

RESEARCH ARTICLE

# Whole Mitochondrial Genome Sequencing and Re-Examination of a Cytoplasmic Male Sterility-Associated Gene in Boro-Taichung-Type Cytoplasmic Male Sterile Rice

Tomohiko Kazama, Kinya Toriyama\*

Graduate School of Agricultural Science, Tohoku University, Sendai, Japan

\* [torikin@m.tohoku.ac.jp](mailto:torikin@m.tohoku.ac.jp)



OPEN ACCESS

**Citation:** Kazama T, Toriyama K (2016) Whole Mitochondrial Genome Sequencing and Re-Examination of a Cytoplasmic Male Sterility-Associated Gene in Boro-Taichung-Type Cytoplasmic Male Sterile Rice. PLoS ONE 11(7): e0159379. doi:10.1371/journal.pone.0159379

**Editor:** Hiroshi Ezura, University of Tsukuba, JAPAN

**Received:** April 7, 2016

**Accepted:** July 3, 2016

**Published:** July 14, 2016

**Copyright:** © 2016 Kazama, Toriyama. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All the nucleotide sequences are available from the DDBJ database (accession numbers AP017385, AP017386). All the other relevant data are within the paper and its Supporting Information files.

**Funding:** This study was partly supported by grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan (<http://www.mext.go.jp/english/>, grant number 24117502 to K.T.), Japan Society for the Promotion of Science (<http://www.jspss.go.jp/english/index.html>, grant numbers 23380002, 26292002 to K.T. and 16H06182 to T. K.), the Ministry of Agriculture Forestry and Fishery,

## Abstract

Nuclear genome substitutions between subspecies can lead to cytoplasmic male sterility (CMS) through incompatibility between nuclear and mitochondrial genomes. Boro-Taichung (BT)-type CMS rice was obtained by substituting the nuclear genome of *Oryza sativa* subsp. *indica* cultivar Chinsurah Boro II with that of *Oryza sativa* subsp. *japonica* cultivar Taichung 65. In BT-type CMS rice, the mitochondrial gene *orf79* is associated with male sterility. A complete sequence of the Boro-type mitochondrial genome responsible for BT-type CMS has not been determined to date. Here, we used pyrosequencing to construct the Boro-type mitochondrial genome. The contiguous sequences were assembled into five circular DNA molecules, four of which could be connected into a single circle. The two resulting subgenomic circles were unable to form a reliable master circle, as recombination between them was scarcely detected. We also found an unequal abundance of DNA molecules for the two loci of *atp6*. These results indicate the presence of multi-partite DNA molecules in the Boro-type mitochondrial genome. Expression patterns were investigated for Boro-type mitochondria-specific *orfs*, which were not found in the mitochondria from the standard *japonica* cultivar Nipponbare. Restorer of fertility 1 (RF1)-dependent RNA processing has been observed in *orf79*-containing RNA but was not detected in other Boro-type mitochondria-specific *orfs*, supporting the conclusion that *orf79* is a unique CMS-associated gene in Boro-type mitochondria.

## Introduction

Plant mitochondria have large genomes with many repeat sequences and complex secondary structures. Because of its sequence complexity, the plant mitochondrial genome can create new sequences and new open reading frames (*orfs*) via homologous recombination. Expression of such newly formed *orfs* in mitochondria sometimes leads to cytoplasmic male sterility (CMS). Thus, such *orfs* are known as CMS-associated genes. A CMS-associated gene, in general, is

Japan (<http://www.maff.go.jp/e/index.html>, Technology Research Promotion Program for the Agriculture, Forestry, Fisheries and Food Industry, grant number 26010A to K.T.), and Council for Science, Technology and Innovation (<http://www8.cao.go.jp/cstp/english/>, Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for creating next-generation agriculture, forestry and fisheries" (funding agency: Bio-oriented Technology Research Advancement Institution, NARO; <http://www.naro.affrc.go.jp/english/brain/>) to K.T.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

composed of a part of a house-keeping mitochondrial protein-coding gene and a sequence from unknown origin, and encodes a peptide with a trans-membrane domain. In land plants, more than 150 species with CMS have been found in nature [1]. In plant breeding programs, CMS lines can be artificially obtained by successive backcrossing to substitute the nucleus of a female parent with that of a male recurrent parent. A female cytoplasm donor cultivar itself is fertile when it possesses *Restorer of fertility (Rf)* genes in its nuclear genome, which suppress the expression of CMS-associated genes. Loss of *Rf* genes by nuclear substitution enables the expression of CMS-associated genes and results in male sterility. Thus, this phenotype is most likely caused by incompatibility between the mitochondrial and nuclear genomes. The cloning of *Rf* genes thus far has revealed that most are involved in the regulation of transcripts of CMS-associated genes [2]. For example, the *Rf* genes of petunia CMS and radish CMS encode an RNA-binding protein known as pentatricopeptide repeat protein, which was found to bind the designated RNA of a CMS-associated gene and to post-transcriptionally suppressed the accumulation of a CMS-causing protein [3, 4, 5].

CMS is a useful phenotype in the breeding of F<sub>1</sub> hybrid varieties, which provide 15%–20% higher yields than conventional inbred lines [6]. Employment of CMS eliminates the needs for hand emasculation to prevent the contamination of self-pollinated seeds. In rice, most of commercial F<sub>1</sub> hybrid seeds are obtained based on a three-line system consisting of CMS, maintainer and restorer lines. The wild abortive (WA)-type CMS, which consists of the cytoplasm of *Oryza rufipogon* and the nuclear genome of *O. sativa* subsp. *indica*, has been most widely used for F<sub>1</sub> hybrid rice production, and WA-type CMS-based hybrid rice has been planted on 6,000 kha in China [6]. The Boro-Taichung 65 (BT)-type CMS, which was obtained by combining the cytoplasm of *O. sativa* subsp. *indica* cultivar Chinsurah boro II and the nuclear genome of *O. sativa* subsp. *japonica* cultivar Taichung 65, has been also utilized for F<sub>1</sub> hybrid rice, which has been planted on 130 kha in China [6]. It is also commercially utilized in Japan [7].

Because of its agronomic importance, many efforts have been made to identify the CMS-associated genes in mitochondria. Historically, the first approach in the identification of CMS-associated genes was through a comparison of genes and gene expression patterns between CMS and non-CMS mitochondria. The most conventional method for identifying CMS-associated genes relies on the use of northern blot screening to compare the transcript profiles of CMS and fertility-restored lines. Candidate genes are identified through detection of modified expression patterns in the presence of *Rf* genes (reviewed in [8]). Such examination of mitochondrial genes and gene expression has identified two CMS-associated genes in rice; *orf79* and its sequence variants in BT-type CMS [9–12], Lead rice (LD)-type CMS [13] and Honglian (HL)-type CMS [14], and *orf352* in WA-type CMS [15].

The most well-characterized CMS-associated gene in rice is *orf79* in BT-type CMS [9–13, 16]. It encodes a 79 amino acid peptide with a transmembrane domain and carries a chimeric structure composed of part of a gene for mitochondrial cytochrome oxidase subunit I and a sequence of unknown origin [10]. It has been reported that Boro-type mitochondria contain two duplicated loci of the *atp6* gene, which encodes a subunit 6 of the ATPase synthase. One locus is completely identical to the Nipponbare *atp6* locus (N-*atp6*). Another locus of *atp6* (B-*atp6*) can be distinguished from N-*atp6* from the differences in the sequence at its 3' untranslated region. The *orf79* gene exists 208-nt downstream of B-*atp6*. In Boro-type mitochondria, the B-*atp6* is co-transcribed with *orf79* as a 2.0-kb RNA, whereas a 1.5-kb transcript containing *atp6* and a 0.45-kb transcript containing *orf79* are generated through RNA processing under the presence of a *restorer of fertility* gene, *Rf1*, which encodes a pentatricopeptide repeat protein [12, 16].

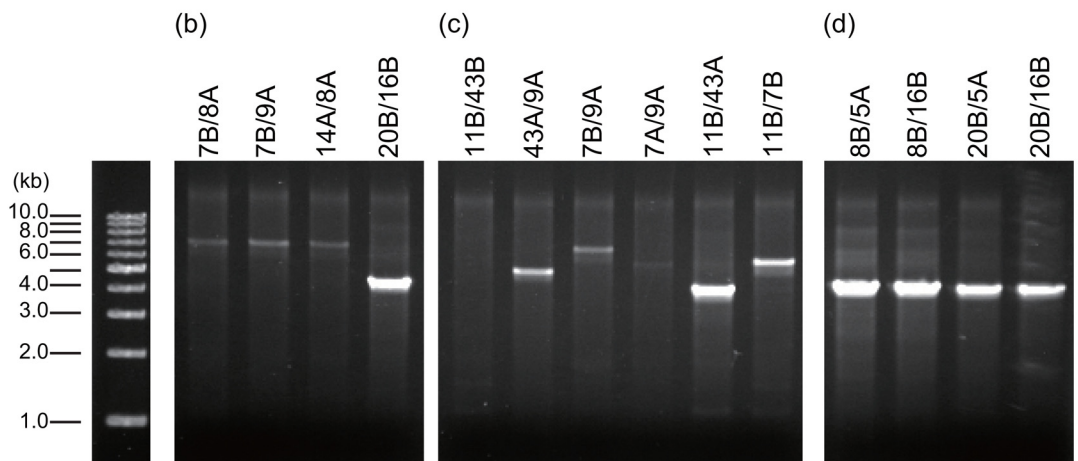
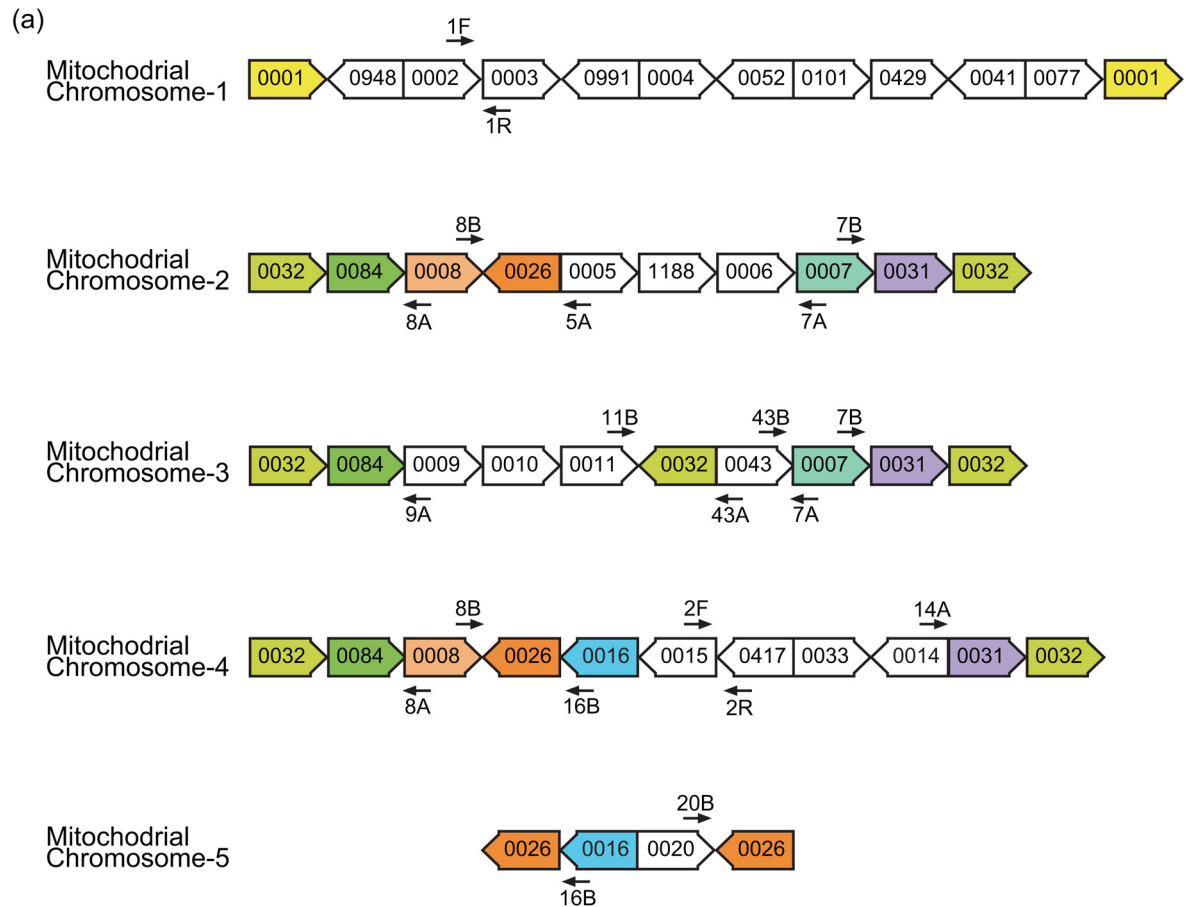
Recently, next-generation sequencing has become a popular method to uncover CMS-associated genes. Whole mitochondrial genome sequences have been compared to those of standard cultivars such as Nipponbare to screen for new *orfs* that were absent in the reference genome [17–20]. CMS-associated gene candidates have been selected based on whether they carry chimeric structures with known mitochondrial genes or encode proteins with a transmembrane domain, and whether or not they exhibit different expression patterns in the presence or absence of *Rf* genes. Whole mitochondrial genome sequences have been reported for Nipponbare [21], PA64S [22], *indica* cultivars 93–11 [22], IR6888B [17], and Hassawai [23], and several CMS lines such as LD-type CMS derived from *Oryza sativa*, WA-type CMS, Chinese wild rice (CW)-type CMS, RT98-type CMS, and RT102-type CMS derived from *Oryza rufipogon* [17–20]. Whole genome sequences of Boro-type mitochondria, however, have not been reported to date despite its economic importance in the production of F<sub>1</sub> hybrid seeds.

In this study, we determined the Boro-type mitochondrial genome sequence responsible for BT-type CMS using a Roche 454 GS pyrosequencing system, a next-generation sequencing platform. Sequence assembly indicated that the mitochondrial genome has a flexible multi-partite structure. Northern blot analysis was performed on Boro-type mitochondria-specific *orfs* to explore the possibility of RF1-regulated RNA processing in *orfs* other than *orf79*.

## Results and Discussion

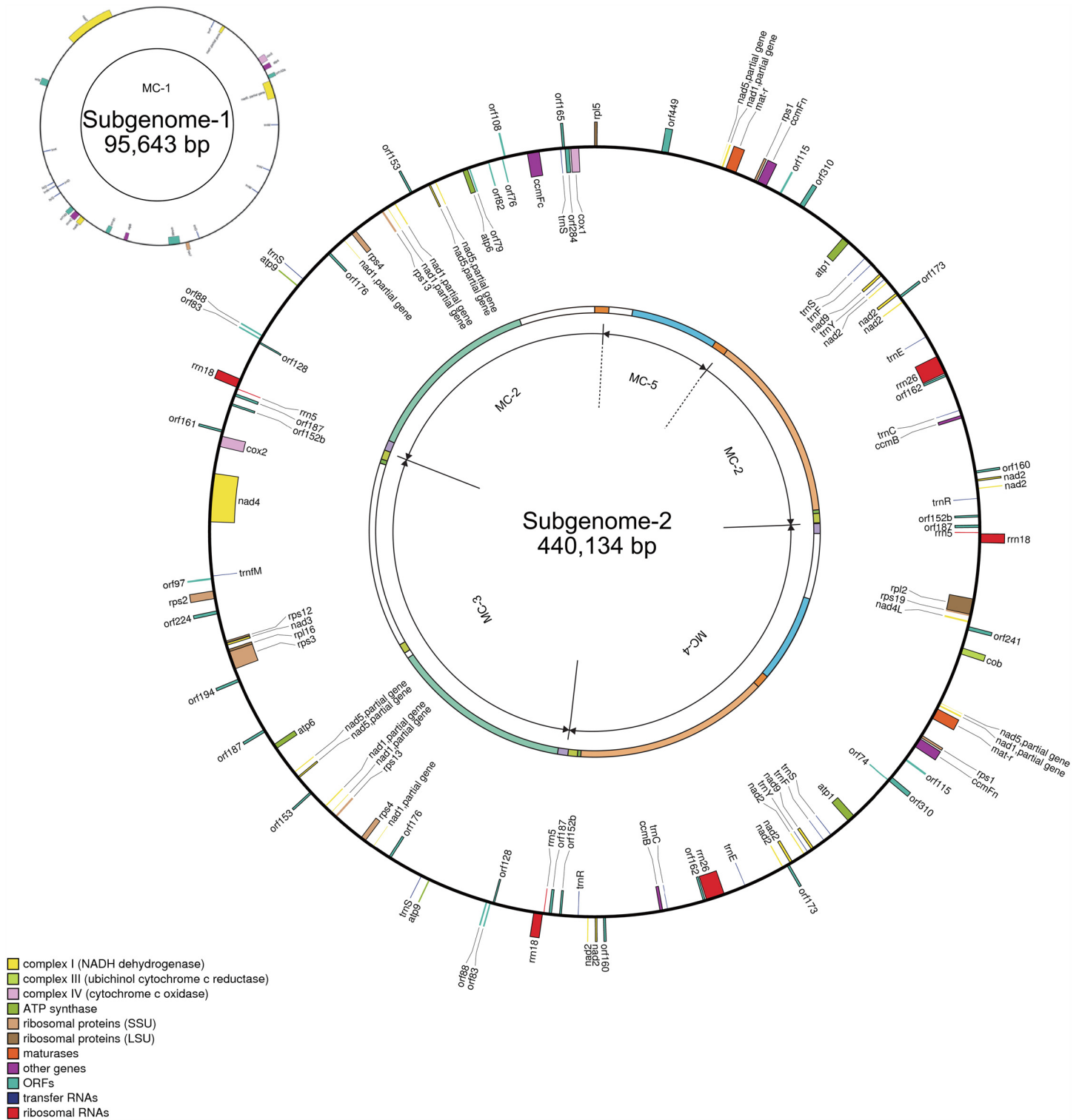
### Genome structure, size, genes and *orfs* of Boro-type mitochondria

Using contiguous sequences (contigs) and bridge sequences obtained from pyrosequencing, we assembled the Boro-type mitochondria genome sequences into five circular DNA molecules (Fig 1a and S1 Table). We named them Mitochondrial Chromosome 1 to 5 (MC-1, MC-2, MC-3, MC-4, and MC-5; Fig 1a). The circular molecules of MC-2 to MC-5 were validated by PCR analysis using sequence-specific primer sets 7B/8A, 7B/9A, 14A/8A and 20B/16B, respectively (Fig 1b). An inverse order of the contiguous contigs flanked by the inverted repeats of contig0032 in MC-3 (contigs 0032-0031-0007-0043-0032) was also possible; however, we adopted the order presented in Fig 1a because this order yielded more intense PCR amplification using primer set 7B/9A compared to that using 7A/9A (Fig 1c). Large, perfect repeats were shared between each mitochondrial chromosome (Fig 1a). MC-5 could be integrated into both MC-2 and MC-4 by recombination at contig0026, a high-read-depth contig. Integration of MC-5 into MC-2 was detected by PCR analysis using primer sets 20B/5A (Fig 1d). Strong band intensity of the amplicon using primer set 20B/5A supported the presence of such an integrated molecule. MC-2 integrated with MC-5, MC-3 and MC-4 can form a large single circular molecule where the components are connected to each other via the high-read-depth contig, contig0032. Amplification of approximately 7.3-kb bands using primer pairs 7B/8A, 7B/9A and 14A/8A (Fig 1b) supported the possibility of the presence of a connected circular molecule of MC-2, MC-3 and MC-4, as well as the possibility of the presence of each circular chromosome. Finally, the Boro-type mitochondrial genome was assembled into two large circular subgenomes named subgenome-1 and subgenome-2 (Fig 2). Subgenome-1 consists of MC-1 and was approximately 95 kb in length. Subgenome-2 was composed of MC-2, MC-3, MC-4, and MC-5 and was approximately 440 kb long (Figs 1 and 2). No common contigs were shared between subgenome-1 and subgenome-2. However, a 450-bp identical sequence was found at the junction of contig0002 and contig0003 in subgenome-1 and at the contig0417 side of contig0015 in subgenome-2 (S1 Fig). This suggests that subgenome-1 and subgenome-2 could form a master circle structure if recombination occurs at these identical sequences (Fig 3a). To confirm this hypothesis, we performed PCR analysis using primer sets targeting this identical sequence to detect a possible recombination event (Figs 1a, 3a and 3b). To obtain a direct



**Fig 1. Boro-type mitochondria sequences are assembled into five circular chromosomes.** (a) Compositions of contigs in five assembled mitochondrial chromosomes (MC). The arrow box indicates a contig and its direction. Arrows represent primers used for confirmation of contig connections. The same contigs are filled with the same colors. (b) Each of the five circular chromosomes was validated by PCR analysis using the primer sets indicated. (c) The inverse order of the contigs in MC-3 was detected by PCR analysis. (d) Integration of MC-5 into MC-2 was validated by PCR analysis.

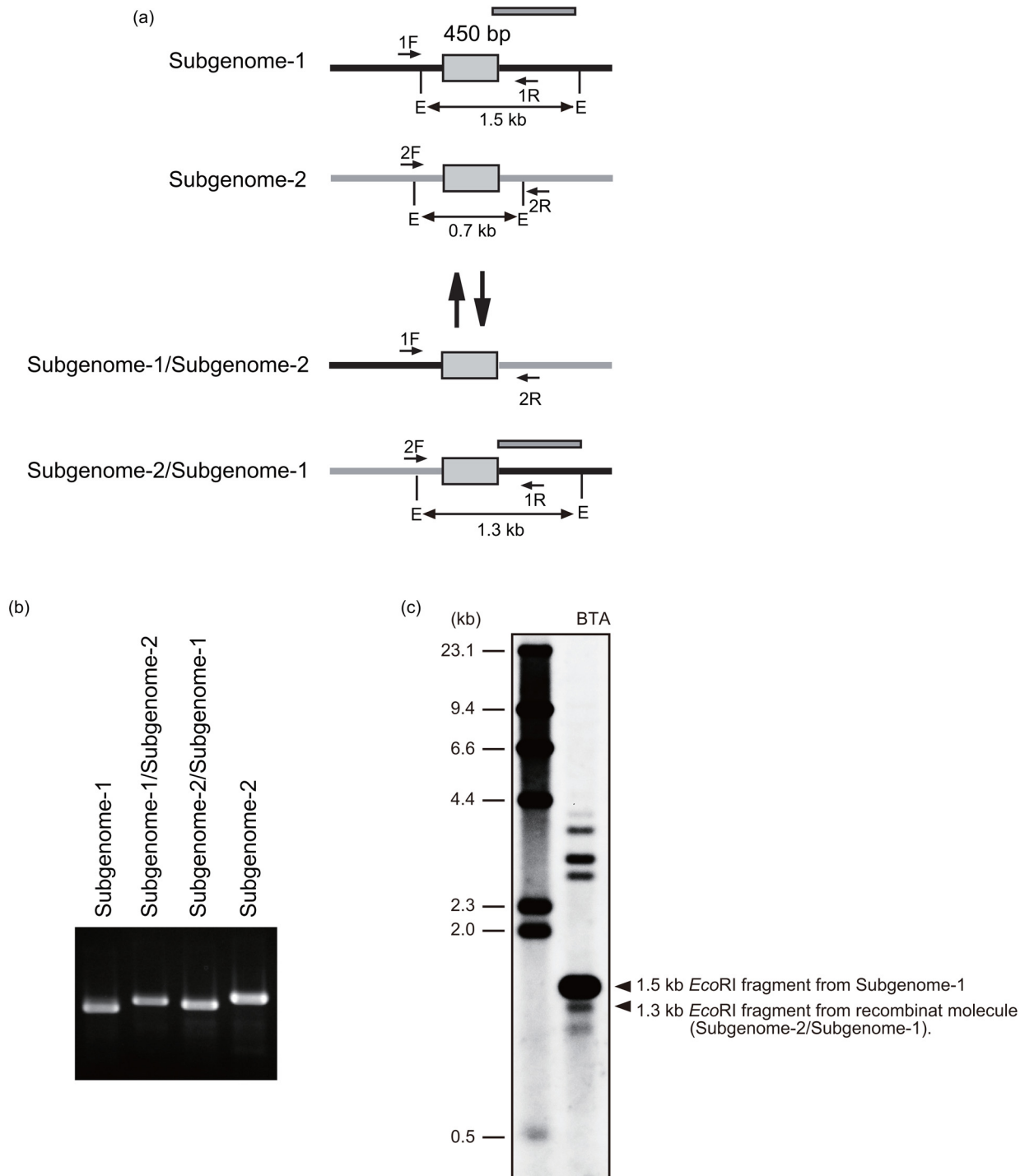
doi:10.1371/journal.pone.0159379.g001



**Fig 2. Graphical genome maps of subgenome-1 and subgenome-2 with all of the known genes and Boro-type mitochondria-specific *orf*s.** The outer circle indicates the graphical genome maps generated by using OGdraw. Colors in the inner circle correspond to those of the contigs listed in Fig 1.

doi:10.1371/journal.pone.0159379.g002





**Fig 3. Recombination between subgenome-1 and subgenome-2 via the 450-bp identical sequence is possible but rare.** (a) Schematic model of recombination between subgenome-1 and subgenome-2. Arrows indicate primers used for detection of recombination. *EcoRI* sites are indicated as E. (b) Recombinant fragments were detected by PCR. (c) Southern blot analysis detected the recombination molecule between subgenome-1 and subgenome-2; however, the frequency of recombination was lower than that of the non-recombinant subgenome-1 molecule.

doi:10.1371/journal.pone.0159379.g003

evidence for this recombination, Southern blot analysis was also performed using a probe targeting the sequence just adjacent to the identical region in subgenome-1 (Fig 3a). After *EcoRI* digestion, 1.5-kb major and 1.3-kb minor bands were detected (Fig 3c), which could have originated from subgenome-1 and the recombinant DNA molecule, respectively. This result indicates that both forms of the genome exist *in vivo*; however, the subdivided form of subgenome-1 is likely to be the more stable form *in vivo*. DNA fragments with sizes of 2.3 kb to 4.4 kb were also detected (Fig 3c), suggesting that another order of contigs could exist or that sub-stoichiometric shifting had occurred. A similar phenomenon has been reported in cucumber mitochondria, in which the large circular chromosomes favor subdivision by homologous recombination via direct repeat sequences [24].

*B-atp6* and *orf79* were found in contig0007 and contig0006 of MC-2 with read depths of 97.8 and 112.2, respectively (S1 Table). In contrast, the depth of contig0043 that corresponded to a sequence 49-bp downstream of *N-atp6* was as low as 12.1 (S1 Table, S2 Fig). The existence of an overlapping sequence in contig0007 and contig0043 was supported by the reported sequences of the *N-atp6* locus (DDBJ accession number S59890.1), which was also present in other mitochondrial genomes in S2 Table, except that of LD-CMS. These results suggest that the abundance of molecules including *N-atp6* is approximately 10% of that for *B-atp6*. A lower abundance of *N-atp6* was supported by the weaker signal intensity of *N-atp6* compared to that of *B-atp6* in the Southern blot analysis of an *EcoRI*-digested mitochondrial genome probed with *atp6* (S2 Fig), as previously reported [9, 10, 13]. These results suggest the presence of multi-partite molecules and the possibility that a fragment flanked by an inverted repeat of contig0032 in MC-3 (contigs 0032-0043-0007-0031-0032) might exist as a linear molecule.

It must be noted that a 3,041-bp-long sequence completely identical to the one found 78 bp upstream of the *atp6* gene to the end of contig0043 (S2 Fig) was also found in the chromosome 1 of Nipponbare (position 33,552,397..33,549,357 in The Rice Annotation Project Database). In ancestral species, the *N-atp6* sequence may have been transferred from mitochondria to the nucleus and integrated into the nuclear genome. In Boro-type mitochondria, the *N-atp6* sequence may be evolutionarily in the process of disappearing from the mitochondrial genome, starting from the time at which the *B-atp6-orf79* locus was created. The *N-atp6* locus is lost in LD-type mitochondria, which only have a *B-atp6-orf79* locus [18]. Further studies are required to define how *N-atp6* and *B-atp6* are maintained in Boro-type mitochondria.

We also obtained five plasmid-like small molecules (S1 Table) that have been previously reported as B1 (2,135 bp; DDBJ accession D00293), B2 (1,548 bp; DDBJ accession D00564), B3 (1,514 bp; DDBJ accession X16154) and B4 (969 bp; DDBJ accession X07904). The GC content of each plasmid-like molecule was 44.2%, 42.9%, 42.1%, and 49.8%, respectively, which was almost identical to that of subgenome-1 (44.0%) and subgenome-2 (43.9%). These values are almost comparable to the GC content of 43.8% in the mitochondrial genome of Nipponbare [22], indicating that the Boro-type mitochondrial genome has no distinct feature with regard to GC content.

Almost all of the mitochondrial genes reported by Notsu et al. [21] exist in the Boro-type mitochondrial genome, although exceptions include *orf288* and pseudo-*rps14*. *orf288* in Nipponbare mitochondria, changed into *orf310* because of a nucleotide insertion. pseudo-*rps14* lost six nucleotides in Boro-type mitochondria. The exon and intron structures of each gene have been conserved. The *nad1* and *nad5* genes are composed of five exons; each exon is transcribed separately and is processed via trans-splicing to form the functional message. Interestingly, in Boro-type mitochondria, exon 1 of *nad1* and exon 1 and exon 2 of *nad5* are in subgenome-1. The residual exons (exon 2 to exon 5 of *nad1* and exon 3 to exon 5 of *nad5*) are in subgenome-2. These findings suggest that when the mitochondrial genome is transmitted to

the next generation or daughter cells, both of the two subgenomes should be transmitted equally to maintain the expression of these genes.

### Some Boro-type *orfs* are not found in the Nipponbare mitochondrial genome

Once constructed, the Boro-type mitochondrial genome was then screened for the presence of *orfs* that encoded proteins longer than 70 amino acids because the CMS-associated gene *orf79* encodes a 79-amino-acid protein. More than 600 *orfs* were predicted in two subgenomes. A BLAST search identified 17 *orfs* that were not found in Nipponbare mitochondria (S2 Table). We selected all of these *orfs* for analysis as Boro-type mitochondria-specific *orfs*. The *orfs* were named based on the length of the proteins they encoded. Ten of these *orfs* were conserved as full-length coding regions in other types of CMS mitochondria (S2 Table). *orf76*, *orf79*, *orf82*, and *orf310* were only found in LD-type mitochondria [20]. We found *orf115* in RT98-, LD-, ZS97-, and WA-type CMS mitochondria [13, 20, 21]. *orf83*, *orf88*, and *orf128* were detected in RT98-, RT102-, LD-, CW-, and WA-type CMS mitochondria [15]. The partial fragments of *orf74*, *orf97*, *orf108*, *orf310*, and *orf449* completely matched those found in the Nipponbare mitochondrial genome, however. Additionally, the plasmid-like small molecules carried *orf80\_B1*, *orf95\_B1*, *orf138\_B2*, *orf73\_B3* and *orf83\_B4*, all of which were found in the plasmid-like small molecules in LD-type and CW-type CMS rice [18]. These results indicate that the *orfs* characteristic for Boro-type mitochondria are restricted to a set of *orf79*, *orf82*, *orf76* and *orf108*, which are adjacent to each other between the *B-atp6* and *ccmFc* loci in contig0006 (Fig 2).

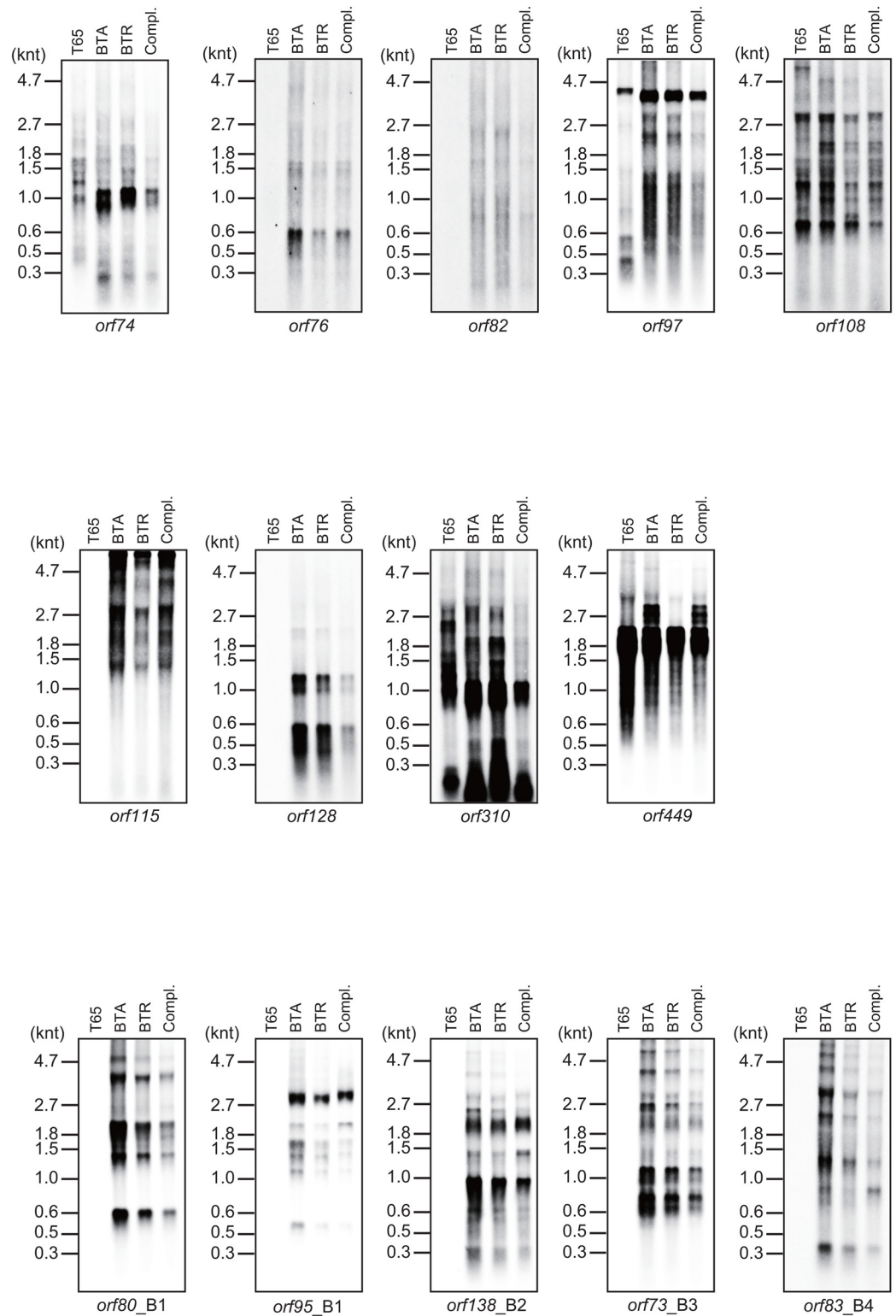
Chimeric features of gene structures were found in *orf449*, *orf310*, *orf97*, and *orf79* (S2 Table). *orf449* contains a partial sequence of the *cob* gene that encodes cytochrome b. This *orf449* has already been deposited in DNA databases as a gene for pseudo apocytochrome b (DDBJ accession X53711). *orf310* contains a partial sequence identical to *orf288* and is predicted to encode a hypothetical protein. *orf97* contains a partial sequence almost identical to that of *orf25* (*atp4*), which encodes a subunit of ATP synthase.

### Expression analysis of *orfs* specific to Boro-type mitochondria

To determine whether or not the expression pattern of these *orfs* is dependent on the presence of the fertility restorer gene, *Rf1*, we performed northern blot analyses on the 16 *orfs* (*orf74*, *orf76*, *orf82*, *orf83*, *orf88*, *orf97*, *orf108*, *orf115*, *orf128*, *orf310*, *orf449*, *orf80\_B1*, *orf95\_B1*, *orf138\_B2*, *orf73\_B3*, and *orf83\_B4*) using the BT-type CMS rice (BTA), fertility restorer rice (BTR) with *Rf1*, a transgenic BTA line transformed with *Rf1* (complemented line) and a maintainer line (T65) (Fig 4). Transcripts of the 14 *orfs* were detected, except for *orf83* and *orf88* in BTA, BTR, and the complemented line carrying Boro-type mitochondria. We expected that CMS-associated genes would show different transcript profiles for the CMS BTA line compared to those of the fertility-restored BTR line and the complemented line, depending on the absence or presence of *Rf1*. However, we did not detect an *Rf1*-dependent alteration of expression patterns. These results suggest that the *orfs* shown in Fig 4 are not likely to be CMS-associated genes, thus supporting the previous reports that *orf79* is the unique CMS-associated gene in Boro-type mitochondria [9, 10].

In our previous study, we reported that the recombinant RF1 protein bound to a 370-nt long RNA probe termed BD4, whose sequence corresponded to the intergenic region between *atp6* and *orf79* RNA [16]. Thus, there was a possible RF1 target sequence in this 370-nt sequence. A similarity search comparing the BD4 sequence with the Boro-type mitochondria sequence revealed that a 67-bp-long sequence in the BD4 region was identical to the 5' region





**Fig 4. Expression patterns of orfs specific to Boro-type mitochondria were not altered in the presence or absence of the restorer of fertility gene, *Rf1*.** Northern blot analysis was performed using a BT-type CMS rice (BTA), a fertility restorer with *Rf1* (BTR), a transgenic BTA complemented with *Rf1* (Compl), and a maintainer line (T65).

doi:10.1371/journal.pone.0159379.g004

of *orf97* that was 16 bp upstream from the start codon (S3 Fig). This implied that the transcript pattern of *orf97* could be altered by the presence of *Rf1*. However, we did not detect any differences when northern blot analysis was performed using an *orf97* probe (Fig 4). We have concluded that the 67-bp region does not contain a binding sequence for RF1. This indicates that *orf97* is not a CMS-associated gene of BT-type CMS rice, further supporting the previous finding that *orf79* is a unique CMS-associated gene post-transcriptionally affected by RF1 [9, 10].

## Conclusion

Pyrosequencing of the Boro-type mitochondrial genome revealed the presence of multipartite structures. The abundance of N-*atp6* molecules is quite low compared to that of B-*atp6* molecules, which suggests these structures are a unique feature of the Boro-type mitochondrial genome. We conclude that *orf79* is a unique CMS-associated gene in BT-type CMS. The whole-genome sequencing data of Boro-type mitochondria will provide useful information for comparative studies of CMS-causing genetic factors in rice.

## Materials and Methods

### Plant materials

The BT-type CMS rice, BTA, was made from the cytoplasm of the *indica* cultivar ‘Chinsurah Boro II’ with the nuclear background of the *japonica* cultivar ‘Taichung 65’. BTR is an isogenic restorer line, except for the *Rf1* locus, and contains the *Rf1* allele [16].

### DNA extraction and Southern blot analysis

Total DNA was extracted from a leaf blade using the DNeasy Plant Mini Kit (Qiagen, <https://www.qiagen.com/>). Mitochondrial isolation was performed according to the description provided in a previous study [25]. Mitochondrial DNA was extracted using a DNeasy Plant Mini Kit (Qiagen). To investigate recombination events between subgenome-1 and subgenome-2, 1 µg of total DNA was digested with *EcoRI*. Gene-specific DNA probes were produced using the PCR DIG Labeling Mix (Roche) by amplification of each gene using the following primer sets: sub1\_rec\_F1 (5′-CCATGGAAATGACGGAATTCC-3′) / sub1\_rec\_R1 (5′-AGGATAAGGGGGAGTAGTAG-3′). Hybridization, washing, and detection were performed according to the manufacturer’s instructions for the DIG DNA labeling and the detection kit (Roche). Southern blot analysis for the detection of *atp6* was performed using 300 ng of mitochondrial DNA by a method described previously [25].

### Sequence assembly

Pyrosequencing was performed using a GS-FLX system with 8-kb paired-end adoptees (Roche, <https://roche-biochem.jp/>) at TaKaRa Bio (<http://www.takara-bio.co.jp/>). This analysis resulted in 67,784,793 bases for the Boro-type mitochondrial genome and the sequence were assembled into 1,412 contigs and 26 scaffolds, using GS De Novo Assembler version 2.3 (Roche). After eliminating nuclear genomic contamination, we adopted a reference sequence-based assembly using the Nipponbare mitochondrial genome (DDBJ accession No. BA000029; [21]) as a reference genome, and as described previously [17]. The contigs were then manually assembled based on the scaffolds and the presence of bridge contigs, and validation was performed by PCR analysis in the same manner as described previously [18]. Ambiguous sequences of protein-coding regions that were distinct from those of Nipponbare, were confirmed by direct sequencing of the PCR products using a CEQ8000 genetic analyzer (Beckman Coulter, <http://www.beckmancoulter.co.jp/>). The PCR primers used for the confirmation of

contig linkages are listed in [S3 Table](#). The PCR reaction was performed using the leaf-derived total genomic DNA as a template and Ex-Taq polymerase (TaKaRa Bio). All of the contiguous contigs, together with each read depth, are listed in [S1 Table](#). Two hypothetical master circles were developed using a 'parsimonious' method, such that each contig appeared at least once when constructing the smallest genome. Other configurations can be considered from rearrangements at sites identical to those of the high-read-depth contigs. The graphical genome map was generated using OGDRAW software (<http://ogdraw.mpimp-golm.mpg.de/>; [26]).

### Specific *orfs* in Boro-type mitochondrial genome

The potential *orfs* were identified using Artemis software [27], which uses a threshold to identify ORFs, as previously used in an analysis of the WA-type CMS mitochondrial genome [17]. In the present study, we used 70 amino acids as the threshold, because the shorter mitochondrial peptides that cause CMS in rice such as *orf79* contain 79 amino acids. Common *orfs* with >99% identity to those in Nipponbare were identified using Sequencher software (Gene Codes Corporation, <https://www.genecodes.com/>) and were subsequently discarded. The TMHMM server v3.0 (<http://www.cbs.dtu.dk/services/TMHMM/>; [28]) was used to predict transmembrane domains in proteins encoded by *orfs*.

### RNA extraction and northern blot analysis

Secondary calli were transferred into fresh medium every 2 weeks, and 4-day-cultured calli were subjected to total and mitochondrial RNA extraction. Total RNA was extracted using an RNeasy Plant Mini Kit (Qiagen). DNA was eliminated using RNase-free DNase I (TaKaRa Bio). RNA (1 µg) was treated with Superscript III reverse transcriptase (Invitrogen, <http://www.thermofisher.co.jp/>) to produce complementary DNA. The mitochondrial isolation was performed according to a previous description [29], and RNA was extracted by RNAiso-Plus (TaKaRa Bio). Northern blot analysis was performed as previously described [12]. Gene-specific DNA probes were produced with PCR DIG Labeling Mix (Roche) by amplification of each *orf* using the primer sets listed in [S3 Table](#). Signals were detected and measured using ImageQuant LAS 4000 mini (GE Healthcare Life Sciences, <http://www.gelifesciences.co.jp/>).

### Supporting Information

**S1 Fig. An identical 450-bp-long sequence shared between subgenome-1 and subgenome-2.** Sequence alignment of the identical genomic region between subgenome-1 and subgenome-2. (EPS)

**S2 Fig. Comparison of N-*atp6* and B-*atp6-orf79* indicates a lower abundance of N-*atp6* molecules in the Boro-type mitochondrial genome.** (a) Schematic representation of N-*atp6* and B-*atp6-orf79* loci in contig0006-contig0007 and contig00043-contig00007, respectively. (b) Southern blot analysis of the *Eco*RI-digested mitochondrial genome probed with *atp6* and *coxI*. The *coxI* probe was used as an internal control. (EPS)

**S3 Fig. A sequence included in the intergenic region of *atp6-orf79* RNA was found in the 5' upstream region of *orf97*.** BD4 is a 370-nt-long RNA probe that binds the RF1 protein. A 67-bp-long sequence in the BD4 region is identical to a 16-bp upstream region of *orf97*. (EPS)

**S1 Table. Compositions of mitochondrial chromosomes (MC-1 to MC-5) in subgenome-1 and subgenome-2, as well as plasmid-like molecules.**

(XLSX)

**S2 Table. Characteristics of *orfs* specific to Boro-type mitochondria.**

(XLSX)

**S3 Table. Primers used for validation of the contig linkage.**

(XLSX)

**S4 Table. Primer sets used for synthesizing probes in Southern and northern blot analyses of *orfs* specific to Boro-type mitochondria.**

(XLSX)

## Author Contributions

Conceived and designed the experiments: KT TK. Performed the experiments: TK. Analyzed the data: TK KT. Contributed reagents/materials/analysis tools: TK KT. Wrote the paper: TK KT.

## References

1. Schnable PS, Wise RP. The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci.* 1998; 3: 175–180. doi: [10.1016/S1360-1385\(98\)01235-7](https://doi.org/10.1016/S1360-1385(98)01235-7)
2. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* 2008; 13(12):663–670. doi: [10.1016/j.tplants.2008.10.001](https://doi.org/10.1016/j.tplants.2008.10.001) PMID: [19004664](https://pubmed.ncbi.nlm.nih.gov/19004664/)
3. Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci U S A.* 2002; 99: 10887–10892. doi: [10.1073/Pnas.102301599](https://doi.org/10.1073/Pnas.102301599) PMID: [12136123](https://pubmed.ncbi.nlm.nih.gov/12136123/)
4. Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, et al. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* 2003; 35: 262–272. doi: [10.1046/J.1365-313X.2003.01799.X](https://doi.org/10.1046/J.1365-313X.2003.01799.X) PMID: [12848830](https://pubmed.ncbi.nlm.nih.gov/12848830/)
5. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, et al. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosen radish. *Plant J.* 2003; 34: 407–415. doi: [10.1046/j.1365-313X.2003.01735.x](https://doi.org/10.1046/j.1365-313X.2003.01735.x) PMID: [12753581](https://pubmed.ncbi.nlm.nih.gov/12753581/)
6. Huang JZ, E ZG, Zhang HL, Shu QY. Workable male sterility systems for hybrid rice: Genetics, biochemistry, molecular biology, and utilization. *Rice.* 2013; 7:14.
7. Tsushiya T, Bastawisi A, Yang ZY, Moon HP, Mann JA, Ikehashi H. Opportunities for and challenges to developing and using hybrid rice technology for temperate countries. In: Virmani SS, Mao CX, Hardy B, editors. *Hybrid rice for food security, poverty alleviation, and environmental protection*: International Rice Research Institute; 2002. pp. 53–82.
8. Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell.* 2004; 16: S154–S69. doi: [10.1105/Tpc.015966](https://doi.org/10.1105/Tpc.015966) PMID: [15131248](https://pubmed.ncbi.nlm.nih.gov/15131248/)
9. Iwabuchi M, Kyoizuka J, Shimamoto K. Processing followed by complete editing of an altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice. *The EMBO J.* 1993; 12: 1437–1446. PMID: [7682174](https://pubmed.ncbi.nlm.nih.gov/7682174/)
10. Akagi H, Sakamoto M, Shinjyo C, Shimada H, Fujimura T. A unique sequence located downstream from the rice mitochondrial *atp6* may cause male sterility. *Curr Genet.* 1994; 25: 52–58. PMID: [7915966](https://pubmed.ncbi.nlm.nih.gov/7915966/)
11. Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, et al. Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell.* 2006; 18: 676–687. doi: [10.1105/tpc.105.038240](https://doi.org/10.1105/tpc.105.038240) PMID: [16489123](https://pubmed.ncbi.nlm.nih.gov/16489123/)
12. Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K. Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. *Plant J.* 2008; 55: 619–628. TPJ3529 [pii]doi: [10.1111/j.1365-313X.2008.03529.x](https://doi.org/10.1111/j.1365-313X.2008.03529.x) PMID: [18435825](https://pubmed.ncbi.nlm.nih.gov/18435825/)

13. Itabashi E, Kazama T, Toriyama K. Characterization of cytoplasmic male sterility of rice with Lead Rice cytoplasm in comparison with that with Chinsurah Boro II cytoplasm. *Plant Cell Rep.* 2009; 28: 233–239. doi: [10.1007/s00299-008-0625-7](https://doi.org/10.1007/s00299-008-0625-7) PMID: [18956194](https://pubmed.ncbi.nlm.nih.gov/18956194/)
14. Yi P, Wang L, Sun QP, Zhu YG. Discovery of mitochondrial chimeric gene associated with cytoplasmic male sterility of HL rice. *Chin Sci Bull.* 2002; 47: 744–747.
15. Luo D, Xu H, Liu Z, Guo J, Li H, Chen L, et al. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat Genet.* 2013; 45: 573–577. doi: [10.1038/ng.2570](https://doi.org/10.1038/ng.2570) PMID: [23502780](https://pubmed.ncbi.nlm.nih.gov/23502780/).
16. Kazama T, Toriyama K. A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Lett.* 2003; 544:99–102. PMID: [12782297](https://pubmed.ncbi.nlm.nih.gov/12782297/)
17. Bentolila S, Stefanov S. A reevaluation of rice mitochondrial evolution based on the complete sequence of male-fertile and male-sterile mitochondrial genomes. *Plant Physiol.* 2012; 158: 996–1017. doi: [10.1104/pp.111.190231](https://doi.org/10.1104/pp.111.190231) PMID: [22128137](https://pubmed.ncbi.nlm.nih.gov/22128137/)
18. Fujii S, Kazama T, Yamada M, Toriyama K. Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. *BMC Genomics.* 2010; 11. doi: [10.1186/1471-2164-11-209](https://doi.org/10.1186/1471-2164-11-209)
19. Igarashi K, Kazama T, Motomura K, Toriyama K. Whole genomic sequencing of RT98 mitochondria derived from *Oryza rufipogon* and northern blot analysis to uncover a cytoplasmic male sterility-associated gene. *Plant Cell physiol.* 2013; 54: 237–43. doi: [10.1093/pcp/pcs177](https://doi.org/10.1093/pcp/pcs177) PMID: [23248202](https://pubmed.ncbi.nlm.nih.gov/23248202/)
20. Okazaki M, Kazama T, Murata H, Motomura K, Toriyama K. Whole mitochondrial genome sequencing and transcriptional analysis to uncover an RT102-type cytoplasmic male sterility-associated candidate gene derived from *Oryza rufipogon*. *Plant Cell Physiol.* 2013; 54: 1560–1568. doi: [10.1093/pcp/pct102](https://doi.org/10.1093/pcp/pct102) PMID: [23852329](https://pubmed.ncbi.nlm.nih.gov/23852329/)
21. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, et al. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genom.* 2002; 268: 434–445. doi: [10.1007/s00438-002-0767-1](https://doi.org/10.1007/s00438-002-0767-1) PMID: [12471441](https://pubmed.ncbi.nlm.nih.gov/12471441/)
22. Tian X, Zheng J, Hu S, Yu J. The rice mitochondrial genomes and their variations. *Plant Physiol.* 2006; 140: 401–410. doi: [10.1104/pp.105.070060](https://doi.org/10.1104/pp.105.070060) PMID: [16384910](https://pubmed.ncbi.nlm.nih.gov/16384910/)
23. Zhang T, Hu S, Zhang G, Pan L, Zhang X, Al-Mssallem IS, et al. The organelle genomes of Hassawi rice (*Oryza sativa* L.) and its hybrid in Saudi Arabia: genome variation, rearrangement, and origins. *PLoS One.* 2012; 7: e42041. doi: [10.1371/journal.pone.0042041](https://doi.org/10.1371/journal.pone.0042041) PMID: [22870184](https://pubmed.ncbi.nlm.nih.gov/22870184/)
24. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell.* 2011; 23: 2499–2513. doi: [10.1105/tpc.111.087189](https://doi.org/10.1105/tpc.111.087189) PMID: [21742987](https://pubmed.ncbi.nlm.nih.gov/21742987/)
25. Kazama T, Itabashi E, Fujii S, Nakamura T, Toriyama K. Mitochondrial ORF79 levels determine pollen abortion in cytoplasmic male sterile rice. *Plant J.* 2013; 85: 707–716. doi: [10.1111/tpj.13135](https://doi.org/10.1111/tpj.13135) PMID: [26850149](https://pubmed.ncbi.nlm.nih.gov/26850149/)
26. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet.* 2007; 52: 267–274. doi: [10.1007/s00294-007-0161-y](https://doi.org/10.1007/s00294-007-0161-y) PMID: [17957369](https://pubmed.ncbi.nlm.nih.gov/17957369/)
27. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinfo.* 2000; 16: 944–945. PMID: [11120685](https://pubmed.ncbi.nlm.nih.gov/11120685/)
28. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001; 305: 567–580. doi: [10.1006/jmbi.2000.4315](https://doi.org/10.1006/jmbi.2000.4315) PMID: [11152613](https://pubmed.ncbi.nlm.nih.gov/11152613/)
29. Zeltz P, Kadowaki K, Kubo N, Maier RM, Hirai A, Kossel H. A promiscuous chloroplast DNA fragment is transcribed in plant mitochondria but the encoded RNA is not edited. *Plant Mol Biol.* 1996; 31: 647–656. PMID: [8790296](https://pubmed.ncbi.nlm.nih.gov/8790296/).