A lineage-specific parologue of Oma1 evolved into a gene family from which a suppressor of male sterility-inducing mitochondria emerged in plants

Takumi Arakawa 1, 2, Hiroyo Kagami 1, Takaya Katsuyama 1, Kazuyoshi Kitazaki 1, Tomohiko Kubo 1,*

1 Research Faculty of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-ku, Sapporo 060-8589, Japan
2 Gifu Prefectural Research Institute for Agricultural Technology in Hilly and Mountainous Areas, Nakatsugawa, Gifu, 508-0203, Japan

* Corresponding author: Tomohiko Kubo, Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan, tel/fax +81-11-706-2484, tomohiko@abs.agr.hokudai.ac.jp
Abstract

Cytoplasmic male sterility in plants is caused by male sterility-inducing mitochondria, which have emerged frequently during plant evolution. Nuclear Restorer-of-fertility (Rf) genes can suppress their cognate male sterility-inducing mitochondria. Whereas many Rfs encode a class of RNA binding protein, the sugar beet (Caryophyllales) Rf encodes a protein resembling Oma1, which is involved in the quality control of mitochondria. In this study we investigated the molecular evolution of Oma1 homologues in plants. We analyzed 37 plant genomes and concluded that a single copy is the ancestral state in Caryophyllales. Among the sugar beet Oma1 homologues, the orthologous copy is located in a syntenic region that is preserved in Arabidopsis thaliana. The sugar beet Rf is a complex locus consisting of a small Oma1 homologue family (RF-Oma1 family) unique to sugar beet. The gene arrangement in the vicinity of the locus is seen in some but not all Caryophyllalean plants and is absent from A. thaliana. This suggests a segmental duplication rather than a whole genome duplication as the mechanism of RF-Oma1 evolution. Among the positively selected codons in RF-Oma1, many are located in predicted transmembrane helices. Phylogenetic network analysis indicated that homologous recombination among the RF-Oma1 members played an important role to generate protein activity related to suppression. Together, our data illustrate how an evolutionarily young Rf has emerged from a lineage-specific paralogue. Interestingly, several evolutionary features are shared with the RNA binding protein type Rfs. Hence, the evolution of the sugar beet Rf is representative of Rf evolution in general.

Keywords

Cytoplasmic male sterility, nuclear-mitochondrial interaction, plant mitochondria, positive selection, restorer-of-fertility, sugar beet

Significance statement

Plant mitochondria sometimes evolve to induce male sterility because the male gamete is not essential for the maternal inheritance of the mitochondria, and this has driven the evolution of nuclear genes that restore male fertility (Rfs). We found that, although the sugar beet Rf belongs to a unique gene family that differs from most other plant Rfs, many of its molecular evolutionary features are strikingly similar to those of other Rfs. Such similarity indicates a common evolutionary mechanism associated with the plant genome’s ability to overcome the male sterility caused by selfish mitochondria.
Introduction

The endosymbiotic theory proposes that the mitochondrion originated as an endosymbiont resembling an alpha-proteobacterium (Scheffler 2007), and is supported by molecular phylogenetic analyses of mitochondrial DNA (Lynch, 2007). The mitochondrial DNA (mtDNA) is, however, not only a remnant of the ancestral genome but contains genes important for the eukaryotic host. Although the number of mitochondrial genes is small (e.g. 37 in mammals and 50–60 in flowering plants) (Gualberto and Newton 2017), their impact on the host’s phenotype and fitness cannot be ignored (Wallace 2018).

In many cases, the mtDNA is inherited maternally, unlike the nuclear DNA (Perlman et al. 2015). According to evolutionary theory, this situation invokes differential interests between the mitochondrial and nuclear DNAs: for the mtDNA, mutations that reduce the number of male gametes can be beneficial if they render some advantage to female fitness. However, such mutations are deleterious for inheritance of the nuclear DNA, and thus they drive the evolution of nuclear genes that suppress the male-harming mitochondrial mutations (Touzet 2012; Rice 2013). This situation is called mitonuclear conflict (Havird et al. 2019).

Cytoplasmic male sterility (CMS) in flowering plants offers a good example of the mitonuclear conflict (Havird et al. 2019). Some flowering plant mitochondria are known to induce male sterility (MS) that inhibits pollen production but does not affect other organs (Schnable and Wise 1998). The MS-inducing mitochondria produce unique proteins that are absent in non-inducing mitochondria (Hanson and Bentolila 2004; Kazama et al. 2019), and the genes encoding such unique proteins have been identified in multiple plant species (Chen and Liu 2014). These gene sequences are composed of fragments of duplicated mitochondrial genes and/or sequences with unknown origins, and their compositions vary, indicating that their occurrences are evolutionarily independent. Considering that CMS has been reported in more than 140 species (Laser and Lersten 1972), it appears that MS-inducing mitochondria have evolved independently and frequently in flowering plants.

A nuclear-encoded suppressor of CMS is referred to a restorer-of-fertility (Rf) gene, and generally, a dominant Rf allele suppresses the MS phenotype (Chase 2007). Genetic analyses suggest that each Rf gene is a ‘specialist’ that can suppress a specific MS-inducing mitochondrion but not others (Duvick 1965). The repeated evolution of MS-inducing mitochondria and their cognate Rf genes constitutes an intra-genomic arms race (Touzet and Budar 2004), and their evolutionary mechanisms can provide key insights into the mitonuclear conflict.

Rf genes have been identified and analyzed in many plant species, and they encode several different types of proteins (Kubo et al. 2020). The most predominant type is pentatrico peptide repeat.
(PPR) proteins, which bind RNA and recognize specific sequences using arrays of ~35 amino acid motifs, each of which has high affinity for one of the four nucleotide residues (Barkan and Small 2014). RF proteins belonging to the PPR class are imported into mitochondria and participate in post-transcriptional mechanisms to repress the expression of MS-inducing genes (Dahan and Mireau 2013; Gaborieau et al. 2016). The PPR-encoding Rs are members of the Restorer-of-fertility-like (RFL) gene family, which is a subset of the larger PPR gene family (Fujii et al. 2011). The RFL genes appear to have diverged from a single origin that was established before the split of the major flowering plant lineages, and they are maintained in flowering plant genomes as clusters at one or several loci (Geddy and Brown 2007; Fujii et al. 2011). The organizational diversity of the RFL family within species is well known (e.g. Melonek et al. 2019). Molecular phylogenetic studies have revealed that members of the RFL family form species-specific clusters that are paralogous to each other and exhibit signatures of positive selection, features that are reminiscent of the evolutionary patterns of pathogen resistance (R) genes (Geddy and Brown 2007; Fujii et al. 2011). These observations support the concept of an arms race in the mitonuclear conflict (Fujii et al. 2011). On the other hand, the selection patterns of the other type of Rf genes (i.e. non-PPR Rfs) have not been examined in such detail, and it is not known if they also exhibit the signature of an arms race.

In sugar beet (Beta vulgaris), several different types of MS-inducing mitochondria have been reported, and among these, the Owen, I-12CMS(3)/E, and G types have been characterized at the molecular level (Kitazaki et al. 2015; Yamamoto et al. 2008; Meyer et al. 2018). The molecular actions of their cognate Rfs are unclear except for that of the Owen type. A protein unique to the Owen-type mitochondria is encoded by the gene preSatp6 (whose origin is unknown), and the preSATP6 protein is the target of the cognate Rf (Kitazaki et al. 2015). Interestingly, a post-translational mechanism is involved in this interaction, in contrast with the post-transcriptional mechanisms associated with the PPR-encoding Rs (Kitazaki et al. 2015) (see below).

Oma1 was first identified in yeast as a gene encoding a protease (overlapping activity with m-AAA protease) (Käser et al. 2003), but now its various roles in the quality control of mitochondria are well known in animals and plants (e.g. Guo et al. 2020 and Migdal et al. 2017). In sugar beet, a duplicated Oma1 copy evolved into an Rf, which was found by our positional cloning of a locus designated as Rf1 (Matsuhira et al. 2012; Arakawa et al. 2019a). The sugar beet Rf1 locus is in fact a complex locus with multiple clustered Oma1-like genes that participate in fertility restoration, with each gene contributing differently. Hence Rf1 (in terms of a genetic factor identified by a classical genetic approach) does not necessarily correspond to a single open reading frame (T. Arakawa et al., manuscript...
submitted). We therefore designated the clustered Oma1-like genes as the RF-Oma1 gene family, after the Oma1-like gene in the Rf1 locus. We have observed molecular diversity within the Rf1 locus among sugar beet lines: the copy number of clustered RF-Oma1 genes varies from one to four, and the nucleotide sequences of the genes have diverged (Moritani et al. 2013; Arakawa et al. 2019b). Our genetic analyses have identified different Rf1 alleles with variations in dominance, including dominant, semi-dominant, hypomorphic, and recessive (Arakawa et al. 2018; Arakawa et al. 2019b). These differences can be explained by the suite of RF-Oma1 genes clustered in each Rf1 allele. Namely, some RF-Oma1 genes encode proteins that can bind the preSATP6 protein of sugar beet to alter its higher order structure (Kitazaki et al. 2015), and the total amount of transcript from these genes is one of the determinants of the allele’s dominance (T. Arakawa et al., manuscript submitted). This molecular chaperone-like activity toward preSATP6 is unique to some RF-Oma1 genes and is absent from other Oma1 genes such as sugar beet bvOma1 and Arabidopsis atOma1 (Arakawa et al. 2019a). This suggests the evolution of a novel function within the RF-Oma1 gene family (Arakawa et al. 2019a). Some RF-Oma1 genes are transcribed but their protein products are unable to bind the preSATP6 protein (Kitazaki et al. 2015). If the locus consists of such genes, the allele is recessive. Interestingly, no sugar beet line examined to date has lost all its RF-Oma1 genes, even when the line is judged as rfl homozygous recessive by a genetic analysis (Ohgami et al. 2016). If the locus has genes that encode both binding and non-binding products, the allele exhibits (strong or weak, depending on the expression level) restoration ability. The strongest allele that we identified had four RF-Oma1 genes, each with the binding activity (T. Arakawa et al., manuscript submitted). We have identified a total of twelve RF-Oma1 genes that comprise six Rf1 alleles, and clarified their binding activity (Arakawa et al. 2018, 2019b; T. Arakawa et al., manuscript submitted).

In addition to bvOma1 and the RF-Oma1 family, there are other Oma1 homologous genes, LOC104888056 and LOC104906603, but they exhibit presence/absence polymorphisms in the beet genetic resources (Arakawa et al. 2019a) and their functional significance is unknown.

The objective of this study is to detail the molecular evolution of the RF-Oma1 genes. Here we show that the RF-Oma1 genes are paralogues of Oma1 that have evolved relatively recently. There are similarities in the patterns of molecular diversity between the RF-Oma1 and RFL families, and in contrast, the evolutionary pattern of the RF-Oma1 genes differs from that of other Oma1 genes such as bvOma1. This suggests that similar mechanisms were involved in the evolution of the RF-Oma1 and RFL families, even though their gene products are different. The evolutionary pattern of the RF-Oma1 family suggests that paralogue genes played significant roles in the foundation of Rf, and this has been
less stressed in the case of the RFL family. We previously showed that RF-Oma1 gene expression is predominant in the anther, unlike the expression pattern of bvOma1 (Arakawa et al. 2019a). It is interesting to note that some features of the RF-Oma1 family (such as specific expression in male reproductive organs, paralogues encoding mitochondrial proteins, and many positively selected codons) seem to be shared with animal genes that counteract male-harming mitochondrial variants (Havird et al. 2019).

Materials and methods

Nucleotide sequence analysis

Homology searches were performed using the databases at NCBI (https://www.ncbi.nlm.nih.gov/), Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), miyakogusa.jp (http://www.kazusa.or.jp/lotus/index.html), and CuGenDB (http://cucurbitgenomics.org/organism/1). The transmembrane helix was predicted using the TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Emanuelsson et al. 2007). Reference sequences are summarized in Supplementary Table S1. Gene maps were drawn using Easyfig (Sullivan et al. 2011) and Microsoft PowerPoint.

Phylogenetic analysis

Phylogenetic trees were drawn using MEGA7 (Kumar et al. 2016) based on amino acid sequences aligned by ClustalW (Supplementary Dataset 1). The partial-deletion option was used with the site coverage cutoff set to 90%. Phylogeny was tested using 1000 bootstrap replications. Evolutionary distances were determined using the neighbor-joining method and estimated by the Poisson correction distance, with the assumption that the amino acid substitution rate varied among sites according to the gamma distribution (shape number: 5). The maximum-likelihood tree was constructed under the Jones-Taylor-Thornton model assuming the gamma distribution for variation in the amino acid substitution rate (number of discrete parameters: 5). For the tree inference option, the initial tree was made automatically with default settings. The maximum-likelihood heuristic method was the Nearest-Neighbor-Interchange.

Selection pattern analysis

Codons were aligned using PAL2NAL v. 14 (Suyama et al. 2006). The ratios of synonymous (dS) and nonsynonymous (dN) nucleotide substitution rates were estimated using PAML v. 4.9h (Yang 2007). A
A comparison of one-ratio and two-ratio models was performed using the branch model of codeml implemented in PAML. Variations in the dN/dS ratios among sites were examined by comparing five evolutionary models provided by the codeml: M0 (single dN/dS ratio for all sites), M1a (dN/dS is < 1 in some sites and = 1 in the others), M2a (dN/dS of each site is either < 1, = 1 or > 1), M7 (dN/dS ratios of codons follow a beta distribution and are < 1 or =1), and M8 (a certain portion of sites evolve as M7 but the dN/dS ratios of the others are > 1). Posterior probabilities for the site classes were calculated using the Bayes Empirical Bayes method (Yang et al. 2005).

Nucleotide sequencing

The *B. vulgaris* accessions used in this study are summarized in Supplementary Table S2. Plants were grown in the field at the Field Science Center for Northern Biosphere, Hokkaido University. Total cellular RNA was isolated from fresh green leaves using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). After treating with RNase-free DNase I (Takara Bio, Kusatsu, Japan), RNA samples were reverse transcribed using the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA, USA) and an oligo dT primer. PCR was then used to amplify the *bvOma1* sequences with the primers 5'-GTAAAACGACGGCCAGTAGCATATCTTCTCCTTCAAATATCATG-3' and 5'-GGAAACAGCTATGACCATGCGAAGTTGCAAAGAAGATGATGTTAATGGTCA-3'. The nucleotide sequences of the RT-PCR products were determined by Sanger sequencing in an ABI3130 Genetic Analyzer (Applied Bio systems, Foster City, CA, USA). The sequence data are available at the DDBJ/EMBL/GenBank database with the accession numbers shown in Supplementary Table S2.

Network analysis

Nucleotide sequences were aligned using ClustalW, implemented in the MEGA7 software package (Supplementary Dataset 2). The phylogenetic network was constructed using the SplitsTree software v. 4.14.8 (Huson and Bryant 2006) with the Neighbor-Net method (Bryant 2003).

Results

*Oma1* homologues in flowering plants

The *bvOmal* and *RF-Oma1* genes are ubiquitous in sugar beet (Arakawa et al. 2019a), but it was unknown whether multiple *Oma1* copies are usual in other flowering plants. Therefore, we searched for *Oma1* homologues in other flowering plant genomes. The amino acid sequence of atOMA1 (TAIR accession At5g51740.1) was submitted as a query in tBLASTN searches of the publicly available
databases (see Materials and methods). The identified sequences were translated in silico according to the annotated gene models. The obtained amino acid sequences were then compared with that of atOMA1 using BLASTP, to confirm their homology.

Table 1 shows the results for 37 flowering plant genomes from 16 orders. Included are *Amborella trichopoda* (the basal lineage of flowering plants), three monocots, eleven superasterids including four Caryophyllales genomes, and 22 superrosids, including one Saxifragales genome (detailed in Supplementary Table S1). In 24 genomes from eight orders, the copy number of Oma1 homologues is one, and the other 13 genomes have multiple copies (Table 1). The highest number of copies is eight in *Medicago truncatula*, and the other two Fabales genomes have two copies (Table 1). Five orders contain at least one single copy genome and at least one multicopy genome.

In the genomes with multiple Oma1 copies, the sequence homology with atOMA1 varies among the copies. For example, for the three Oma1 homologues in *Fragaria vesca*, their E-values (BLSTP) are 1.00E-175, 3.00E-43 and 2.00E-74 (Supplementary Table S1). In rice, a second Oma1 copy that was overlooked by Matsuhira et al. (2012) has an E-value of 2.00E-19, which is higher than the E-value for the other copy (5.00E-141) (Supplementary Table S1). We divided the plant Oma1 homologues into conserved and diverged types based on their homology with atOMA1, using 5.00E-141 as the threshold to discriminate between the two types. Accordingly, all but one genome contains at least one conserved Oma1, and 29 of the 36 genomes have only one copy of the conserved type. Six genomes (*Gossypium raimondii*, *M. truncatula*, Glycine max, Lactuca sativa, Sesamum indicum, and Chenopodium quinoa) have two conserved types (Table 1). On the other hand, both of the two copies in *Lotus japonicus* are divergent based on our threshold criterion (1.00E-135 and 1.00E-107) (Supplementary Table S1).

Among the four groups of Oma1 homologues in sugar beet (*bvOma1*, the RF-Oma1 family, LOC104888056, and LOC104906603), only bvOma1 is of the conserved type based on our present threshold (1.00E-179), and the remaining three groups are diverged (4.00E-109 to 3.00E-33) (Supplementary Table S1). All the other Oma1 homologues in the Caryophyllalean genomes that we investigated are conserved, including the two Oma1 copies in *C. quinoa* (2.00E-177 for each copy). Our data suggest that the conserved type of Oma1 is nearly ubiquitous in flowering plant genomes, and that the diverged copies sporadically occur in a lineage specific manner.

**Phylogenetic relationships between bvOma1 and the RF-Oma1 genes**

From the results in the previous section, we inferred that the RF-Oma1 genes evolved from a lineage-
specific duplicate of Oma1 rather than from an earlier duplicate, such as in the common ancestor of flowering plants (note that the RFL genes evolved before the split of monocot and dicot lineages; Fujii et al. 2011). We next investigated the phylogenetic relationships between the RF-Oma1 genes and other Oma1 homologues from flowering plants. From the twelve RF-Oma1 genes identified to date, we selected two, one from a dominant and one from a recessive allele (orf20\textsubscript{NK-198-2} and orf20\textsubscript{TK-81}, respectively). The protein product of orf20\textsubscript{NK-198-2} can bind the preSATP6 protein, whereas the product of orf20\textsubscript{TK-81} does not (Kitazaki et al. 2015). Our analysis also included bvOma1, LOC104888056, and LOC104906603. A total of 46 amino acid sequences, including the five sequences from sugar beet, 40 Oma1 copies of the conserved type from the other genomes, and the Marchantia polymorpha Oma1 (as an outgroup) were subjected to phylogenetic analysis.

In both the neighbor-joining and maximum-likelihood trees, the Oma1 homologues form order-specific clusters (Figures 1A and B), although the superrosids and superasterids are not as clearly separated in the neighbor-joining tree. The bootstrap values tend to be lower in nodes closer to the root.

The orf20\textsubscript{NK-198-2} and orf20\textsubscript{TK-81} genes are clustered (hereafter referred to as the RF-Oma1 clade) in both trees within the Caryophyllales clade, suggesting that they are duplicates specific to this order (or the lower taxonomic class). In both trees, the internal branches of the RF-Oma1 clade are longer than the branches separating the other Oma1 genes (Figures 1A and B). In the neighbor-joining tree, the sugar beet Oma1 homologues form a species-specific clade with the genes and RF-Oma1 clade branching in the following order: first bvOma1, then LOC104906603, LOC104888056, and the RF-Oma1 clade (Figure 1A). In the maximum-likelihood tree, LOC104888056 and the RF-Oma1 clade form a sister group within a larger group containing all the other Caryophyllalean Oma1 genes except that of Amaranthus hypochondriacus (Figure 1B).

Gene arrangements associated with bvOma1 and the RF-Oma1 family

Among the B. vulgaris Oma1 homologues, bvOma1 is the likely orthologue of atOma1. In such cases, genes linked with the orthologue would exhibit synteny. We investigated whether the genomic region surrounding bvOma1 is conserved in the A. thaliana genome.

In the sugar beet reference genome, a total of 30 genes, including bvOma1, span a chromosomal region of ~487 kbp (Supplementary Table S3). This region, between LOC104906573 and LOC104906596, is illustrated in Figure 2. We used the amino acid sequences of the 30 genes as queries to find homologous genes in the reference genome of A. thaliana, and found twelve homologues, including atOma1, clustered on chromosome 5 (Supplementary Table S3 and Figure 2). The order and
orientations of the homologues in *A. thaliana* are consistent with those in *B. vulgaris* except for *atOma1*, which is inversed. (Figure 2).

We conducted the same analysis on other Caryophyllalean genomes (Figure 2). The gene arrangement found in *B. vulgaris* is well conserved in *S. oleracea*. We found two syntenic regions in the *C. quinoa* genome, which may be associated with the polyploidization of this species (see Discussion). *A. hypochondriacus* has a different arrangement, with a similar syntenic region on Scaffold_7 but with no *Oma1* homologue (Figure 2). In *A. hypochondriacus*, the *Oma1* homologue is located on Scaffold_13, where only two genes are common to the syntenic region (Figure 2). We think this gene arrangement is a derived morph (see Discussion).

Although the *RF-Oma1* genes are unique to *B. vulgaris*, it is possible that other plants have DNA regions with similar gene arrangements to that of the *RF-Oma1* region. We selected 22 sugar beet genes linked to the *RF-Oma1* genes as queries (from LOC104888042 to LOC104888100 in Supplementary Table S4 and Figure 3), and used them to investigate the *S. oleracea* genome. In *B. vulgaris* strain KWS2320, which has only one *RF-Oma1* copy, the genes LOC104888049, LOC104888050, *RF-Oma1*, and LOC104888089 are clustered together. A similar gene arrangement is also seen in an *S. oleracea* contig, but the *RF-Oma1* (or *Oma1*-like) gene is missing (Supplementary Table S4 and Figure 3). Instead, three other genes (marked by an asterisk in Figure 3) are located between the LOC104888050 and LOC104888089 homologues in *S. oleracea*, but their gene products are annotated as abscisic stress-ripening protein 3-like, long non-coding (Inc) RNA, and lysine-rich arabinogalactan protein 19-like, respectively. None of these genes are related to *Oma1*.

In *C. quinoa*, the contig NW_018742205 contains a similar gene arrangement to the *RF-Oma1* region (Supplementary Table S4 and Figure 3). As in *S. oleracea*, the genes homologous to *B. vulgaris* LOC104888049, LOC104888050, and LOC104888089 are clustered together but a different gene is situated in the position corresponding to that of the *RF-Oma1* gene (marked by a plus sign in Figure 3). Although it is annotated as an IncRNA, its nucleotide sequence is not closely related to the *S. oleracea* counterpart in the same position. Another *C. quinoa* contig, NW_018743021, contains a much shorter syntenic region containing five genes homologous to genes located near *RF-Oma1* in *B. vulgaris* (Supplementary Table S4 and Figure 3).

We found no gene arrangement homologous to the *B. vulgaris* region LOC104888049-LOC104888050-LOC104888089 in *A. hypochondriacus* or *A. thaliana* (Table S5). Both the *bvOma1* and *RF-Oma1* regions of *B. vulgaris* contained members of the PPR protein gene family and the *R* gene family (Supplementary Tables S3 and S4). However, no other homologous genes were shared.
Differences in the selection patterns between the RF-Oma1 genes and the conserved type of Oma1

The RF-Oma1 genes appear to have evolved faster than the conserved type of Oma1, as indicated by the longer internal branches within the RF-Oma1 clade in both phylogenetic trees (Figures 1A and B). We therefore investigated the selection patterns for the RF-Oma1 genes and the conserved type of Oma1. The ratio of non-synonymous to synonymous substitutions in codons (dN/dS) was used as an index of selection. We selected orf20NK-198-2 (whose product had the binding activity) and orf20TK-81 (whose product had no binding activity) to represent the RF-Oma1 family, and their codons were aligned with those of the 42 conserved Oma1 genes used in the phylogenetic analysis (Supplementary Dataset 3). We compared two evolutionary models: one assumes that dN/dS is constant (\(\omega_0\)) among all the branches in a phylogenetic tree (Supplementary Dataset 3) (one-ratio model), and the other assumes dN/dS is different between the RF-Oma1 lineage (\(\omega_{RF1}\)) and the conserved type Oma1 lineage (\(\omega_{cons-Oma1}\)) (two-ratio model). As shown in Table 2, the values for \(\omega_0\), \(\omega_{RF1}\) and \(\omega_{cons-Oma1}\) are 0.24, 1.23 and 0.22, respectively. Our likelihood ratio test indicated that the difference between the models is significant (\(p = 2.0E^{-35}\)).

In Table 2, the \(\omega_{RF1}\) value is larger than the \(\omega_{cons-Oma1}\) value, suggesting that the RF-Oma1 genes contain sites that undergo positive selection. Therefore, we investigated whether such positively selected sites (with dN/dS > 1) are found in the RF-Oma1 genes. Among the twelve RF-Oma1 genes identified in our previous studies, one is an apparent pseudogene (orf20NK-219-1), and it was excluded from our analysis. The remaining eleven genes were subjected to further study. Their codons were aligned according to the alignment of the amino acid sequences, from which a phylogenetic tree was constructed (Supplementary Dataset 4). We tested five evolutionary models: Model M0 assumes constant dN/dS ratios at all sites, and under this model, the average dN/dS is calculated as 2.45. Models M1a and M8 allow positive selection, and the calculated dN/dS ratios are 3.57 and 3.58, respectively. Our likelihood tests comparing M1a with M2a and M7 with M8 resulted in the rejection of both M1a and M7 (\(p = 2.6E^{-33}\) and \(p = 2.1E^{-33}\), respectively). Therefore, we found evidence that the RF-Oma1 genes include some sites that have undergone positive selection. Our Bayes Empirical Bayes analysis (Yang et al. 2005) revealed a total of 37 amino acid residues that are potential positively selected sites.
selected sites with high posterior probability (> 0.95) (Supplementary Figure S1 and Figure 4). Twenty-six of these sites reside in the predicted transmembrane helices (Supplementary Figure S1 and Figure 4). Note that RF-Oma1 translation products were detected in the mitochondrial membrane fraction (Kitazaki et al. 2015). We compared the transmembrane helices of all the RF-Oma1 gene products. Of the three transmembrane helices, the N-terminal one is less clear in ORF20NK-198-3, ORF20PI 615522, ORF20NK-305-2 and ORF20NK-219-2, but this does not appear to affect the proteins’ ability to alter the higher order structure of preSATP6 (Supplementary Figure S2).

We also compared the selection pattern of bvOma1 with that of the RF-Oma1 genes. Because our phylogenetic trees suggested bvOma1 evolution is slower than RF-Oma1, we thought identification of its polymorphic sites would require larger sample size. We chose cultivated B. vulgaris species, which are inter-crossable (Biancardi et al. 2020). We amplified cDNA copies of bvOma1 from 57 plants, representing 50 accessions of B. vulgaris (Supplementary Table S2), and subjected them to direct sequencing. Among the 57 sequences, 19 contained one to several heterozygous sites and 38 were entirely homozygous. We found several groups of identical sequences within these sequences so that the total number of different bvOma1 sequences was 11. Unlike the 11 RF-Oma1 genes, the 11 bvOma1 sequences can be aligned with no indels (Supplementary Dataset 5). The codons were aligned to test the evolutionary models (Supplementary Dataset 5), and the statistics are shown in Table 2. Our likelihood tests of M0-M1a and M0-M7 rejected M0 (p = 0.0002 and p = 0.0003, respectively). However, in the tests of M1a-M2a and M7-M8, the null hypotheses (i.e. M1a and M7, respectively) were not rejected (p = 0.153 and p = 0.111, respectively). This suggests differences in the numbers of positively selected codons between bvOma1 and the RF-Oma1 genes.

**Phylogenetic network of the RF-Oma1 gene copies**

As mentioned previously, the number of clustered RF-Oma1 gene copies varies among sugar beet lines (Moritani et al. 2013; Arakawa et al. 2019b). Unequal crossing over was suggested to explain this polymorphism (Arakawa et al. 2019b). In the present study, we adopted a phylogenetic network analysis (Huson and Bryant 2006) to analyze the relationships among the 11 functional RF-Oma1 genes. The results showed a reticulate network, in which five RF-Oma1 genes clustered together while the remaining genes were separated from each other (Figure 5). Huson and Bryant (2006) pointed out that recombination can be one of the mechanisms to explain such topology.

The 11 RF-Oma1 genes have been classified into two classes based the properties of their translation products: those capable of altering the higher order structure of preSATP6 protein via a
molecular chaperone-like activity (6 copies), and those not capable (5 copies) (T. Arakawa et al., manuscript submitted). The former class is separated into two groups that are distantly positioned in the phylogenetic network: one includes orf20\textsubscript{NK-198-1}, orf20\textsubscript{NK-198-2}, orf20\textsubscript{NK-198-4}, orf20\textsubscript{NK-105-1}, and orf20\textsubscript{Fukkoku} and the other includes only orf20\textsubscript{NK-198-3} (Figure 5). The amino acid sequence homology between ORF20\textsubscript{NK-198-3} and the proteins encoded by former group is 88–89% (T. Arakawa et al., manuscript submitted).

**Discussion**

Unlike in model eukaryotes such as yeast and mouse, the flowering plant *Omal1* does not always occur as a single copy gene. We found that about 35% of the plant genomes listed in Table 1 contain multiple *Omal* genes. Whole genome duplication can partly explain these multiple copies, because polyploidization is common in flowering plants (Soltis et al. 2015). For example, several rounds of whole genome duplication are known to have occurred in the Fabales (Sato et al. 2008; Schmutz et al. 2010; Young et al. 2011), and this seems to be consistent with the results shown in Table 1. Whole genome duplication may also explain the occurrence of two *Omal* copies in *C. quinoa*, an allotetraploid species (Jarvis et al. 2017), because two chromosomal segments containing conserved type *Omal* homologues have very similar gene contents and arrangements (Figure 2). *A. hypochondriacus* is a paleoallotetraploid species but has only one *Omal* homologue. Although Scaffold\_7 and the Scaffold\_13 in *A. hypochondriacus* (depicted in Figure 2) are derived from homeologous chromosomes, it is possible that a redundant *Omal* gene was lost from the Scaffold\_7.

On the other hand, we speculate that whole genome duplication does not explain the evolution of the *RF-Omal* genes, because no recent whole genome duplication event is evident in *B. vulgaris* (Dohm et al. 2014), and our phylogenetic data suggest that the *RF-Omal* genes evolved relatively recently. Interestingly, *bvOma1*, the *S. oleracea Oma1*, and *atOma1* are located on chromosomal segments that are syntenic with one another (Figure 2). Therefore, this gene arrangement seems to represent the ancestral organization, and *bvOma1* is the likely orthologue of *atOma1*. In contrast, the gene arrangement in the neighborhood of the *RF-Omal* genes is dissimilar to that of *bvOma1*. Although both regions contain members of the PPR protein and *R* gene families, these families are known to be very large, with hundreds of copies per plant genome (Fujii et al. 2011). Therefore, the occurrence of such genes does not necessarily indicate a syntenic relationship. We favor the notion that the first *RF-Omal* gene was generated via a segmental duplication of a small chromosomal region. Considering that all the *RF-Omal* gene copies identified to date are interrupted by two introns whose
number and position are similar to those of bvOma1, retrotransposition (Hurles 2004) as the mechanism of duplication is unlikely.

Our phylogenetic analysis suggests that the RF-Oma1 lineage diverged after the establishment of the Caryophyllales. Although we have no phylogenetic data to further delimit when the chromosomal segment was duplicated, it may be possible to gain further information from the gene arrangement surrounding the RF-Oma1 genes. We found that the gene arrangement near the locus is conserved in S. oleracea and C. quinoa but not in A. hypochondriacus. A phylogenetic study of the Amaranthaceae suggests that Beta, Spinacia, and Chenopodium are closely related to one another and more distantly related to Amaranthus (Kadereit et al. 2000; Fuentes-Bazan et al. 2012). Given this, it seems unlikely that the duplication occurred in the common ancestor of the four species, because this scenario requires a series of events including multiple losses of the duplicated Oma1. The most parsimonious explanation may be that the framework of the gene arrangement surrounding the RF-Oma1 genes had been gradually established before the duplication occurred. However, this is not conclusive because the resolution of our phylogenetic analysis is relatively low due to the lack of genome information for other Caryophyllalean plants, and further investigations are needed. In either of the scenarios, the original RF-Oma1 gene appears to be a lineage-specific paralogue.

The evolutionary patterns of bvOma1 and the RF-Oma1 family are very different: whereas bvOma1 is a conserved, single copy gene (Arakawa et al. 2019a; this study), the RF-Oma1 genes have diverged in terms of copy number and nucleotide sequence (e.g. Arakawa et al. 2019b). Our network analysis provides additional data suggesting the involvement of recombination in the molecular diversity of the RF-Oma1 family. This implies that unequal crossing over between alleles has produced various RF-Oma1 genes. For an RF-Oma1 gene to suppress Owen-type mitochondria, its translation product should bind to the preSATP6 protein (Kitazaki et al. 2015). Our previous study revealed that this activity is an acquired trait of some RF-Oma1 copies because neither bvOma1 nor atOma1 has such activity (Arakawa et al. 2019a). Therefore, the role of recombination in this acquisition is interesting. Our network analysis suggests that the RF-Oma1 genes whose translation products lack the binding activity are more various (Figure 5), raising the possibility that the majority of recombinants are non-suppressive. This may suggest that the number of non-suppressive RF-Oma1 genes would increase with each recombination event. As a result, non-suppressive RF-Oma1 genes might be generated and stacked within alleles, leading to many novel non-restoring alleles (i.e. recessive rf1 alleles). However, molecular variation among recessive rf1 alleles in sugar beet is very small (Ohgami et al. 2016). In addition, previous genetic studies have shown that the restoring genotype (presumably containing
dominant \textit{Rf1} is prevalent in the \textit{B. vulgaris} genetic resources (Touzet 2012). This might indicate selection to maintain suppressive \textit{RF-Oma1} copies in the allele, but the details of this have not been investigated. Furthermore, we found two independent groups of suppressive \textit{RF-Oma1} genes (Figure 5), and it is unknown whether other types of suppressive \textit{RF-Oma1} genes exist in the \textit{B. vulgaris} genetic resources. Further studies are necessary to clarify the entire molecular diversity of the \textit{RF-Oma1} gene family, and to understand what selection pressures have shaped the molecular variation within the family.

The molecular evolution of the \textit{Rf} gene families has not been studied in detail except for the \textit{RFL} genes, which form a subset of the PPR gene family (Fujii et al. 2011). Therefore, it is valuable to compare the evolutionary patterns of the \textit{RFL} and \textit{RF-Oma1} families. Although their gene products and functions are completely different, they share the features of gene clustering and copy number variation (Kubo et al. 2020). Our present study adds another shared feature: the ratios of nonsynonymous to synonymous substitutions is high in both families. As is the case for \textit{RFL} genes, the \textit{RF-Oma1} genes include codons that showed high probabilities of positive selection. In the \textit{RFL} genes, such codons correspond to amino acid residues that are directly associated with substrate recognition: they are involved in recognizing specific RNA sequences (Fujii et al. 2011). Many of the positively selected codons in the \textit{RF-Oma1} alleles are confined to the transmembrane helices. Yeast YTA12 is an \textit{m-AAA} protease that is also involved in the quality control of mitochondria, and one of its transmembrane helices plays an important role in substrate recognition for dislocating the substrate from the membrane (Lee et al. 2017). Similarly, it may be that the positively selected codons in the \textit{RF-Oma1} genes are associated with recognition of the preSATP6 protein to exert the molecular chaperone-like activity. The alteration of amino acid residues in the N-terminal transmembrane helix may change the topology of the protein, but this alteration appears not to affect the suppression function (Supplementary Figure S2). Hence, the functional significance of the amino acid substitutions is unclear. Further studies are needed to analyze the relationship between the positively selected codons and substrate specificity, and this will contribute to our understanding of how the dominant \textit{Rf1} allele emerged. Knowledge of the OMA1 protein structure would benefit such studies.

Positive selection in the \textit{RFL} family is considered to be adaptive, because variations in substrate specificity resulting from changes in the recognition motifs would be beneficial for coping with novel MS-inducing mitochondria that might emerge over time (Gaborieau et al. 2016). This notion predicts a PPR-type \textit{Rf} locus with multiple alleles, each counteracting a different MS-inducing mitochondrion. The role of positive selection in the case of the \textit{RF-Oma1} family is currently obscure because little is known about how substrate specificity is determined (see above). It is possible that...
various \textit{RF-Oma1} copies could be beneficial to cope with various types of MS-inducing mitochondrial proteins, but to date, \textit{Rf1} alleles for different MS-inducing mitochondria have not been found in \textit{B. vulgaris}. Other \textit{Rf}s that counteract different kinds of MS-inducing mitochondria have been located on different chromosomes (Laporte et al. 1998; Touzet et al. 2004).

The \textit{RF-Oma1} and \textit{RFL} families are sources of \textit{Rf} genes. Their evolutionary patterns are very similar, although the former appears to have evolved more recently than the latter. It is difficult to infer how the \textit{RFL} genes originated, and there are questions about what type of duplication was involved, which PPR gene evolved into the first \textit{RFL} gene (the PPR genes constitute a large family), and what expression pattern was exhibited by the original \textit{RFL} gene. These questions are related to the initial stages of \textit{Rf} evolution (i.e. how \textit{Rf} emerged), and studies of evolutionarily young \textit{Rf}s may be valuable because remnants of their initial evolutionary processes may be preserved in the genome. We think the \textit{RF-Oma1} family represents an early stage of \textit{Rf} evolution, and the evolutionary pattern of this family could provide some answers to these questions. Considering that \textit{RF-Oma1} expression in anthers is more than 480 times higher than in vegetative organs (Arakawa et al. 2019a), the evolution of the \textit{RF-Oma1} family recalls that of certain \textit{Drosophila} genes: paralogues that are expressed specifically in the testes, that undergo positive selection, and encode mitochondrial proteins that possibly counteract male-harming mitochondrial variants (reviewed in Havird et al. 2019). We find it very interesting that paralogues specifically expressed in male-reproductive organs are the source of \textit{Rf}, and play key roles in the mitonuclear conflict in flowering plants.

Acknowledgements

This work is supported in part by the JSPS KAKENHI (grant number 18K05564 to KK and TK); the NARO Bio-oriented Technology Research Advancement Institution (BRAIN) research program on the Development of Innovative Technology (grant number 30001A to KK and TK); and the JSPS Research Fellowship for Young Scientists (16J01146 to TA).

Supplementary data

Supplementary Dataset 1: Supporting file_1.fas
Supplementary Dataset 2: Supporting file_2.fas
Supplementary Dataset 3: Supporting file_3.fas, Supporting file_4.txt and Supporting file_5.nwk
Supplementary Dataset 4: Supporting file_6.fas, Supporting file_7.txt and Supporting file_8.nwk
Supplementary Dataset 5: Supporting file_9.fas, Supporting file_10.txt and Supporting file_11.nwk
Figure S1
Figure S2
Table S1
Table S2
Table S3
Table S4
Table S5

Data availability statement
All sequence data generated for this study has been deposited at the DDBJ/EMBL/GenBank under accession numbers LC550296–LC550352.

References
Arakawa T, Uchiyama D, Ohgami T, Ohgami R, Murata T, et al. 2018. A fertility-restoring genotype of beet (*Beta vulgaris* L.) is composed of a weak *restorer-of-fertility* gene and a modifier gene tightly linked to the *Rf1* locus. PLoS ONE. 13:e0198409.

Arakawa T, Sugaya H, Katsuyama T, Honma Y, Matsui K, et al. 2019a. How did a duplicated gene copy evolve into a *restorer-of-fertility* gene in a plant? The case of *Oma1*. R Soc Open Sci. 6:190853.

Arakawa T, Ue S, Sano C, Matsunaga M, Kagami H, et al. 2019b. Identification and characterization of a semi-dominant *restorer-of-fertility* 1 allele in sugar beet (*Beta vulgaris*). Theor Appl Genet. 132:227-240.

Barkan A, Small I. 2014. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol. 65, 415-442.

Boriani E, Panella LW, McGrath JM, editors. 2020. *Beta maritima*. Gewerbestrasse: Springer Science+Buisiness Media.

Bryant D. 2003. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 21:255–265.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. 23:81-90.

Chen L, Liu Y.G. 2014. Male sterility and fertility restoration in crops. Annu Revs Plant Biol. 65:579-606.

Dahan J, Mireau H. 2013. The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR
genes. RNA Biol. 10:1469-1476.

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, et al. 2014. The genome of recently domesticated crop plant sugar beet (Beta vulgaris). Nature 505:546-549.

Duvick DN. 1965. Cytoplasmic pollen sterility in corn. Adv Genet. 13:1-56.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc. 2:953–971.

Fuentes-Bazan S, Mansion G, Borsch T. 2012. Towards a species level tree of the globally diverse genus Chenopodium (Chenopodiaceae). Mol Phylogenet Evol. 62:359-374.

Fujii S, Bond CS, Small ID. 2011. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci USA. 108:1723-1728.

Gaborieau L, Brown GG, Mireau H. 2016. The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. Front Plant Sci. 7:1816.

Geddy, R., Brown, G.G., 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics 8:130.

Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 68:225-252.

Guo X, Aviles G, Liu Y, Tian R, Unger BA. et al. 2020. Mitochondrial stress is relayed to the cytosol by an OMA1–DELE1–HRI pathway. Nature 579:427-432.

Hanson MR, Bentolila S. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:S154-S169.

Havird JC, Forsythe ES, Williams AM, Werren JH, Dowling DK, et al. 2019. Selfish mitonuclear conflict. Curr Biol. 29:R496-R511.

Hurles M. 2004. Gene duplication: the genomic trade in spare parts. PLoS Biol. 2:0900.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 23:254–267.

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, et al. 2017. The genome of Chenopodium quinoa. Nature 542:307-312.

Kadereit G, Hohmann S, Kadereit JW. 2000. A synopsis of Chenopodiaceae subfam. Betoideae and notes on the taxonomy of Beta. Wildenowia 36:9-19.

Kazama T, Okuno M, Watari Y, Yanase S, Koizuka C, et al. 2019. Curing cytoplasmic male sterility
via TALEN-mediated mitochondrial genome editing. Nat Plants 5:722-730.

Käser M, Kambacheld M, Kisters-Woike B, Langer T. 2003. Oma1, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. J Biol Chem. 278:46414-46423.

Kitazaki K, Arakawa T, Matsunaga M, Yui-Kurino R, Matsuhira H, et al. 2015. Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet (Beta vulgaris). Plant J. 83:290-299.

Kubo T, Arakawa T, Honma Y, Kitazaki K. 2020. What does the molecular genetics of different types of restorer-of-fertility genes imply? Plants 9:361.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874.

Laporte V, Merdinoglu D, Saumitou-Laprade P, Butterlin G, Vernet P, et al. 1998. Identification and mapping of RAPD and RFLP markers linked to a fertility restorer gene for a new source of cytoplasmic male sterility in Beta vulgaris ssp. maritima. Theor Appl Genet. 96:989–996.

Laser KD, Lersten NR. 1972. Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot Rev. 38:425-454.

Lee S, Lee H, Yoo S, Kim H. 2017. Molecular insights into the m-AAA protease–mediated dislocation of transmembrane helices in the mitochondrial inner membrane. J Biol Chem. 292:20058-20066.

Lynch M. 2007. The origins of genome architecture. Sunderland:Sinauer Associates. p. 494.

Matsuhira H, Kagami H, Kurata M, Kitazaki K, Matsunaga M, et al. 2012. Unusual and typical features of a novel restorer-of-fertility gene of sugar beet (Beta vulgaris L.). Genetics. 192:1347-1358.

Melonek J, Zhou R, Bayer PE, Edwards D, Stein N, et al. 2019. High intraspecific diversity of Restorer-of-fertility-like genes in barley. Plant J. 97:281-295.

Meyer E, Lehmann C, Boivin S, Brings L, De Cauwer I, et al. 2018. CMS-G from Beta vulgaris ssp. maritima is maintained in natural populations despite containing an atypical cytochrome c oxidase. Biochem J. 475:759–773.

Migdal I, Skibior-Blaszczyk R, Heidom-Czama M, Kolodziejczak M, Gurbic A, et al. 2017. AtOMA1 affects the OXPHOS system and plant growth in contrast to other newly identified ATP-independent proteases in Arabidopsis mitochondria. Front Plant Sci 8:1543.

Moritani M, Taguchi K, Kitazaki K, Matsuhira H, Katsuyama T, et al. 2013. Identification of the predominant nonrestoring allele for Owen-type cytoplasmic male sterility in sugar beet (Beta vulgaris L.): development of molecular markers for the maintainer genotype. Mol Breed. 32:91–
100.

Ohgami T, Uchiyama D, Ue S, Yui-Kurino R, Yoshida Y, et al. 2016. Identification of molecular variants of the nonrestoring restorer-of-fertility 1 allele in sugar beet (Beta vulgaris L.). Theor Appl Genet. 129:675-688.

Perlman SJ, Hodson CN, Hamilton PT, Opit GP, Gowen BE. 2015. Maternal transmission, sex ratio distortion, and mitochondria. Proc Natl Acad Sci USA. 112:10162-10168.

Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikita S, et al. 2017. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nat Commun. 8:14953.

Rice WR. 2013. Nothing in genetics makes sense except in light of genomic conflict. Annu Rev Ecol Evol Syst. 44:217-237.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, et al. 2008. Genome structure of the legume, Lotus japonicus. DNA Res. 15:227-239.

Scheffler IE. 2007. Mitochondria (Second edition). New York: Wiley-Liss. p. 462.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. 2010. Genome sequence of the paleopolyploid soybean. Nature 463:178-83.

Schnable PS, Wise RP. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci. 3:175-180.

Soltis PS, Marchant DB, van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. Curr Opin Genet Dev. 35:119-125.

Sullivan MJ, Petty NK, Beatson BA. 2011. Easyfig: a genome comparison visualiser. Bioinformatics 27:1009-1010.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.

Touzet P. 2012. Mitochondrial genome evolution and gynodioecy. In: Marechal-Drouard L, editor. Mitochondrial Genome Evolution. Oxford: Academic Press. p. 71-98.

Touzet P, Budar F. 2004. Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? Trends Plant Sci. 9:568-570.

Touzet P, Hueber N, Bürkholz A, Barnes S, Cuguen J. 2004. Genetic analysis of male fertility restoration in wild cytoplasmic male sterility G of beet. Theor Appl Genet. 109:240–247.

Wallace DC. 2018. Mitochondrial genetic medicine. Nat Genet. 50:1642-1649.

Wang K, Wang Z, Li F, Ye W, Wang J, et al. 2012. The draft genome of a diploid cotton Gossypium raimondii. Nat Genet. 44:1098-1103.
Yamamoto MP, Shinada H, Onodera Y, Komaki C, Mikami T, et al. 2008. A male sterility-associated mitochondrial protein in wild beets causes pollen disruption in transgenic plants, Plant J. 54:1027-1036.

Yang Z. Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 24:1586–1591.

Young ND, Debellé F, Oldroyd GED, Geurts R, Cannon SB, et al. 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480:520-524.
Table 1 Copy numbers of *Oma1* homologues in flowering plant genomes

| Taxon                  | Number of *Oma1* homologues | Type of *Oma1* |
|------------------------|------------------------------|----------------|
|                        |                              | Conserved  | Diverged | Total |
| **Eudicots**           |                              |             |          |       |
| - Superrosids          |                              |             |          |       |
| Brassicales            | *Arabidopsis thaliana*       | 1           | 0        | 1     |
|                        | *Arabidopsis lyrata*         | 1           | 0        | 1     |
|                        | *Capsella rubella*           | 1           | 0        | 1     |
|                        | *Brassica rapa*              | 1           | 0        | 1     |
|                        | *Carica papaya*              | 1           | 0        | 1     |
| Malvales               | *Gossypium raimondii*†        | 2           | 0        | 2     |
|                        | *Theobroma cacao*            | 1           | 0        | 1     |
| Sapindales             | *Citrus sinensis*            | 1           | 0        | 1     |
| Myrtales               | *Eucalyptus grandis*         | 1           | 1        | 2     |
| Cucurbitales           | *Cucumis melo*               | 1           | 0        | 1     |
|                        | *Citrullus lanatus*          | 1           | 0        | 1     |
| Rosales                | *Prunus persica*             | 1           | 0        | 1     |
|                        | *Malus domestica*            | 1           | 0        | 1     |
|                        | *Fragaria vesca*             | 1           | 2        | 3     |
| Fabales                | *Lotus japonicus* ††         | 0           | 2        | 2     |
|                        | *Medicago truncatula* ††     | 2           | 6        | 8     |
|                        | *Glycine max* ††             | 2           | 0        | 2     |
| Malpighiales           | *Ricinus communis*           | 1           | 0        | 1     |
|                        | *Manihot esculenta*          | 1           | 0        | 1     |
|                        | *Populus trichocarpa*        | 1           | 0        | 1     |
| Vitales                | *Vitis vinifera*             | 1           | 0        | 1     |
| Saxifragales           | *Kalanchoe fedtschenkoi*     | 1           | 0        | 1     |
| - Superasterids        |                              |             |          |       |
| Asterales              | *Helianthus annuus*          | 1           | 3        | 4     |
|                        | *Lactuca sativa* †††         | 2           | 0        | 2     |
|                        | *Cynara cardunculus*         | 1           | 0        | 1     |
| Solanales              | *Capsicum annuum*            | 1           | 0        | 1     |

http://mc.manuscriptcentral.com/gbe
| Species                        | Numbers |
|-------------------------------|---------|
| **Solanum lycopersicum**      | 1 0 1   |
| **Solanum tuberosum**         | 1 0 1   |
| **Lamiales**                  |         |
| **Sesamum indicum**           | 2 0 2   |
| **Caryophyllales**            |         |
| **Beta vulgaris**             | 1 3 4   |
| **Spinacia oleracea**         | 1 0 1   |
| **Chenopodium quinoa**        | 2 0 2   |
| **Amaranthus hypochondriacus**| 1 0 1   |
| **Monocots**                  |         |
| **Poales**                    |         |
| **Zea mays**                  | 1 0 1   |
| **Oryza sativa**              | 1 1 2   |
| **Hordeum vulgare**           | 1 2     |
| **Basal Magnoliophyta**       |         |
| **Amborellales**              |         |
| **Amborella trichopoda**      | 1 0 1   |

†1 to †6 In these cases, taxon-specific whole genome duplication is suggested (Wang et al. 2012; Satoh et al. 2008; Young et al. 2011; Schmutz et al. 2010; Reyes-Chin-Wo et al. 2017; and Jarvis et al. 2017, respectively).

*1 Data adopted from Arakawa et al. (2019a).

*2 Part of the coding region is missing in the gene model.
| Lineage of gene | Type of model | dN/dS (ω) | lnL | 2ΔlnL |
|----------------|--------------|-----------|-----|-------|
| **Conserved type** | Branch (one ratio) | ω₀ = 0.24 | -36059.34 | 154.28 |
| Oma1 and RF-Oma1 | and Branch (two ratios) | ω₁RF = 1.23 | ω₁cons-Oma₁ = 0.22 | (p = 2.0E-35) |
| RF-Oma1 | M0 | 2.45 | -2870.16 | - |
| | M1a | 0.44 | -2856.59 | 27.14 [M0-M1a] (p = 2.6E-33) |
| | M7 | 0.40 | -2856.82 | 26.68 [M0-M7] (p = 2.4E-07) |
| RF-Oma1 | M2a | 3.57 | -2781.58 | 150.02 [M1a-M2a] (p = 2.6E-33) |
| | M8 | 3.58 | -2781.58 | 150.48 [M7-M8] (p = 2.1E-33) |
| bvOma1 | M0 | 0.096 | -2159.93 | - |
| | M1a | 0.071 | -2153.13 | 13.6 [M0-M1a] (p = 0.0002) |
| | M7 | 0.10 | -2153.50 | 12.86 [M0-M7] (p = 0.0003) |
| | M2a | 0.10 | -2151.25 | 3.76 [M1a-M2a] (p = 0.153) |
| | M8 | 0.10 | -2151.30 | 4.4 [M7-M8] (p = 0.111) |
Figure legends

Fig. 1. Phylogeny of the flowering plant Oma1 genes, inferred from their amino acid sequences. Trees were drawn using the neighbor-joining (A) and maximum-likelihood (B) methods. Scale bars indicate evolutionary distances. Superrosids, superasterids, monocots and basal magnoliophyta are indicated by R, A, M, and BM, respectively, in parentheses. Bootstrap values are shown near the nodes. Abbreviations of scientific names are: Al, Arabidopsis lyrata; At, Arabidopsis thaliana; Cr, Capsella rubella; Br, Brassica rapa; Cp, Carica papaya; Tc, Theobroma cacao; Gr, Gossypium raimondii; Me, Manihot esculenta; Rc, Ricinus communis; Pt, Populus trichocarpa; Cs, Citrus sinensis; Cl, Citrullus lanatus; Cm, Cucumis melo; Vv, Vitis vinifera; Kf, Kalanchoe fedtschenkoii; Eg, Eucalyptus grandis; Si, Sesamum indicum; St, Solanum tuberosum; Sl, Solanum lycopersicum; Ca, Capsicum annuum; Cc, Cynara cardunculus; Ha, Helianthus annuus; Ls, Lactuca sativa; Pp, Prunus persica; Md, Malus domestica; Fv, Fragaria vesca; Mt, Medicago truncatula; Gm, Glycine max; Ah, Amaranthus hypochondriacus; Cq, Chenopodium quinoa; So, Spinacia oleracea; Bv, Beta vulgaris; Zm, Zea mays; Os, Oryza sativa; and Mp, Marchantia polymorpha.

Fig. 2. Microsynteny of chromosomal regions associated with Oma1 homologues. Annotated genes are shown as horizontal arrows. Genes homologous to the B. vulgaris queries are linked by thin lines. A 50 kbp scale bar is shown on the bottom right. The Oma1 homologue in each segment is shown as a red arrow. Note that the Amaranthus Scaffold_7 has no Oma1 homologue. The chromosomal regions with NCBI accession numbers are: Arabidopsis, Arabidopsis thaliana (NC_003076, from 20979520 to 21080594); Beta, Beta vulgaris (NW_017567367, from 71170 to 557770); Spinacia, Spinacia oleracea (NW_018931419, from 994825 to 1414068); Chenopodium, Chenopodium quinoa (NW_0142475 to 1649963); and Amaranthus Scaffold_7, Amaranthus hypochondriacus (from 22662158 to 23111136); and Amaranthus Scaffold_13, A. hypochondriacus (from 4279078 to 4673615).

Fig. 3. Microsynteny of chromosomal segments associated with the RF-Oma1 genes. Annotated genes within the segments are shown by horizontal arrows. Genes homologous to the B. vulgaris queries are linked by thin lines. A 20 kbp scale bar is shown at the bottom right. The RF-Oma1 gene is shown as a purple arrow. We refer to this copy as orf20_AWSS220. An asterisk denotes the S. oleracea genes LOC110796815, LOC110796811, and LOC110797175 (from left to right) that are annotated as abscisic stress-ripening protein 3-like, long non-coding (Inc) RNA, and lysine-rich arabinogalactan protein 19-
like, respectively. A plus sign denotes *C. quinoa* LOC110725390, which is annotated as an IncRNA but is not related to *S. oleracea* LOC110796811. The chromosomal regions with NCBI accession numbers are: Beta, *B. vulgaris* (NC_025814, from 2300755 to 2586046); Spinacia, *S. oleracea* (NW_018931398, from 1613247 to 1362360); Chenopodium NW_018742205, *C. quinoa* (from 265231 to 482238); and Chenopodium NW_018743021, *C. quinoa* (from 3318973 to 3346398).

Fig. 4 Positional relationships between positively selected amino acid residues and transmembrane helices in the *RF-Oma1* genes. Vertical and horizontal axes show probability and residue number, respectively. **A.** Positively selected codons revealed by Bayes Empirical Bayes analysis. Each vertical line indicates the posterior probability of positive selection on each residue, and is orange if the posterior probability is more than 0.95. **B.** Plot showing the probability that each amino acid residue sits within a transmembrane helix (red vertical line), inside the membrane (blue graph line) or outside the membrane (purple graph line).

Fig. 5 Phylogenetic network of eleven *RF-Oma1* genes in *B. vulgaris*. Genes capable of altering the higher order structure of the cognate MS-inducing mitochondrial protein are highlighted in green or red. A scale bar indicates evolutionary distances. The origins of the *RF-Oma1* genes are as follows: *orf20_NK-198-1, orf20_NK-198-2, orf20_NK-198-3, and orf20_NK-198-4* are from sugar beet line NK-198; *orf20_NK-305-1 and orf20_NK-305-2* are from sugar beet line NK-305; *orf20_NK-219-2 and orf20_NK-219-3* are from sugar beet line NK-219mm-O; and *orf20_TK-81, orf20_Pi615522, and orf20_takkoku* are from sugar beet line TK-81mm-O, sugar beet line PI 615522, and leaf beet accession 'Fukkoku-ouba', respectively (Matsuhira et al. 2012; Ohgami et al. 2016; Arakawa et al. 2018, 2019b).
Fig. 1 / Phylogeny of the flowering plant Oma1 genes, inferred from their amino acid sequences.

275x190mm (300 x 300 DPI)
Fig. 2 / Microsynteny of chromosomal regions associated with Oma1 homologues.

275x190mm (300 x 300 DPI)
Fig. 3 / Microsynteny of chromosomal segments associated with the RF-Oma1 genes.
Fig. 4 / Positional relationships between positively selected amino acid residues and transmembrane helices in the RF-Oma1 genes.
Fig. 5 / Phylogenetic network of eleven RF-Oma1 genes in B. vulgaris.
126x84mm (300 x 300 DPI)