interactions.
CMS is a maternally inherited pollen/stamen dysfunction observed in >150 plant species, and it is often caused by cytoplasmic/nuclear substitutions derived from inter- or intra-subspecies backcrossing. It should be noted that in many cases of CMS, no apparent defects in vegetative development are observed. Unlike the mitochondrial–nuclear incompatibility seen in cœpods, CMS is considered to be caused by a gain-of-function mutation in the mitochondrial genome rather than by loss of function. It is well documented that an unusual open reading frame (ORF) consisting of the chimeric structure of endogenous mitochondrial genes in CMS mitochondrial genome is responsible for CMS (Schnable and Wise 1998, Hanson and Bentolila 2004). The proteins encoded by most such ORFs are predicted to be hydrophobic, and several of them have been demonstrated to be associated with mitochondrial inner membranes (Hack et al. 1994, Yamamoto et al. 2005, Yamamoto et al. 2008). These lines of evidence support the dominant nature of CMS, and one might speculate that mitochondrial membrane potential may be essential for proper pollen development, since CMS is likely to be caused by an aberrant protein inserted into the mitochondrial inner membrane. However, the underlying mechanism of male-specific dysfunctions in CMS remains unresolved, although CMS is a unique case of the mitochondrial–nuclear relationship representing the clear epistacy of the mitochondria over the nucleus in pollen development. The analysis of downstream reactions under CMS conditions is obviously necessary to understand the mitochondrial–nuclear incompatibility in CMS.

Rice is one of the ideal plant species for CMS analysis because >60 types of CMS have been reported (for reviews, see Virmani 1994, Kinoshita 1997). For example, BT-type CMS originating from an indica rice variety (Oryza sativa L.) Chinsurah Boro II (Shinjyo 1969, Shinjyo 1975), LD-type CMS from a Burmese variety of indica rice Lead Rice (Watanabe 1971), WA-type CMS from a wild abortive line found in a natural population of the wild rice O. rufipogon Griff. (Lin and Yuan 1980) and CW-type CMS from a Chinese wild rice O. rufipogon Griff. strain W1 (Katsuo and Mizushima 1958) have been identified. W1-type CMS was first discovered from the discrepancy in fertility between reciprocal F1 hybrids of an Indian wild rice strain W11 O. rufipogon (O. sativa L. var. fatua in the original literature; Katsuo and Mizushima 1958), and its phenotypes have not yet been studied in detail prior to this study.

The extent of pollen phenotypes of the rice CMS lines is broad; for instance, there is a CMS line that exhibits damaged microspores without starch accumulation, and, in addition, another CMS line produces visibly normal pollen but with no germination ability on its stigma (for reviews, see Li et al. 2007, Fuji et al. 2008). The phenotypic differences in these lines are a valuable resource, and comparative analysis among cytoplasms could be effective for elucidating the common and distinct features of each CMS type.

To gain further insight into the downstream cellular reactions under CMS cytoplasm, in this study we compared the anther transcriptomic status of five independent rice CMS lines, BT-, CW-, LD-, W11- and WA-type CMS lines, each of which shows distinct pollen abortion phenotypes. We conclude that the phylogenetic relationships of the mitochondrial genomes are strongly correlated with nuclear gene expression patterns, and these data suggest the dependency of pollen development on mitochondrial genotypes. As CMS is often caused as a consequence of successive backcrossing, we hypothesize that there is a genomic barrier between mitochondria and an alien nucleus, and in this manuscript we term this as ‘nuclear–mitochondrial genomic barrier’ or ‘nuclear–cytoplasmic genomic barrier’. The role of cytoplasmic epistasy in gene regulation during the nuclear–mitochondrial genomic barrier in CMS is discussed. It should be noted here that ‘epistasy’ in this case is the dominance in control of nuclear genes by mitochondrial-encoded factors.

**Results**

**Genetic characterization of CMS mitochondrial genomes**

BT-, CW-, LD-, W11- and WA-type CMS lines were chosen for analysis in this study. Each CMS line was backcrossed with an O. sativa L. ssp. japonica cultivar, Taichung 65 (T65), and each was confirmed to possess nuclear chromosomes nearly isogenic to T65 by marker-aided selection. Although BT-, CW-, LD- and WA-CMS lines were completely seed-set sterile, the W11-CMS line retained 24% of its seed-set and was semi-male sterile (Supplementary Fig. S1). Thus W11-CMS was defined as the semi-male sterile phenotype.

To categorize these CMS cytoplasmas by phylogenetic distance of the mitochondrial genomes, restriction fragment length polymorphism (RFLP) of around 76 mitochondrial ORF regions was investigated (Supplementary Fig. S2). Distances among the haplogroups were estimated by calculation based on the rate of identity of the RFLPs (Supplementary Fig. S3), which was detected using combinations of four restriction enzymes and 76 probes of mitochondrial genes (Supplementary Table S1). BT, LD and W11 mitochondrial genomes were clustered into the same groups, showing an average of 67.4% identity among the lines, whereas the WA mitochondrial genome was outgrouped, as only 49.5% was conserved on average compared with the other mitochondrial genomes (Fig. 1A and Supplementary Fig. S3).

A mitochondrial CMS-associated gene, which is responsible for CMS in general, is a chimeric ORF unique to a respective CMS line (Schnable and Wise 1998, Hanson and Bentolila 2004, Fuji et al. 2008). The best candidate for the mitochondrial CMS-associated gene of BT-CMS has been reported to be orf79 located downstream from one of the atp6 genes (Iwabuchi et al. 1993, Akagi et al. 1994, Wang et al. 2006, Kazama et al. 2008). This genomic structure was termed B-atp6-orf79, and we have previously identified a similar structure in the LD mitochondrial genome (Itabashi et al. 2009). From the RFLP
In the CMS lines (Fujii and Toriyama 2005, Itabashi et al. 2009), their phenotypes are described here again for the sake of comparison and restoration in BT, LD and W11 are closely related.

Variation of pollen abortion phenotypes in the CMS lines

The next step towards a transcriptomic study was to distinguish CMS pollen phenotypes morphologically, as it was advantageous to conduct a transcriptomic comparison with tissues without obvious morphological phenotypes to reduce the secondary or tertiary side effects. We have previously reported studies on the pollen development of the BT-, LD- and CW-CMS lines (Fujii and Toriyama 2005, Itabashi et al. 2009). Their phenotypes are described here again for the sake of comparison with the W11-CMS line, in which the phenotype of the male reproductive tissues has not been characterized before. WA-CMS is characterized here for the same reason, although WA-CMS has been known for its pollen abolition at the early microspore stage (Lin and Yuan 1980, Li et al. 2007, Fujii et al. 2008). The morphologies of microspores at the early uninucleate microspore stage in all of the CMS lines were indistinguishable from those of T65 under light microscopy (Fig. 2A). In the WA-CMS line, the majority of microspores were shrunken at the late uninucleate microspore stage and did not develop further without any accumulation of starch (Fig. 2A). Thus, the

analysis, we noticed the presence of a B-atp6-orf79-like structure in the W11 mitochondrial genome (Supplementary Fig. S2, panel 2, arrowheads in the atp6 probe; see Itabashi et al. 2009 for the methods for detecting the B-atp6-orf79-like genomic structure). This region was sequenced, and the insertion of four nucleotides was identified in W11, in the same manner as in LD mitochondria, although the rest of the sequence is identical to that of B-atp6-orf79 in BT mitochondria (Fig. 1B). The presence of a similar B-atp6-orf79 structure in BT-, LD- and W11-CMS lines represents the relative phylogenetic closeness of their mitochondrial genomes, which is in agreement with the fact that the orf79-like structure was not found in WA or CW mitochondrial genomes (Supplementary Fig. S2, panel 2).

From the above findings, we considered the possibility that BT-, LD- and W11-CMS lines share a common mechanism not only for the mitochondrial genomic structure, but also for fertility restoration. As the molecular mechanism of the fertility restorer gene Rf1 for the BT-CMS line has already been identified (Kazama and Toriyama 2003, Akagi et al. 2004, Komori et al. 2004, Wang et al. 2006), the effect of Rf1 on fertility restoration of each CMS line was examined by test-cross experiments. An Rf1-carrying transgenic BT-CMS line (A-28) was constructed in our previous study (Itabashi et al. 2009), and this line retained >90% of its seed-set, indicating that the Rf1 in the transgenic line is fully capable of restoring fertility of a BT-CMS line. The seed setting percentage of the F1 hybrid between the W11-CMS line and A-28 was 89.8%, while that of the W11-CMS line was 24.9%, demonstrating that fertility is remarkably restored by the cloned Rf1 gene, in the same manner as previously demonstrated for restoration of the BT-CMS line and the LD-CMS line by the cloned Rf1 (Kazama et al. 2008, Itabashi et al. 2009). On the other hand, the test cross of A-28 into the WA- or the CW-CMS line did not restore any fertility at all. These results are consistent with the mitochondrial phylogenetic results presented (Supplementary Figs. S2, S3), and it is clear that the molecular features of both CMS induction and restoration in BT, LD and W11 are closely related.

Variation of pollen abortion phenotypes in the CMS lines

The next step towards a transcriptomic study was to distinguish CMS pollen phenotypes morphologically, as it was advantageous to conduct a transcriptomic comparison with tissues without obvious morphological phenotypes to reduce the secondary or tertiary side effects. We have previously reported studies on the pollen development of the BT-, LD- and CW-CMS lines (Fujii and Toriyama 2005, Itabashi et al. 2009). Their phenotypes are described here again for the sake of comparison with the W11-CMS line, in which the phenotype of the male reproductive tissues has not been characterized before. WA-CMS is characterized here for the same reason, although WA-CMS has been known for its pollen abolition at the early microspore stage (Lin and Yuan 1980, Li et al. 2007, Fujii et al. 2008). The morphologies of microspores at the early uninucleate microspore stage in all of the CMS lines were indistinguishable from those of T6S under light microscopy (Fig. 2A). In the WA-CMS line, the majority of microspores were shrunken at the late uninucleate microspore stage and did not develop further without any accumulation of starch (Fig. 2A). Thus, the

Fig. 1 Genotyping of mitochondrial haplogroups. (A) Phylogenetic tree based on percentage identity of RFLPs of the mitochondrial genome. (B) Schematic structure of the atp6-orf79 region in BT-, LD- and W11-CMS lines.

Fig. 2 Morphological appearance of the pollen from the CMS lines. (A) A microspore and pollen observed by light microscopy. (B) Pollen germination on stigmas 6 h after anthesis. Fluorescence indicates a pollen tube stained with aniline blue. Bar = 50 µm.
anther development stages of the WA-CMS line were judged from the developmental stages of their panicles based on the distance between the last two auricles (auricle distance). The morphological pollen development of the other CMS lines was relatively normal until the bicellular pollen stage. Pollen dysfunctions were observed at the tricellular pollen stage in the BT- and LD-CMS lines (Fig. 2A), in which an apparent reduction in starch accumulation was observed, as previously reported (Li et al. 2007, Itabashi et al. 2009). The CW-CMS line retained tricellular pollen without optical dysfunctions as previously reported (Fujii and Toriyama 2005). We observed the pollen grains of each line stained with I$_2$–KI for the observation of starch accumulation (Supplementary Fig. S1). The rate of darkly stained pollen grains, i.e. pollen stainability, is shown in Supplementary Fig. S1. The pollen stainability was 33% for the W11-CMS line, while it was nearly 90% for T65 and the CW-CMS line, nearly 0% for the BT-CMS and the LD-CMS lines, and 0% for the WA-CMS line (Supplementary Fig. S1).

For those CMS lines that did not exhibit complete pollen dysfunction, namely the CW-CMS and W11-CMS lines, pollen germination ability on the stigma was tested (Fig. 2B). Approximately 80% of the pollen on the stigma germinated 6 h after anthesis in wild-type T65 plants. In contrast, none of the pollen in the CW-CMS line germinated, as previously reported (Supplementary Fig. S1), and only 8–36%, with an average of 17%, in the W11-CMS line (Fig. 2B). Thus, the compromised pollen germination ability in the W11-CMS line was able to account for its semi-seed-set sterility (Supplementary Fig. S1).

Microarray analysis

From the above experiment, we reckoned that visible pollen defects are not present in the uninucleate microspore or bicellular pollen stages of the CMS lines, with the exception of the WA-CMS line (Fig. 2). Therefore, we considered that we could compare nuclear gene expression in anthers at the uninucleate microspore or bicellular pollen stages, minimizing the possible secondary effects of pollen morphological changes on gene expression. In total, 10,559 probes, which correspond to 8,199 genes, were significantly altered in their hybridization level (one-way analysis of variance (ANOVA) \(P < 0.01\)) in tissues of any one of the CMS lines in comparison with T65 (Supplementary Table S2). As expected from its failure to develop normal microspores, the largest number of the genes was affected in the WA-CMS line amongst the CMS lines, in which the hybridization levels of 3,348 probes were altered at the uninucleate microspore stage, and 5,333 probes were altered at the bicellular pollen stage.

Hierarchical clustering of the nuclear gene expression profiles of T65 and the CMS lines in each developmental stage outgrouped the WA-CMS line, whereas it intimately grouped the BT-CMS and the LD-CMS lines in the same expression pattern cluster (Supplementary Fig. S4). At the bicellular pollen stage, the W11-CMS line joined the BT and LD expression cluster (Supplementary Fig. S4), and this relationship could have been expected from the phylogenetic closeness of the mitochondrial haplogroups of the three CMS lines (Fig. 1A).

By using non-hierarchical k-means clustering, probes were categorized into 100 clusters based on their hybridization patterns among the CMS lines (Supplementary Table S2). Among the clusters, 51 involved the expression changes in the WA-CMS line (Supplementary Table S2), which are the gene expression alterations that could have been affected by the morphological changes in pollen development. For instance, cluster 17 is a good example, representing the secondary effects of male sterility, in which these cluster genes were highly overexpressed at the uninucleate microspore stage of the WA-CMS line and overexpressed at the bicellular pollen stages of all of the CMS lines except the CW-CMS line (Fig. 3, Supplementary Table S2). Cluster 17 included genes such as peroxidase (LOC_Os01g73170), catalase (LOC_Os02g02400) and glutathione-S-transferase (LOC_Os10g25590) that are directly responsible for reactive oxygen species (ROS) scavenging, the typical genes expressed in stressed conditions. These genes were obviously linked to the extent of pollen deficiency, as displayed in Fig. 2, and quantitatively represent the effects of CMS. Clusters 23, 57, 61, 66, 67, 69, 70, 72, 74, 75, 78, 80, 99 and 100 were characterized by their high expression levels in anthers at both the uninucleate and bicellular pollen stages (Supplementary Table S2), whereas clusters 73, 77, 79 and 98 included genes that were highly expressed only in the anthers at the bicellular pollen stage of the WA-CMS line (Supplementary Table S2). In contrast, clusters 21, 50, 63, 83, 85, 87, 89, 90, 91, 92, 93 and 95 were the genes suppressed in either of the tissues of the WA-CMS line (Supplementary Table S2). Gene expression images of representative clusters are shown in Fig. 3.

Clusters 13, 14 and 19 were the members that were misregulated only in the CW-CMS line (Fig. 3). Clusters 13 and 19 were the genes that were overexpressed in anthers at the bicellular pollen stage of the CW-CMS line. We noticed a curious tendency in the probes listed for cluster 13; namely 54 out of 62 of them corresponded to the genes that were predicted by Predotar (Small et al. 2004) to encode putative plastid proteins, such as chlorophyll a/b-binding proteins (LOC_Os01g41710, LOC_Os01g52240, LOC_Os03g39610, LOC_Os07g37240 and LOC_Os11g13890) and PSI reaction center subunits (LOC_Os03g56670, LOC_Os07g05480, LOC_Os09g30340 and LOC_Os12g08770), and subunits of PSII multicomplexes (LOC_Os08g02630 and LOC_Os07g04840) (Supplementary Table S2). In terms of gene loci after ignoring the redundant probes, 26 out of 32 genes were considered to encode plastid proteins (Supplementary Table S2, cluster 13 highlighted in green, plastid protein-encoding genes in dark green). It was also interesting that cluster 19 contained 11 independent genes encoding heat shock protein (HSP) and molecular chaperones out of 31 genes listed in this gene cluster (Supplementary Table S2, cluster 19 highlighted in yellow, HSP-related genes in dark yellow). Seventeen genes in cluster 14 were the genes suppressed in the bicellular pollen stage of the
Although BT-CMS, LD-CMS and W11-CMS lines have relatively similar mitochondrial genomes and all three of them contained a B-atp6-orf79-like structures (Fig. 1B), we were able to identify nuclear gene clusters by which these CMS cytoplasm could be distinguished. Clusters 9 and 52 were the genes up-regulated only in the W11-CMS line specifically, and the members of cluster 9 were the genes aberrantly up-regulated in the W11-CMS line in anthers at both the uninucleate microspore and the bicellular pollen stages (Fig. 3, Supplementary Table S2, highlighted in blue), whereas cluster 52 was the group in which aberrant expression was evident only in the bicellular pollen stage (Fig. 3, Supplementary Table S2, highlighted in black). Genes in cluster 2, in which only seven were assigned, were strongly up-regulated in the LD-CMS line in anthers at both the uninucleate microspore and bicellular pollen stages (Fig. 3, Supplementary Table S2, highlighted in red). Genes in cluster 59 were only up-regulated in anthers at the bicellular pollen stage of the LD-CMS line (Fig. 3, Supplementary Table S2, highlighted in pink). Lastly, genes up-regulated in the anthers at the uninucleate microspore stage of the BT-CMS line were found in cluster 55 (Fig. 3, Supplementary Table S2, highlighted in purple). Clusters with genes down-regulated specifically in BT-, LD- or W11-CMS lines were not found. Thus, we were able to identify nuclear markers which enabled us to make qualitative distinctions between the cytoplasmic–nuclear gene regulation occurring in each CMS line.

**Mitochondrial gene regulation networks in the CMS lines**

We paid special attention to the 853 probes that corresponded to the genes predicted by Predotar (Small et al. 2004) to encode a putative mitochondrial protein within the 10,559 probes that were significantly altered in their expression levels in any of the CMS lines (Supplementary Table S3). Among the 10,559 probes, three corresponded to Alternative oxidase 1a (AOX1a, LOC_Os04g51150), which we found in our previous study to be overexpressed in the CW-CMS line in a mature anther-specific manner (Fujii et al. 2007). AOX is a gene well known for its mitochondrial stress responsiveness, such as to antimycin A treatment (Djajanegara et al. 2002), and its gene products are also reported to overaccumulate in mitochondrial gene deletion mutants (Karpova et al. 2002, Vidal et al. 2007). Since AOX1a was overexpressed specifically in the mature anthers of the CW-CMS line (Fujii et al. 2007), we have been focusing on AOX1a as the marker gene for rice CMS. Including AOX1a, we categorized the expression patterns of the 853 probes into seven clusters by k-means clustering; the seven clusters are hereafter termed as mt_clusters (Supplementary Table S3). Three probes representing AOX1a were all slotted into mt_cluster 3, and 94 independent genes were included in mt_cluster 3 (Supplementary Table S3, highlighted in blue). Mt_cluster 3 includes the group of genes highly expressed in anthers at the bicellular pollen stage of W11-, LD- and BT-CMS lines, and this group included genes encoding proteins related to various

**Fig. 3** Visualization of relative expression levels of the representative gene clusters. Expression levels are presented as relative log2 values against T65 based on the microarray analysis.
extents to mitochondrial processes, such as acetyl-CoA production from pyruvate (LOC_Os04g32330 and LOC_Os07g49520), the NADH dehydrogenase complex subunit (LOC_Os01g61410 and LOC_Os05g26660) and pentatricopeptide repeat proteins (LOC_Os01g65840 and LOC_Os10g21470). It should be noted here that DCW11, down-regulated gene 11 in CW-type CMS (LOC_Os02g15594), was included in this cluster. DCW11 encodes a mitochondrial protein phosphatase 2C (Fujii and Toriyama 2008).

To understand further the regulation of the members in mt_cluster 3, including AOX1a, we constructed a network of genes that have positive and negative relationships within the cluster. A total of 244 data sets were collected from the Affymetrix probe set in the Gene Expression Omnibus (GEO) database at the NCBI (http://www.ncbi.nlm.nih.gov/geo/), and Pearson correlations between the expression patterns were calculated. Gene expression networks were constructed on the basis of the correlation values of the genes, and connections within mt_cluster 3 were chosen and visualized. Mt_cluster 3 was grouped into three major subclusters with distinct gene expression regulatory behaviors (sub1, sub2 and sub3). AOX1a was included in sub1. Interestingly, sub1 and sub2 were apparently in a negative expressional relationship with each other in terms of their global set of microarrays (Fig. 4, Supplementary Table S3, highlighted in blue). In contrast, sub1, sub2 and sub3 showed identical expression profiles in our study, as shown in Fig. 5, without the negative regulatory correlation relationships between sub1 and sub2 (Fig. 5). These data apparently show that mt_cluster 3 genes in the BT-, LD- and W11-CMS lines are irregularly overexpressed (Fig. 5), without regard to the fact that sub1 and sub2 genes are usually negatively regulated (Fig. 4).

**Discussion**

In this study, we regarded the CMS lines as a series of cytoplasmic substitution lines and sought to determine the characteristics of the CMS phenotype expression in conjunction with the mitochondrial genomic structures and gene expression patterns. Cytoplasmic effects on plant development have been extensively studied in maize (Allen 2005) and wheat (Tsunewaki et al. 1996, Tsunewaki et al. 2002). In maize, 17 cytotypes have been studied from 24 accessions, and various extents of morphological phenotypes have been reported in alloplasmic lines derived from cytoplasmic substitution, such as altered plant height, number of tassel branches and fertility (Allen 2005). In wheat, the cytoplasts of eight Triticum species (10 accessions) and 24 Aegilops species (36 accessions) have been introduced by repeated backcrosses to 12 genotypes of common wheat. The effects of cytoplasm on characteristics such as fertility, heading date and plant height have been investigated (Tsunewaki et al. 1996, Tsunewaki et al. 2002). However, comparative studies on nuclear gene expression have not yet been reported for these alloplasmic lines.

Few studies, including our previous study (Fujii et al. 2007), have analyzed the nuclear gene expression patterns in CMS lines (Carlsson et al. 2007). To the best of our knowledge, the current study presents the first comparison of the nuclear gene expression profiles of the multiple cytoplasmic substitution lines derived from natural variations in any organisms. Despite its clear contribution to sound pollen development as displayed by CMS, the effect of maternally inherited genomic information for plant development has been overlooked from the point of view of its global gene expression.

The details of the downstream reactions, which might be evoked by the presence of CMS-associated protein, have been mainly characterized in alloplasmic CMS lines, derived from interspecific or intergenic crosses of distantly related species.
(Zubko 2004). The wheat APETALA3 homolog WAP3 was suppressed in alloplasmic wheat suffering from the pistillody phenotype (Murai et al. 2002). Linke et al. (2003) found that the MADS box genes _DeMADS2_ and _DeMADS3_, which are orthologous to _Antirrhinum GLOBOSA_ and _DEFICIENS_, were down-regulated in CMS carrot. The APETALA3 gene was aberrantly expressed in CMS _Brassica napus_ (Geddy et al. 2005).

Teixeira et al. (2005) proposed evidence of a retrograde influence of the expression levels of various homeotic genes. These studies present agreeable gene expression features of the alloplasmic CMS lines with a link to its flower organ conversion phenotypes, and it is considered only that a signal(s) emitted from the mitochondria is related to flower morphology. It is obvious that the coordinated action of the nucleus and mitochondria is required for proper plant reproductive development.

First, we clarified the phylogenetic relationships of the CMS mitochondrial genomes to better understand the common or the specific structures in the CMS genomes. Amongst the most noticeable is the phylogenetic (Fig. 1) and transcriptomic (Supplementary Fig. S4) relevance of W11-, BT- and LD-CMS lines. All of these CMS mitochondria possessed _orf79_ (Fig. 1B), which was reported as the best candidate for the mitochondrial BT-CMS causative gene (Iwabuchi et al. 1993, Akagi et al. 1994, Wang et al. 2006, Kazama et al. 2008). Recently, collective research on the _orf79_ gene from the wild relatives of rice has succeeded in identifying nine independent CMS lines with different _orf79_ sequences, with similar types of pollen sterility (Li et al. 2008). Although protein accumulation of ORF79 has not been studied in the research of Li et al. (2008), it would be valuable to know whether _orf79-carrying_ cytoplasm caused common CMS effects. Our present data support _orf79-carrying_ CMS cytoplasm sharing common features, and it is possible that their less starch-accumulating pollen sterility phenotypes (Supplementary Fig. S1) might be caused by similar mechanisms. In practice, the seed-sets of all BT-, LD- and W11-CMS lines were restored by _Rf1_, previously reported as the fertility restorer for BT-CMS. However, a mystery remains regarding the W11-CMS line, which apparently exhibited partial pollen and seed-set sterility phenotypes (Fig. 2, Supplementary Fig. S1). This semi-CMS type might provide a clue for further understanding of CMS, because of its intermediate phenotype. It is possible that genes clustered into clusters 9 and 52 misregulated in a W11-specific manner are involved in this specific semi-CMS mechanism of W11-CMS. For example, one of the cluster 9 genes, mitochondrial NADH dehydrogenase subunit N (LOC_Os01g66000), is overexpressed in anthers of the W11-CMS line (Supplementary Fig. S1), perhaps resulting in some compensation for the mitochondrial defects in the CMS which might be occurring for partial male fertility.

There were also noticeable unique features in the gene expression patterns of the BT-CMS and the LD-CMS lines. One of the interesting gene family members included in cluster 55, overexpressed in the BT-CMS line, was the heavy metal transporter (LOC_Os01g55320, LOC_Os01g61070 and LOC_Os09g09930). It is known that heavy metal treatment induces ROS (Mithofer et al. 2004), and the mitochondrion is the major source of ROS. The downstream reaction of BT-CMS may include effects mimicking heavy metal stress, which can be linked to mitochondrial function by ROS signaling. Cluster 59 was comprised of genes solely up-regulated in the LD-CMS line (Fig. 3), and one of the described genes in this cluster was tRNA synthetase class I (LOC_Os01g03020), which encodes the rice ortholog for Arabidopsis plastid leucine-tRNA ligase (AT4G04350, Duchêne et al. 2005). This gene is one of the _Embryo defective_ genes essential for embryo development (Berg et al. 2005). It is not possible to explain how this gene is misregulated only in the LD-CMS line but, given the plastid–mitochondrial dual-targeting rule for the tRNA synthetase family (Berg et al. 2005), it is likely that specific mitochondrial dysfunction in LD-CMS caused this gene to be overexpressed.

There was also one characterized gene in cluster 2, peptide methionine sulfoxide reductase _msrB_ (LOC_Os05g33510), highly overexpressed in the LD-CMS line (Supplementary Table S2). This enzyme is responsible for the thioredoxin-linked peptide methionine oxidation (EC 1.8.4.12), and Arabidopsis encodes functional methionine sulfoxide reductases (Vieira Dos Santos et al. 2007). It can be speculated that the thioredoxin-linked redox state is altered in LD-CMS, and the gene regulation of the relevant enzyme was altered correspondingly.

On the other hand, unique features of pollen developmental and the molecular phenotypes were observed in the CW- or the WA-CMS lines (Figs. 1–3). For CW-CMS, in addition to its unique non-pollen-germinating CMS phenotype without visible pollen developmental defects, the genes overexpressed in the CMS type were unique. Cluster 19 included 11 independent genes relevant to HSPs, including heat shock factors (HSFs). HSP-encoding genes are often considered as the hallmark of multiple stresses, due to their multiple stress responsiveness (Kotak et al. 2007). The important point in our study is that these HSP genes were solely up-regulated in the CW-CMS line and not in others. This clearly indicates that these HSP genes are responsive only to the CW-CMS cytoplasm background, and CW-CMS can be distinguished from other CMS types by this characteristic. As well as the members of cluster 19, most of the members in cluster 13 were predicted to encode plastid proteins uniquely overexpressed in anthers at the bicellular pollen stage of the CW-CMS line (Fig. 3). There is a well-known signaling process that passes GENOMES UNCOUPLED1 (GUN1), a pentatricopeptide protein, in plant plastid retrograde signaling (Koushevitzky et al. 2007). Under Plastid-stressed conditions, a nucleus-encoded plastid PSI1 subunit gene for light harvesting complex b ( _Lhcb_ ) is suppressed, which gives the idea of coordinated nuclear–plastid expression (Koushevitzky et al. 2007). Including _Lhcb_, genes governed by plastid signaling were shown to possess an ABA-responsive element in their promoter regions. ABSCISIC ACID-INSENSITIVE4 (ABI4), an APETALA2-type transcription factor, was required to suppress the genes upon plastid stress, and it has been described that GUN1 can be upstream of ABI4 (Koushevitzky et al. 2007). We were curious as to whether our genes slotted into cluster 13...
carry ABA-responsive elements at their promoter sequences; thus, the abundant nucleotide motifs were searched using the Element program (Nemhauser et al. 2004). As a result, 28 out of 32 non-redundant genes that belong to cluster 13 were shown to possess at least one ABA-responsive element (ACGTG) in their promoter regions. It would be possible, if the GUN1–ABI4 system was also conserved in rice pollen, for the CW-CMS system to have a deficiency in the GUN1–ABI4 system to suppress the plastid genes in cluster 13.

It is also noteworthy that HSFs (included in cluster 19) and plastid genes (included in cluster 13) were both found to be overexpressed in a rice tapetum degradation retardation (tdr) mutant (Zhang et al. 2008). The tapetum cell layer is known to experience programmed cell death (PCD) during pollen maturation to provide lipid compounds to coat the pollen. In addition, premature PCD is implicated in sunflower CMS (Balk and Leaver 2001). tdr appeared to be the PCD mutant (Li et al. 2006, Zhang et al. 2008), and HSF is known as the anti-apoptotic factor (Yamanouchi et al. 2002). Considering that CW-CMS is restored by a fertility restorer gene which functions in a gametophytic manner (Fujii and Toriyama 2005), it is unlikely that tapetum dysfunction is the cause of CW-CMS. The CW-CMS line might be deficient in some PCD processes usually also taking place in mature pollen grains and progressively in vegetative cells (Varnier et al. 2005), and vegetative cell dysfunction might be a reasonable explanation for its non-pollen germinating phenotype (Fig. 2B). Meanwhile, the TDR gene itself (LOC_Os02g02820, Table S2, cluster 63) was suppressed in the WA-CMS line. Sporophytic CMS including sunflower PET-CMS may be explained by the early degeneration of tapetum by premature PCD (Balk and Leaver 2001), and likewise the WA-CMS line might suffer similar defects. Furthermore, with relevance to PCD, the GAMYB gene (LOC_Os03g38210, Table S2, cluster 30) was suppressed in most of the CMS lines. GAMYB was originally described as being involved in gibberellin signal transduction, but moreover a recent study concluded its relevance in tapetum PCD (Aya et al. 2009). GAMYB was down-regulated in most of the CMS lines (Supplementary Table S2), suggesting that defects in gibberellin-mediated signal transduction can be defective in CMS and PCD as a consequence.

Recently, ABI4 has also been shown to be the candidate regulon for AOX1α suppression, as the abi4 mutant was incapable of repressing AOX1α (Giraud et al. 2009). Based on this idea, we re-analyzed the co-regulation network of AOX1α including the genes for plastid protein present in cluster 13 (Supplementary Fig. S5). As a result, cluster 13 genes were shown to be positively related to sub1 genes as shown in Fig. 4, indirectly suggesting that there is a retrograde suppression mechanism of AOX1α and plastid genes bypassing ABI4 in rice, similar to that of Arabidopsis. AOX1α was highly overexpressed in anthers at the bicellular pollen stage of BT-, LD- and W11-CMS (Fig. 3). In the global sets of microarrays extracted from GEO, the gene clusters co-expressed with AOX1α in the CMS lines (mt_cluster 3, Supplementary Table S3) were clustered into three subgroups (sub1, sub2 and sub3, Fig. 5).

Although sub3 displayed a correlation with both sub1 and sub2, sub1 and sub2 were negatively regulated in the majority of the microarray sets, suggesting that even within mitochondrial genes there are specific expression regulatory mechanisms (Fig. 5). Sub1 included genes involved in mitochondrial metabolism, e.g. adenylate kinase involved in purine biosynthesis (LOC_Os04g01530), the 2-oxoglutarate dehydrogenase complex (LOC_Os04g32330 and LOC_Os06g01630) and dihydroxy- poyl dehydrogenase included in the pyruvate dehydrogenase complex (LOC_Os05g06750). On the other hand, sub3 was comprised of genes that are related to protein–protein interaction, or signaling processes, such as protein phosphatase 2C (DCW11, LOC_Os02g15594), P21-Rho-binding domain-containing protein (LOC_Os07g26480) and GTPase-activating protein (LOC_Os07g31830). Although the epistasis between each subgroup cannot be determined, there seems to be an interesting conflicting regulatory rule between these subgroups, and the effect of BT-, LD- and W11-CMS could be breaking the rules of the relationships among sub1, sub2 and sub3, and overexpressing everything as shown in Fig. 4. From these data we can hypothesize that the effects of CMS-inducing cytoplasm are epistatic to nuclear genotypes. Such cytoplasmic epistasy has also been reported in Drosophila: a mitochondrial genotype is one of the critical determinants of the Drosophila lifespan and is epistatic to nuclear genotypes in some cases of interspecific hybrids (Rand 2005, Rand et al. 2006). Our finding here is a good example of cytoplasmic epistasy over nuclear gene expression in plants, suggesting that there is a positive nuclear gene regulation governed by cytoplasmic genomes during pollen development.

One of the best molecular explanations for cytoplasmic epistasy might be cytoplasm to nucleus retrograde signaling, which is the opposite of anterograde signaling which could be described as the nucleus as the central governing authority of the plastid and mitochondrial genomes. Retrograde signaling has been exclusively studied in yeast and animals, and a number of intermediary factors have been identified (Butow and Avadhanil 2004). These factors were identified as molecules involved in the mitochondrial stress response, and it is not known if a similar mechanism exists in retrograde signaling upon cytoplasmic substitution. Except for ABI4 being related to AOX1α regulation as mentioned above (Giraud et al. 2009), no other retrograde regulatory components in plant mitochondria have been identified. As we categorized our cytoplasmic effects into three types based on our microarray study, (i) BT, LD and W11 type; and (ii) CW type; and (iii) WA type, it is apparent that multiple pathways of retrograde signaling are involved in mechanisms of rice CMS occurrence. A recent study of Arabidopsis mitochondrial stress signaling predicted three independent pathways for up-regulation of AOX (Ho et al. 2008).

Our current study revealed the relevance of mitotypes to the pollen developmental and molecular phenotypes of CMS. From this study, it should be emphasized that maternally inherited genomic information contributes quite importantly...
to plant development in terms of molecular regulation, and this study is one of the frontier research studies that focuses on gene regulation dependent on cytoplasmic genotype.

Materials and Methods

Plant materials

The origin of the BT-CMS, LD-CMS and CW-CMS lines was described previously (Kazama et al. 2008, Fujii and Toriyama 2009, Itabashi et al. 2009). A japonica cultivar ‘Taihung 65 (T65)’ was used as the maintainer line. The WA-CMS line was obtained by backcrossing a commercial hybrid rice cultivar ‘Wei You 6’ (reviewed in Virmani 1994) with a japonica cultivar ‘Reimei’ five times followed by T65 twice. The W11-CMS line was obtained by backcrossing O. rufipogon Griff. strain W11 with ‘Reimei’ six times followed by T65 twice. The rate of nuclear substitution by the T65 genome, which was determined using 63 simple sequence repeat (SSR) markers, was 100% for the BT-CMS line, 91.2% for the CW-CMS line, 85.6% for the WA- and the W11-CMS lines, and 82.4% for the LD-CMS line (data not shown). The rest of the genome consisted of a japonica cultivar ‘Avihikari’ for the LD-CMS line and ‘Reimei’ for the WA- and the W11-CMS lines. These five CMS lines, therefore, are near isogenic lines with the nuclear genome of T65.

Estimating the distance of the mitochondrial haplogroups

The extracted mitochondrial genomic DNA of each CMS line was digested with BamHI, EcoRI, HindIII or XhoI, and RFLP was detected by Southern blot analysis. The probes used were 76 mitochondrial genes generated by PCR using the primers listed in Supplementary Table S4. Southern blotting was performed according to our previous study (Fujii et al. 2007). The percentage identity of the mitochondrial genomes was calculated using the combinations of the four restriction enzymes and 76 probes, namely 308 combinations (Supplementary Table S1). A band pattern of RFLP was treated as one character, and genetic distances among the haplogroups were calculated based on their percentage identity. A phylogenetic tree was constructed using a Neighbor–Joining method (Saitou and Nei 1987).

Microscopic observation of pollen grains

Observation of pollen under light microscopy, pollen germination and seed setting was carried out as previously described (Fujii and Toriyama 2009, Itabashi et al. 2009).

Microarray analysis

Total RNA was extracted from anthers at the uninucleate microspore and bicellular pollen stages using an RNeasy Plant Mini Kit (Qiagen, Tokyo, Japan) and subjected to microarray analysis using the 44K Rice Gene Expression Microarray and One-color Quick Amp Labeling Kit (Agilent Technologies, Santa Clara, CA, USA). The anther development stages were judged from the distance between the last two auricles (auricle distance). An auricle distance of 8–9 cm corresponded to the uninucleate microspore stage and a distance of 12–15 cm corresponded to the bicellular pollen stage. Hybridization and detection were carried out according to the manufacturer’s instructions with a G2545A hybridization oven (Agilent) and G2565BA DNA microarray scanner (Agilent). The scanned tiff image files were analyzed with FeatureExtraction 9.5.3.1 (Agilent). All the stages/lines combinations had three biological replicates. Original transcriptome data generated in this study are deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GEO: GSE18057.

Statistical analysis and gene clustering

All of the normalization, statistical analysis, sample clustering and gene clustering were performed using the R v 2.9.1 program (R Development Core Team 2009). The raw expression value obtained from the microarray experiments was normalized by quantile normalization according to Jeffery et al. (2006). The expression value was defined as the average of three independent hybridizations for each of the CMS lines in each tissue, and statistically significant changes in expression were evaluated by one-way ANOVA (P < 0.01) as described in Norris et al. (2005). After Z-scaling, hierarchical clustering of the expression patterns among the samples was done using the pvclust package in the R library with a bootstrap value of 100. Non-hierarchical k-means clustering was performed after conversion into log2 scale, using the cclust package in the R library. The number of clusters (100) for non-hierarchical clustering had a sufficient amount of cluster validity (Davies-Bouldin Index 4.43, Supplementary Fig. S6) compared with other cluster numbers, and therefore ‘100’ was chosen as the round number. The expression images were created using Multiexperimental Viewer v 4.3 (Saeed et al. 2005).

Gene network construction in rice

We constructed a gene expression network to classify expression patterns of genes in Mt_cluster3. To examine gene expression patterns under various biological conditions, we used publicly available expression data from 244 samples, which were obtained by GeneChip Rice Genome Array (Affymetrix) in the GEO database at the NCBI (http://www.ncbi.nlm.nih.gov/geo/). The expression data were normalized by the Z-score transformation method. For each pair of probes, pairwise Pearson’s correlation coefficients were calculated. Since the GeneChip probe names were assigned with identifiers from the Rice Genome Annotation database (http://rice.plantbiology.msu.edu/), identifiers (RAP loci) of 111 genes included in Mt_cluster3 (Supplementary Table S2) and cluster 13 (Supplementary Table S1) were converted into MSU identifiers according to the ‘ID converter table’ (http://rapdb.dna.affrc.go.jp/rapdownload/rap2/final_osid2tigr.tbl.gz). For the 105 GeneChip probes corresponding to the 111 genes, the expression network was constructed on the basis of correlation.
coefficients, where positive and negative correlation coefficients indicated similar and inverse expression profiles.

**Supplementary data**

Supplementary data are available at PCP online.

**Funding**

This study was supported by the Ministry of Education, Science, Sports and Culture, Japan [Grant-in-Aid for Special Research on Priority Areas (No. 18075002)]; the Japan Society for the Promotion of Science for Young Scientists [Research Fellowships to S.F. and E.I.].

**References**


