

## RESEARCH ARTICLE

## Open Access

# Extensive structural variations between mitochondrial genomes of CMS and normal peppers (*Capsicum annuum* L.) revealed by complete nucleotide sequencing

Yeong Deuk Jo<sup>1,2</sup>, Yoomi Choi<sup>1</sup>, Dong-Hwan Kim<sup>3</sup>, Byung-Dong Kim<sup>1</sup> and Byoung-Cheorl Kang<sup>1\*</sup>

## Abstract

**Background:** Cytoplasmic male sterility (CMS) is an inability to produce functional pollen that is caused by mutation of the mitochondrial genome. Comparative analyses of mitochondrial genomes of lines with and without CMS in several species have revealed structural differences between genomes, including extensive rearrangements caused by recombination. However, the mitochondrial genome structure and the DNA rearrangements that may be related to CMS have not been characterized in *Capsicum* spp.

**Results:** We obtained the complete mitochondrial genome sequences of the pepper CMS line FS4401 (507,452 bp) and the fertile line Jeju (511,530 bp). Comparative analysis between mitochondrial genomes of peppers and tobacco that are included in Solanaceae revealed extensive DNA rearrangements and poor conservation in non-coding DNA. In comparison between pepper lines, FS4401 and Jeju mitochondrial DNAs contained the same complement of protein coding genes except for one additional copy of an *atp6* gene ( $\psi atp6-2$ ) in FS4401. In terms of genome structure, we found eighteen syntenic blocks in the two mitochondrial genomes, which have been rearranged in each genome. By contrast, sequences between syntenic blocks, which were specific to each line, accounted for 30,380 and 17,847 bp in FS4401 and Jeju, respectively. The previously-reported CMS candidate genes, *orf507* and  $\psi atp6-2$ , were located on the edges of the largest sequence segments that were specific to FS4401. In this region, large number of small sequence segments which were absent or found on different locations in Jeju mitochondrial genome were combined together. The incorporation of repeats and overlapping of connected sequence segments by a few nucleotides implied that extensive rearrangements by homologous recombination might be involved in evolution of this region. Further analysis using mtDNA pairs from other plant species revealed common features of DNA regions around CMS-associated genes.

**Conclusions:** Although large portion of sequence context was shared by mitochondrial genomes of CMS and male-fertile pepper lines, extensive genome rearrangements were detected. CMS candidate genes located on the edges of highly-rearranged CMS-specific DNA regions and near to repeat sequences. These characteristics were detected among CMS-associated genes in other species, implying a common mechanism might be involved in the evolution of CMS-associated genes.

\* Correspondence: [bk54@snu.ac.kr](mailto:bk54@snu.ac.kr)<sup>1</sup>Department of Plant Science, Plant Genomics and Breeding Institute, and Institute of Green BioScience and Technology, Seoul National University, Seoul 151-921, South Korea

Full list of author information is available at the end of the article

## Background

Mitochondrial genomes of higher plants are clearly different from their animal counterparts and from plastid genomes in terms of evolutionary dynamics of genome structure [1,2]. Although the rate of synonymous substitution in plant mitochondrial DNA (mtDNA) is 50–100 times and three times lower than in vertebrate mtDNA and plant plastid DNA, respectively, structural variations including changes in gene order, rearrangement, genome expansion and shrinkage, and incorporation of foreign DNAs are more common in plant mitochondria compared to the others [3-5]. The complexity of the plant mtDNA structure has been attributed to existence of a reservoir of low-copy-number subgenomic mtDNA molecules suppressed via nuclear control as well as the presence of repeat sequences dispersed throughout the genome that can mediate recombination [6-8].

Structural variations in mtDNA are associated with several mutant phenotypes such as cytoplasmic male sterility (CMS) and variegated-leaf phenotypes [9-11]. CMS has been studied in many crop species because of its economic importance for hybrid seed production. Most identified CMS-associated genes are novel chimeric open reading frames (ORFs) generated by fusion of several sequence segments due to rearrangement of the mitochondrial genome [9,12]. Although CMS-associated genes in different crop species do not show significant similarity in their sequences, most share several features such as possession of a transmembrane domain and co-transcription with normal mitochondrial genes, which often encode ATP-synthase or cytochrome C oxidase subunits [9,13-15]. The detailed mechanisms of how these genes originated remain unknown, however.

Complete sequencing and comparative analysis of mtDNA of normal and CMS lines has been performed in several crop species including sugar beet (Owen-type CMS), maize (CMS-T, CMS-S, CMS-C), wheat (K-type CMS), rice (CW- and LD-type CMS), rapeseed (pol- and nap-type CMS), and radish (Ogura- and DCGMS-type CMS), revealing structural variation in mtDNA and identifying genes responsible for CMS [16-24]. These studies showed that mitochondrial genome structures in lines exhibiting CMS are extensively rearranged compared to those of fertile lines, whereas gene contents are mostly conserved [12]. For example, in sugar beet, normal and CMS lines have mitochondrial genomes composed of different arrays of fourteen sequence blocks that are syntenic between the two genomes [23]. Recently, Tanaka et al. [24] showed that the mitochondrial genome of a radish with CMS, Ogura cytoplasm, has a large CMS-specific DNA region in addition to syntenic block sequences. This region, containing the CMS-associated *orf138*, was postulated to be inserted into the fertile mitochondrial genome by recombination via inverted repeat sequences located

on its borders. However, the origin of the CMS-associated gene and other CMS-specific sequences in this region are still unknown.

CMS has been widely used in hybrid seed production in chili peppers. Only a single source of cytoplasm has been reported to be responsible for CMS [25]. Kim et al. [14,26] found two candidate CMS-associated genes, *orf456* and  $\psi$ *atp6-2*. The *orf456* gene fused with a mitochondrial target sequence induced male sterility in transgenic Arabidopsis [14]. In later studies, it was shown that the *orf456* gene exists as a longer *orf* named *orf507* [27]. Further analysis of the function of ORF507 protein showed that the interaction of ORF507 with an ATP-synthase subunit (ATP synthase 6 kDa subunit) may cause impaired ATP synthase activity in CMS cytoplasm [28]. The  $\psi$ *atp6-2* gene is the truncated form of *atp6-2* that was generated by rearrangement in the 3' region of *atp6-2*. Differences in the transcription pattern of  $\psi$ *atp6-2* between male-sterile and restorer lines demonstrated a possible association of this gene with CMS [26]. However, complete sequence analyses of mitochondrial genomes to elucidate CMS-specific mtDNA structures and their evolutionary history has not been performed in *Capsicum*.

In this study, we first report the complete mitochondrial genome sequences for *Capsicum*. Comparative analysis between CMS and normal mitochondrial genomes provides insights into the evolution of mitochondrial genome structure associated with CMS in plants.

## Methods

### Plant materials

A pepper CMS line, 'FS4401' (*S/rff*), and a restorer line, 'Jeju' (*N/R/Rf*), which are known to contain CMS and normal cytoplasm, respectively, were provided by Monsanto Korea. 'FS4401' is a breeding line containing a natural CMS cytoplasm that has been used as the stable CMS source in seed companies in Korea and 'Jeju' is a landrace in Korea. For each pepper line, approximately 3,000 seedlings were grown in dark conditions for twenty days and harvested to isolate mitochondria.

### Mitochondrial DNA extraction

The method described by Millar et al. [29] and modified by Kim [14] was used for mitochondrial DNA extraction. Seedlings were homogenized using a mortar with isolation buffer consisting of 0.3 M mannitol, 50 mM Tris-HCl, 3 mM EDTA, 1 mM 2-mercaptoethanol, 0.1% BSA, 1% PVP, and protease inhibitor cocktail (Roche Applied Science, Indianapolis, USA) and adjusted to pH 7.5 with KOH. Homogenized tissue was filtered with one layer of Miracloth and four layers of cheesecloth. Two rounds of centrifugation at 2,000 g for 10 min were then performed to remove cell debris and larger organelles in cells. The supernatant subsequently was centrifuged at

15,000 g for 10 min to obtain a crude mitochondrial pellet. The pellet was gently resuspended with a painter's brush in isolation buffer without PVP, adjusted to 10 mM MgCl<sub>2</sub> and treated with DNase I (50 µg/ml) for one hour to degrade nuclear DNA. The sample was adjusted to 20 mM EDTA and centrifuged at 15,000 g for 10 min. The pellet was gently resuspended in 500 µL buffer II (0.3 M sucrose, 0.05 M Tris-HCl, 0.02 M EDTA, 0.1% BSA, pH 7.5) with a painter's brush. After resuspension, the sample was layered on a 30-mL Percoll cushion (28% Percoll, 0.3 M sucrose, 0.05 M Tris-HCl, 0.02 M EDTA, pH 7.5) and centrifuged at 40,000 g for 90 min. The yellowish mitochondrial ring in the middle of the cushion was collected. The mitochondrial fraction was rinsed with washing buffer (0.3 M mannitol, 50 mM Tris-HCl, 1 mM EDTA, pH 7.5) using three rounds of centrifugation at 15,000 g for 10 min. Mitochondrial DNA was extracted following the method described by Kim [30].

#### DNA sequencing

mtDNA sequencing was performed using the GS-FLX system (Roche Applied Science, Indianapolis, USA) and the resultant single-read sequences were assembled by Newbler Assembler Software Version 2.0 (454 Life Sciences, Branford, USA) in the National Instrumentation Center for Environmental Management (NICEM, Seoul, Republic of Korea) (Additional file 1).

#### Sequence assembly

Contigs were further assembled using the following strategies. First, analysis using the Basic Local Alignment Search Tool (BLAST) was performed against the GenBank nucleotide database (<http://www.ncbi.nlm.nih.gov>) to obtain mitochondrial contig sequences longer than 1 kb. The contig sequences that contained significant matches to known mtDNA sequences from other species were used for the next analyses. In the second step, a DNA library in which the average size of inserts was about 3 kb in length was constructed and the end sequences of inserts were analyzed using the ABI3700 sequencing system (Applied Biosystems, Foster City, USA) in NICEM (Seoul, Republic of Korea; Additional file 1). The mate-pair information of insert end sequences obtained by ABI sequencing was used to order contig sequences. In the third step, primers were designed from end sequences of each contig and all possible combinations of primers were tested by PCR analysis to identify connected contigs. If a gap sequence obtained from PCRs contained only plastid-derived sequence, the gap was considered to be obtained due to contamination with plastids during mitochondrial DNA preparation, not because of the existence of subgenomic mtDNA molecules. Through this step, gaps could be closed and

the number of contigs was reduced. In the fourth step, genome walking from the ends of each assembled scaffold sequence was conducted using the GenomeWalker™ universal kit (Clontech, Mountain View, USA) according to the manufacturer's instructions. In the fifth step, LA-PCRs were performed to fill remaining gaps between scaffolds using TaKaRa LA Taq™ (TaKaRa, Shiga, Japan). Finally, all information obtained by the stepwise approach was used to construct a master circle model that contained at least one copy of every mtDNA contig. The contig sequences that could be connected to multiple other contig sequences were regarded as repeated sequences. These sequences could be contained in a master circle two or more than two times. Insertions of repeated sequence were validated by PCR with primers designed from flanking regions of repeated sequences.

#### Accession numbers of mitochondrial genome sequences

Complete mitochondrial genome sequences of FS4401 (CMS) and Jeju (male-fertile) have been deposited in the GenBank nucleotide sequence database under the accession numbers of KJ865409 and J865410, respectively.

#### Screening a CM334 bacterial artificial chromosome library

BAC clones containing *cox2* or *atp6* was screened from a 12× BAC library of CM334, which is a Mexican land-race of chili pepper (*C. annuum* L.) containing normal cytoplasm [31]. The *cox2* and *atp6* genes of CM334 were amplified from total DNA of CM334 using primer sets designed from sugar beet mitochondrial DNA (GenBank accession number: BA000009). The amplicons were labeled and used as the probes for BAC library screening based on hybridization as described by Yoo et al. [31]. The sequences of the selected BAC clones containing both *cox2* or *atp6* was analyzed by 12× Shot-gun sequencing, which was carried out in NICEM (Seoul, Republic of Korea).

#### Gene annotation and characterization of ORFs

The protein and rRNA genes on mtDNA sequences were identified using BLAST with the nucleotide and protein database in GenBank (<http://www.ncbi.nlm.nih.gov>). The tRNA genes were identified using the tRNA scan-SE program (<http://lowelab.ucsc.edu/tRNAscan-SE/>). ORFs that were predicted to encode hypothetical proteins longer than 100 amino acids were screened using custom-made Perl scripts. The presence of a transmembrane domain in each hypothetical protein was predicted using TMHMM server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>).

#### Sequence comparison and repeat sequence analysis

Alignment between target sequences was performed using the BLASTN algorithm (<http://blast.ncbi.nlm.nih.gov/>). Pools of repeated sequences were obtained by analysis

using BLASTN in which a given target sequence was used as both query and subject sequence. The alignments that met the criteria to be syntenic sequence blocks or repeated sequences in terms of length and similarity were isolated and visualized as Scalable Vector Graphics using custom-made Perl scripts followed by manual modification.

## Results

### Assembly of complete mitochondrial genome sequences

We obtained contigs containing most mitochondrial genome information for a pepper CMS line (FS4401) and a fertile line (Jeju) by sequencing their mitochondrial genomes using the 454 GS-FLX system. However, it appeared that too many mtDNA contigs (>1 kb) were obtained for each line when the high coverage of sequencing (>100×) was considered (Additional file 1). The reasons for this were revealed in the process of further assembly of contigs and analysis of gap sequences. Firstly, contamination with plastid DNA hampered sequence assembly at the positions of mitochondrial genomes where plastid-derived sequences were located. Secondly, ends of large repeated sequences remained unconnected due to the short length of individual reads in the 454 GS-FLX system (Additional file 1). Therefore, we constructed a DNA library containing inserts averaging 3 kb in length and analyzed the end sequences of inserts using ABI3700 for ordering of contigs (Additional file 1). In addition, we performed PCR analysis and genome walking from contig ends to test all of the possible combinations of large repeated sequences with other contig sequences. The final circular molecules for complete mitochondrial genomes included every mtDNA contig

longer than 1 kb at least one time and were consistent with all of the results produced during sequence assembly process.

Meanwhile, screening of a bacterial artificial chromosome (BAC) library of pepper line 'CM334' with normal (fertile) cytoplasm using *cox2* and *atp6* as probes resulted in the isolation of one BAC clone containing both *cox2* and *atp6*. The complete sequence of this BAC clone (74,615 bp) was obtained by shotgun sequencing and assembly.

### Comparative analysis of general features and sequence contents between mitochondrial genomes

General features of mitochondrial genomes were compared between FS4401, Jeju, and tobacco (Table 1). Tobacco is the only species in the Solanaceae for which complete mtDNA sequence has been reported [32]. The complete mitochondrial genomes of FS4401 and Jeju were 507,452 and 511,530 bp in length, respectively, significantly larger than that of tobacco (Table 1). The proportion of protein-coding sequences was similar between pepper mtDNAs, at 7.9% and 7.7%, respectively, while it was higher in tobacco due to repetition of some genes (*trnM*, *rrn26*, *nad2a*, *sdh3*) and smaller genome size. The ratio of chloroplast-derived sequences was slightly higher in Jeju (12.7%) than FS4401 (11.8%), and those values were about 2.5 times higher than in tobacco (4.5%). Tobacco had a larger proportion of repeated sequences, mainly due to containing larger repeat sequences. The total length of repeat sequences was longer in Jeju than in FS4401. The contents of genes encoding proteins and rRNAs were the same between Jeju and tobacco, whereas FS4401 had an additional copy of the *atp6*

**Table 1 General features of mitochondrial genomes of two pepper lines and one tobacco line**

Features	FS4401	Jeju	<i>Nicotiana tabacum</i>
Genome size (bp)	507,452	511,530	430,597
GC content (%)	44.5	44.6	45.0
Coding sequences (bp) <sup>a</sup>	40,085 (7.9%)	39,524 (7.7%)	43,642 (10.1%)
Plastid-derived sequences (bp) <sup>b</sup>	59,873 (11.8%)	64,815 (12.7%)	19,492 (4.5%)
Repeated sequences (bp) <sup>c</sup>	42,505 (8.4%)	70,122 (13.7%)	73,511 (17.1%)
Gene content (number)	66	64	61
Protein coding genes <sup>d</sup>	38	37	37
rRNAs	3	3	3
tRNAs <sup>e</sup>	25 (13)	24 (12)	21 (9)

<sup>a</sup>All of the copies of duplicated genes were included.

<sup>b</sup>Complete sequences of chloroplast genomes of FS4401 [33] and *Nicotiana tabacum* [34] were used to screen for plastid-derived sequences in mitochondrial genomes of FS4401/Jeju and *N. tabacum*, respectively, based on the BLASTN algorithm. The value for *N. tabacum* was different from that of Sugiyama et al. [35] probably due to differences in methodology used to isolate plastid-derived sequence.

<sup>c</sup>Sequences longer than 100 bp and showing similarity higher than 95% between copies were considered repeated. In case repeated sequence units were overlapping, nucleotides that were included in repeated sequences at least one time were counted without repetition.

<sup>d</sup>*rp110*, which was reported to be a conserved mitochondrial gene by Kubo and Arimura [32] was added to the list of genes described by Sugiyama et al. in tobacco [35].

<sup>e</sup>The number of tRNA genes included in plastid-derived sequences is given in parenthesis.

The tobacco line is 'Bright Yellow 4' for which the mitochondrial genome was analyzed by Sugiyama et al. [35].

gene, named  $\psi atp6-2$  [26]. The number of genes encoding tRNAs were different between mtDNAs. FS4401 had one and four additional tRNA genes of chloroplast origin compared to Jeju and tobacco, respectively.

Sequence alignment analysis showed that most sequence content was shared between two mtDNAs of pepper. 95.1% of the sequence of FS4401 could be aligned with Jeju and 98.0% of Jeju with FS4401. In comparative analysis with tobacco sequence, only about 46.3% and 45.4% of each genome could be aligned with tobacco mtDNA, respectively (Additional file 2).

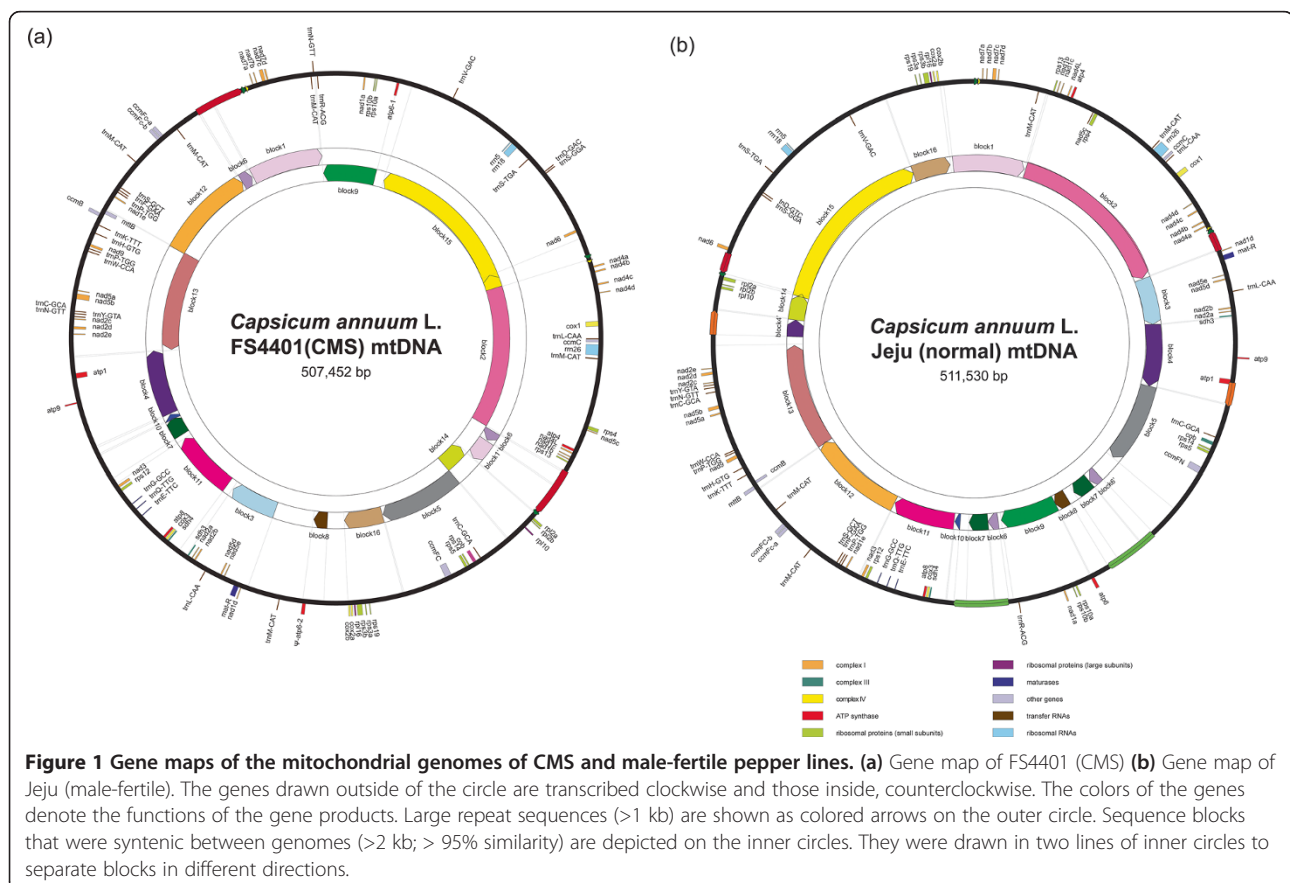
### Distribution of sequence blocks syntenic between pepper and tobacco mitochondrial genomes

Sequence blocks between mitochondrial genomes of the two pepper lines and the tobacco line were defined based on similarity higher than 95% and matching length longer than 2 kb. These sequence blocks were then localized on each genome (Figure 1; Additional file 3). In the alignment between FS4401 and tobacco, a total of 33 sequence blocks that covered 116,598 bp of the FS4401 mitochondrial genome were detected. The comparison between Jeju and tobacco accounted for 122,124 bp of the Jeju mitochondrial genome and revealed the same 33 blocks as well as two additional sequence blocks including

one located downstream of *cox2* and another on which *nad9*, *trnP*, *trnW*, and *ccmB* were localized. Most of the syntenic sequence blocks contained clusters of genes, resulting in high conservation of the gene clustering pattern between pepper and tobacco. However, four of the tobacco gene clusters, including *atp9-rps13-nad1bc*, *nad4-rps1-nad5ab*, *nad3-nad1a*, and *rps4-nad6*, were not conserved in FS4401 and Jeju. The order of the syntenic blocks was extremely different among the mitochondrial genomes, demonstrating extensive rearrangement between non-coding regions.

### Gene contents and localization on mitochondrial genomes

We next annotated and localized the genes with known functions on the FS4401 and Jeju mitochondrial genomes (Figure 1). The protein-coding genes were classified according to the functions of proteins; the 37 protein-coding genes shared by FS4401 and Jeju included nine genes for complex I proteins (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*), two for complex II (*sdh3*, *sdh4*), one for complex III (*cob*), three for complex IV (*cox1*, *cox2*, *cox3*), five for ATP synthase subunits (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*), ten for ribosomal proteins (*rpl2*, *rpl5*, *rpl10*, *rpl16*, *rps3*, *rps4*, *rps10*, *rps12*,



*rps13*, *rps19*), four for proteins involved in cytochrome c biogenesis (*ccmB*, *ccmC*, *ccmFc*, *ccmFN*), one for maturase (*matR*), and one for a protein translocation system subunit (*mttB*). In addition to these genes, FS4401 had another copy of the *atp6* gene ( $\psi$ *atp6-2*). Both FS4401 and Jeju contained three rRNA genes (*rrn5*, *rrn18*, *rrn26*), while FS4401 contained one additional tRNA gene compared to Jeju.

Comparison of protein-coding gene sequences between FS4401 and Jeju revealed six genes exhibiting polymorphism in their nucleotide sequences (Table 2). Sequence polymorphisms in *atp4*, *atp8*, *rpl2*, *sdh3*, and *atp6* resulted in a change in protein sequence, whereas a synonymous substitution was detected in *matR*. Changes in the gene product length were predicted for *atp4*, *atp8*, and *rpl2* due to in-frame indels, whereas length polymorphism in *atp6* was attributable to a structural rearrangement (Figure 2). The *atp6* gene in Jeju showed high similarity with *atp6-1* of FS4401 over three-fifths of the gene, spanning 800 bp including the highly conserved region of *atp6* genes [26], whereas the sequence upstream of the conserved region could be aligned with the additional copy of *atp6* in FS4401,  $\psi$ *atp6-2*. The nucleotide sequence and the structure of the gene-flanking regions of *atp6* of Jeju were different from any *atp6* reported in previous research where two copies of *atp6* genes were present even in normal cytoplasm [26]. Comparison of the genes carrying nucleotide sequence

polymorphism between the pepper species to the corresponding genes in tobacco showed that the polymorphic sites of *matR* and *sdh3* in FS4401 were the same as those of tobacco, indicating that nucleotide substitutions in these genes were probably not related to sequence alteration during the evolution of CMS. By contrast, differences in the sequences of *atp4*, *atp8*, *rpl2*, and *atp6* of FS4401 relative to those of both Jeju and tobacco pointed to the possibility that variations in these genes might underlie CMS (Table 2).

A large number of genes were located close to each other, forming gene clusters that might be co-transcriptional units. In total, sixteen clusters were detected in FS4401 and Jeju, including *rps10ab-nad1a*, *rrn18-rrn5*, *trnD-trnS*, *trnM-rrn26-ccmC-trnL*, *rps4-nad5c*, *rps13-nad1bc-nad4L-atp4*, *rpl2ab-rpl10*, *rpl5-rps14-cob-trnC*, *rps19-rps3ab-rpl16-cox2ab*, *nad1d-matR*, *sdh3-nad2ab*, *atp8-cox3-sdh4*, *nad3-rps12*, *trnC-trnN-trnY-nad2cde*, *nad9-trnP-trnW*, *trnS-trnF-trnP-nad1e* (Figure 1, Additional file 3). Although numerous rearrangements between the two pepper mitochondrial genomes were detected, the clustering pattern of these genes was highly conserved.

#### Rearrangements of genome structure between CMS and normal mtDNA

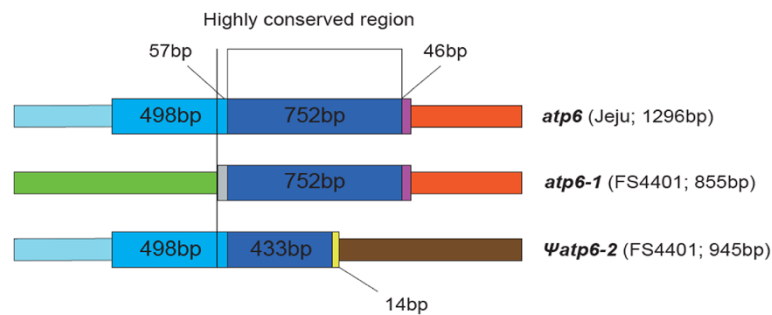
A total of sixteen syntenic blocks were localized on each genome (Figure 1, Additional files 4 and 5). The sizes of blocks ranged from 2.9 kb (block 10) to 78.9 kb (block

**Table 2 Differences between Jeju and FS4401 in sequences of known genes**

Genes in Jeju	Gene length in Jeju	Polymorphism in FS4401	Corresponding sequence in tobacco
<i>matR</i>	1977	904 gcA (A) → 904 gcG (A)	904 gcG (A)
<i>atp4</i>	597	16 acGAATATGCAG (TNMQ) → 16 acg (T)	16 acGAATATGCAG (TNMQ)
<i>atp8</i>	462	178 ccCAACAGTTTg (PNSL) → 178 ccg (P)	178 ccCAACTGTTTg (PNCL)
<i>rpl2</i>	999	337 ccCGGGAAGGGgat (PGKGD) → 337 ccgat (PD)	337 ccCGGGAAGGGgat (PGKGD)
<i>sdh3<sup>a</sup></i>	333	178 tCTtc (FF) → 178 tCTtc (SF)	178 tCTtc (SF)
<i>atp6</i>	1296	$\psi$ <i>atp6-2</i> 283 gGt (G) → gCt (A) 316 ACa (T) → CAa (Q) 454 aaAGaa (KE) → aaCCaa (NQ) no similarity in downstream of 931 <sup>th</sup> bp due to DNA rearrangement <i>atp6-1</i> no similarity in upstream of 497 <sup>th</sup> bp due to DNA rearrangement	Higher similarity with <i>atp6-1</i> in FS4401

<sup>a</sup>The SNP polymorphism can potentially be eliminated if the plant mitochondrial C-to-U RNA editing occur.

The pattern of base changes and corresponding amino acid changes is described. The polymorphic site (the number indicates the position of the first nucleotide), with the polymorphic nucleotide capitalized, and the corresponding amino acid in parentheses, is given to the left of the arrow for Jeju and to the right of the arrow for FS4401.



**Figure 2 Structure of the *atp6* gene copies in Jeju and FS4401.** The sequences correspond to gene-coding region are drawn as the wider rectangles and the upstream or downstream regions as the narrower bars. The sequence units that show high similarity (>99%) to each other and included in the same category of sequence characteristics (non-coding region/coding region, *atp6* region showing high conservation/poor conservation among plant taxa) are depicted as the same color. The overall scheme of figure was adopted from Kim et al. [26].

15; Additional files 4 and 5). Block 6 and a part of block 1, block 7 and parts of blocks 4 and 6 were duplicated in FS4401 and Jeju, respectively (Figure 1, Additional files 4 and 5). The mitochondrial genome of FS4401 had a total of eighteen junctions between blocks, whereas Jeju had nineteen. In FS4401, sequences overlap between blocks were detected in seven junctions while no matching sequences between blocks were found at eleven junctions (Additional file 4). Among sequences located between blocks, the sequences between blocks 16 and 8, and between blocks 8 and 3 were noticeable, accounting for 70.5% of the unique sequence of FS4401 mtDNA. The *orf507* and  $\psi$ *atp6-2* genes known to be responsible for CMS [14,26] were localized at the 5' junctions of the sequence segments between blocks 16 and 8, and blocks 8 and 3, respectively (Figure 3). In Jeju, a total of nine sequences overlapped by adjacent blocks and ten sequences between blocks were detected (Additional file 5). The sequence between blocks 5 and 6' contained the largest portion (37.7%) of sequences unique to Jeju. Most of the large sequence segments between blocks remained specific to each mtDNA in alignment analysis with less strict criteria (BLASTN default) and also could not be aligned with tobacco mtDNA sequence (Figure 3). A sequence segment between blocks 13 and block 4' in Jeju contained a chloroplast-derived sequence specific to Jeju (Figure 3). Such sequence overlaps between syntenic blocks in one mitochondrial genome corresponded to repeated sequences located at the edges of blocks in the other genome. The sizes of this kind of repeated sequence varied from 6 to 7,413 bp (Additional files 4 and 5).

End points of several sequence blocks syntenic between FS4401 and Jeju were located very close (<2 kb) to the edges of blocks shared by each pepper line and tobacco analyzed using the same criteria (>2 kb, > 95%; Additional file 3). There were eight and seven block end

regions included in this category in FS4401 and Jeju, respectively. At least two rearrangement events, including one during speciation between tobacco and pepper and another during evolution of pepper mitochondrial genomes, were expected in these regions. No single sequence blocks that could cover adjacent ends of blocks syntenic between FS4401 and Jeju were detected in the alignment with tobacco mtDNA except for two syntenic sequence block that were specific to the alignment between Jeju and tobacco. This syntenic block connected the 3' junction of block 16, which was downstream of *cox2*, and the 5' junction of block 1 (Additional file 3).

#### ORFs unique to each mitochondrial genome

We next identified novel open reading frames predicted to encode proteins longer than 100 amino acids in length in the mitochondrial genomes. A total of 155 and 142 ORFs (excluding ORFs for known genes) were detected in FS4401 and Jeju, respectively. Comparative analysis of these ORFs showed that 45 and 30 ORFs had polymorphisms with ORF counterparts or were specific to one genome in FS4401 and Jeju, respectively (Additional file 6). FS4401 mtDNA contained 33 ORFs with SNPs or length polymorphism and 12 ORFs that were chimeric, whereas Jeju mtDNA contained 22 polymorphic and 8 chimeric ORFs. When the localization of ORFs that were chimeric or unique to FS4401 was investigated to search for candidate CMS-associated genes, seven ORFs including *orf100d*, *orf102l*, *orf108a*, *orf119c*, *orf141*, *orf300* and *orf507* were found to be close (<2 kb) to the edge of sequence blocks syntenic between FS4401 and Jeju (>2 kb; > 95% similarity). When we searched for ORFs located near (<2 kb) repeat sequences (>100 bp; > 95% similarity between copies) and containing putative transmembrane domains, six (*orf102l*, *orf119c*, *orf262*, *orf300*, *orf338*, *orf507*) and three (*orf262*, *orf300*, *orf507*) ORFs were identified, respectively





### Structure of sequences around *orf507* and $\psi$ *atp6-2*

The genomic region around *orf507* and  $\psi$ *atp6-2* was found to be highly specific to the mitochondrial genome of the line showing CMS in this study as well as previous researches [14,26]. Therefore, the structure of this DNA region was analyzed in detail (Figure 4). The *orf507* gene was located downstream of *cox2*, and  $\psi$ *atp6-2* was about 12 kb from *orf507* in FS4401. In Jeju, however, not only was *orf507* absent, but also *cox2* and *atp6* were distantly located from each other, implying that rearrangements occurred between the two genes. In FS4401, two repeated sequences (R19, Ra) and *orf507*, which were overlapped subsequently by small number of nucleotides, were located downstream of the *cox2*. Sequences showing high similarity to Ra were detected in FS4401, Jeju and tobacco at the 5' upstream region of the *nad9* gene. The part of *orf507* not covered by Ra showed no similarity to other sequences in FS4401, Jeju, or tobacco mtDNA, nor to any sequences registered in GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). However, in Jeju, downstream region of *cox2* consisted of sequence elements, CS2, R21, and CS1 that were located on different regions in FS4401. DNA region around *atp6-2* (or  $\psi$ *atp6-2*) also showed extensive DNA rearrangements. Upstream and 5' regions of  $\psi$ *atp6-2* showed high similarity with the corresponding region in Jeju. However, a repeated sequence Rb and CS2-R21 sequence element composed FS4401-specific DNA structure on downstream of  $\psi$ *atp6-2*. The conserved part of  $\psi$ *atp6-2*, Rb, and CS2 sequence elements were overlapped subsequently by small number of nucleotides. Stretches of sequences present only in FS4401

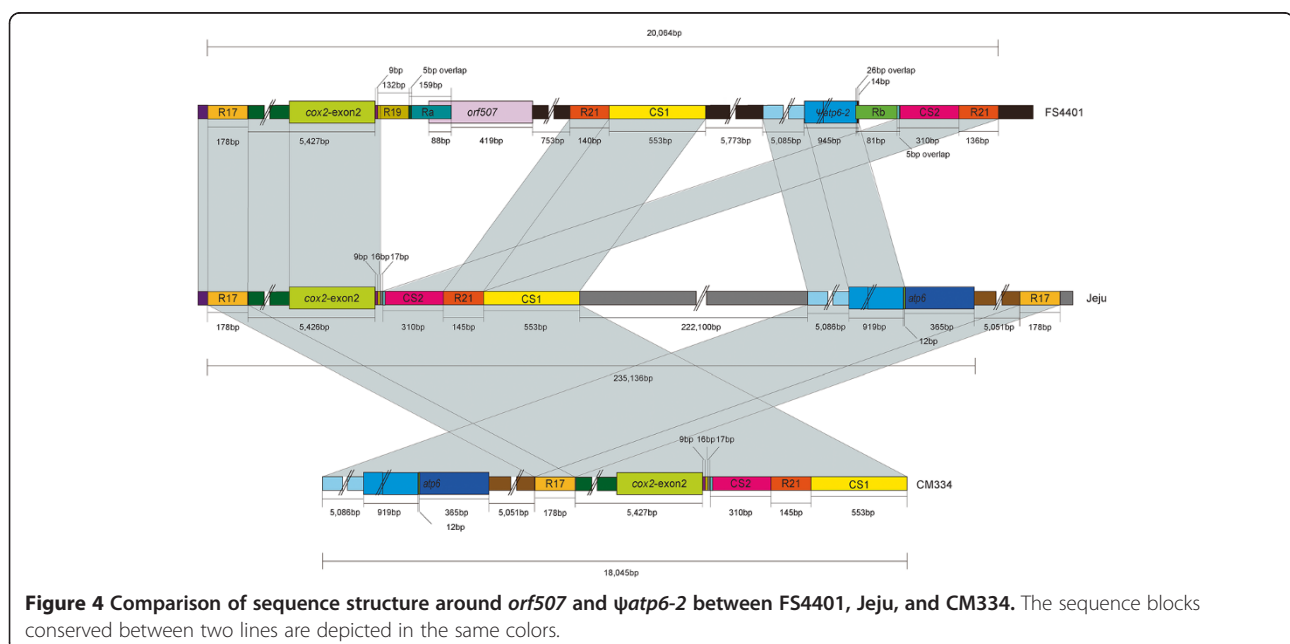
were also detected on the region between *orf507* and  $\psi$ *atp6-2* and the downstream of  $\psi$ *atp6-2*.

The short sequence elements that were detected on and  $\psi$ *atp6-2* in FS4401 were present as repeat sequence only in FS4401, implying that these sequences were duplicated during rearrangement (Additional file 7). In particular, R21 in Jeju was duplicated downstream of  $\psi$ *atp6-2* in FS4401 resulting in generation of a repeat pair around the  $\psi$ *atp6-2* gene. The *orf507* gene and other related sequence elements seemed to be inserted between *cox2* and R21 via multiple DNA rearrangements.

Comparison of mtDNA around *cox2* and *atp6* in Jeju with corresponding regions in CM334, which is a *C. annuum* landrace introduced from distant area (Mexico) and contains normal cytoplasm, showed that the DNA rearrangement involving R17 resulted in close proximity of *cox2* and *atp6* to each other although sequence contents flanking the two genes were highly conserved. In addition, the gene order of *cox2* and *atp6* was opposite in CM334 compared to FS4401 (Figure 4).

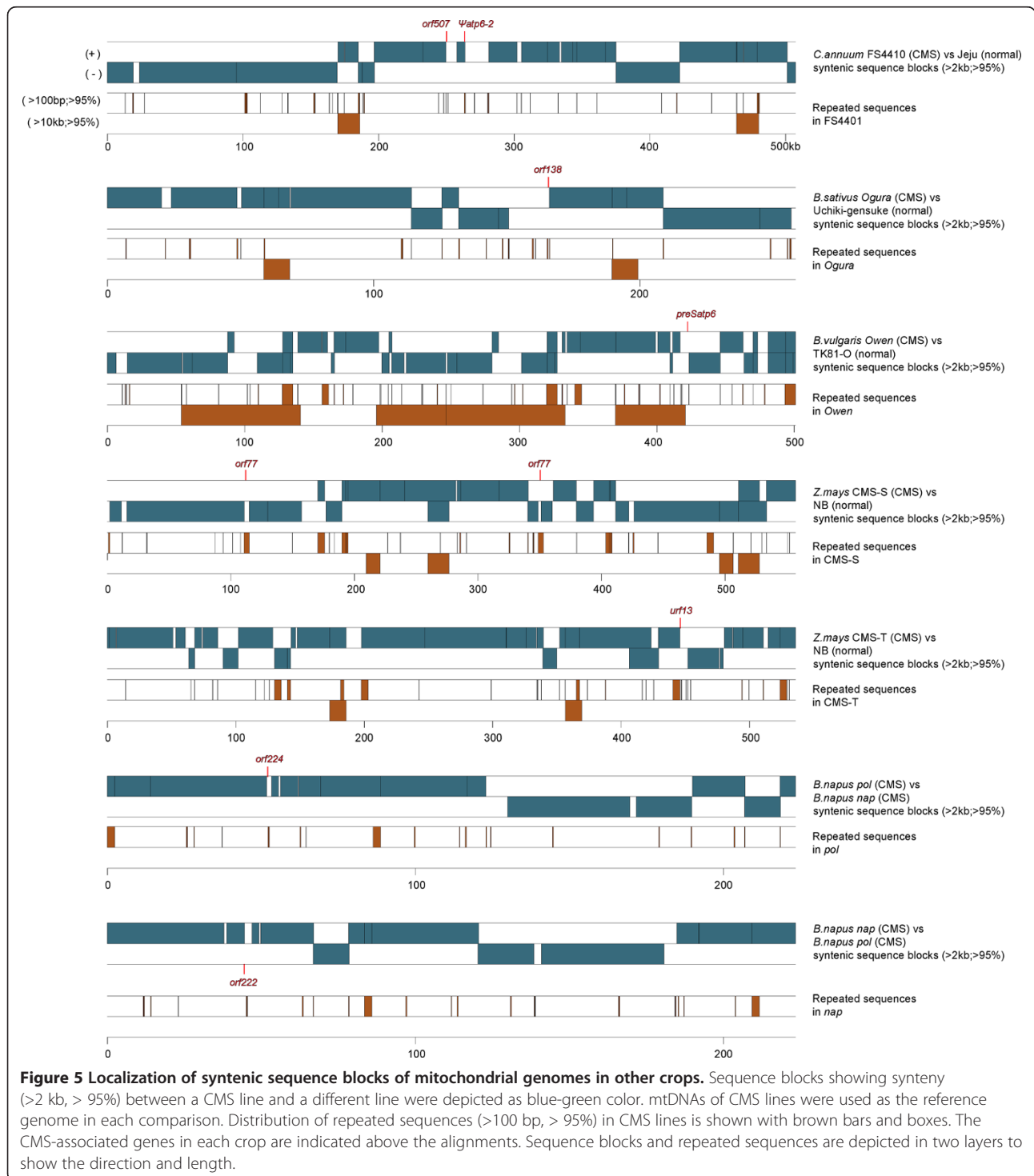
### DNA rearrangement pattern and localization of CMS-associated genes in mitochondrial genomes of other crop species

The rearrangement pattern of mitochondrial genomes was investigated in pepper and other crop species including *Brassica sativus*, *Beta vulgaris*, *Zea mays*, and *Brassica napus* for which complete sequences of a CMS-associated mitochondrial genome and at least one mitochondrial genome from different cytoplasm were available



and for which the genes responsible for CMS were identified (Figure 5). Alignment of CMS-associated mtDNA with available mtDNA from a different source for seven mitochondrial genomes identified numerous syntenic sequence blocks (>2 kb; > 95%), as reported by many other

studies [16,18,21-24]. All of the genes known to be associated with CMS were localized close to the edge of syntenic sequence blocks. In particular, CMS-associated genes in pepper and radish were near the end of long sequences located between syntenic blocks. Analysis of repeat



sequences (>100 bp; > 95%) showed that CMS genes were always located near or in the repeat sequences.

## Discussion

Here, we report the complete mitochondrial genome sequence of pepper (*C. annuum* L.). This represents only the second mitochondrial genome from the Solanaceae to be fully sequenced, following that of tobacco [35]. Therefore, the mtDNA sequence of pepper described herein is a valuable resource for studying the evolution of mitochondrial genomes in the Solanaceae. The contents and sequences of protein-coding genes were mostly conserved between tobacco and pepper mtDNA although small number of SNPs and indels were found in two and three genes, respectively. It was noticeable that the frequency of in-frame indel polymorphisms was higher in pepper than other crops such as rice [17] and radish [24]. However, the overall structure and the non-coding sequences were extensively changed, resulting in less than 50% coverage of pepper mitochondrial genome sequence by tobacco mtDNA. Similar widespread rearrangements within a plant family have been reported from comparative analysis of Arabidopsis and rapeseed mitochondrial genomes [19], in which only one-third of Arabidopsis and two-third of rapeseed mtDNA could be aligned to each other. By contrast, small DNA regions containing clusters of gene sequences were mostly protected from rearrangement events. The conservation of gene sequence clusters might reflect a requirement for co-regulation of gene expression on each cluster. However, rearrangements were detected even in the small number of gene clusters, including *atp9-rps13-nad1bc*, *nad4-rps1-nad5ab*, *nad3-nad1a*, and *rps4-nad6*, that have been reported to be putative co-transcribed units in tobacco [35]. In particular, co-transcription of gene cluster *nad3-nad1a* has been confirmed experimentally in tobacco [36,37], whereas we found that *nad3* and *nad1a* were incorporated into different clusters in pepper mtDNA. Therefore, a change in co-transcription units has resulted from DNA rearrangement during speciation or independent evolution of tobacco and pepper mitochondrial genomes after speciation.

Numerous rearrangements of mitochondrial DNA were also detected even in the comparison of CMS-associated and normal mtDNA within *C. annuum* species. Conservation of gene coding sequences and clustering patterns indicated that maintenance of clusters may be essential for normal expression of genes or those sequences in transcribed regions have characteristics that efficiently suppress rearrangement. However, multiple rearrangements occurring outside of gene clusters resulted in the fragmentation of alignment units between genomes. Several sequence blocks that were syntenic between genomes contained overlapping repeat sequences that might

mediate homologous recombination and result in genome rearrangement. However, many sequence blocks were connected with sequences unique to each genome or had overlapping sequences that were shorter than 50 bp which was known as the lower limit of homologous sequence length required for recombination that mediates double-strand break repair [7,38,39]. This might be explained by loss of larger repeats during evolution after rearrangements occurred or involvement of nonhomologous end-joining (NHEJ) and/or microhomology-mediated recombination. Comparative analyses on mitochondrial genomes of Arabidopsis ecotypes and mutants revealed that DNA rearrangement results from nonhomologous end-joining (NHEJ) and asymmetric recombination via intermediate-sized repeats followed by randomly occurring double-strand breaks of DNA [7,40]. Sequences lacking homology were joined by NHEJ, while asymmetric recombination is accompanied by repeat sequences longer than 50 bp [7]. Recombinations via microhomologous repeats (ranging from 6 to 31 bp) have been also reported in pearl millet and maize mutants showing nonchromosomal stripe (NCS) phenotype [41-43]. Microhomology-mediated break-induced replication (MMBIR) have been proposed as one of the mechanism for microhomology-mediated rearrangements in plastids and mitochondria [44,45].

A significant number of ends of sequence blocks syntenic between FS4401 and Jeju were located close to the junctions of sequence blocks syntenic between FS4401 and tobacco or Jeju and tobacco. Therefore, those regions experienced at least two rearrangement events within a very short distance: one between pepper and tobacco, and the other between CMS and normal pepper lines. Why recombination frequently occurs in specific regions is still unknown although localization of DNA cruciforms, localized melting due to high transcriptional activity, and stalling replication forks have been suggested as possible explanations [40,46]. Further investigation on a large number of mitochondrial genome sequences in diverse plant families is required to identify and characterize recombination hotspots.

The *orf507* and *ψatp6-2* genes are known to be associated with CMS in pepper, based on genetic and functional analyses [14,26,28]. Comparison between complete mtDNA sequences from CMS-associated and normal cytoplasm in this study reinforced these genes as CMS candidate genes. Although a large number of ORFs were specific to the CMS-associated mitochondrial genome, only one ORF (*orf300*; Figure 3) in addition to *orf507* and *ψatp6-2* had the typical characteristics shared by most CMS-associated genes in other species: formation of the ORF by novel DNA rearrangement [9], presence of a transmembrane domain [12], and localization close to a junction of syntenic sequence blocks and repeat sequences

(discussed below). However, the potential association of other screened *orfs* including *orf300* with CMS also needs to be investigated because discrepancies in cytoplasm types and haplotypes of markers based on *orf507* and  $\psi atp6-2$  have been reported in several germplasms [47,48], which may be due to incorrect identification of candidate CMS genes or to the existence of a different CMS source.

The genomic regions around *orf507* and  $\psi atp6-2$  were clearly distinguished from other DNA regions because they were included among the largest sequence fragments highly specific to the CMS line and matched no other known sequences. Recently, similar results were reported for radish *Ogura* type cytoplasm in which the CMS-associated gene *orf138* was located on an edge of the largest genomic region unique to the CMS line [24]. Although the insertion of the DNA region containing *orf138* could be demonstrated to result from homologous recombination using a pair of inverted repeat sequences on the ends, the region around *orf507* and  $\psi atp6-2$  in pepper contained more complicated structure, hampering the elucidation of the mechanism by which the genomic structure arose. The origin of the 3' part of *orf507* and a large portion of the region around *orf507* and  $\psi atp6-2$  remains unknown. One possible explanation for how these sequences came to be specifically present in the CMS line might be substoichiometric shift (SSS), which has been reported in several plant species [6,41,49,50]. According to the SSS model, subgenomic molecules of mtDNA are present at very low copy number under normal conditions, in which recombination of intermediate-sized repeat sequences is suppressed, and if this recombination is activated (e.g., under certain conditions), these molecules can be efficiently amplified by recombination-dependent replication and maintained as the predominant form of subgenomic molecules even in subsequent generations [40]. In fact, small amounts of *orf507* and  $\psi atp6-2$  were detected by PCR even in fertile pepper lines [47]. Therefore, the subgenomic molecule containing CMS candidate genes that had been generated by rearrangements via microhomology-mediated recombination using short sequences overlapped between sequence elements (ranging from 5 to 40 bp; Figure 4) and/or NHEJ of sequence elements from diverse sources might be maintained at low copy number even in normal pepper lines. If the suppression of ectopic recombination is released under certain conditions, amplification of the CMS-specific DNA structure containing *orf507* and  $\psi atp6-2$  might occur by recombination-dependent replication via intermediate-sized repeat sequences around these regions. A pair of intermediate-sized repeats (R21) of which one copy is located downstream of *orf507* and the other downstream of  $\psi atp6-2$  might be candidates for mediating this procedure. However, prediction of

the precise mechanism of the rearrangement is limited by the lack of information on the evolutionary relationship between the CMS cytoplasm and a normal type cytoplasm in Jeju. In fact, the corresponding DNA region found in another type of normal cytoplasm from CM334 showed structural differences when compared to Jeju implying the presence of multiple cytoplasm types that have undergone different levels of rearrangement (Figure 4). Analyses of the mtDNA region containing *orf507-ψatp6-2* from different cytoplasms of pepper may provide clues to the detailed steps of the rearrangements. In addition, further analysis using high-coverage paired-end sequencing may facilitate identification of possible structures of subgenomic molecules to elucidate dynamic processes related to origin of CMS in *Capsicum*.

Considering the specific characteristics of the CMS-associated region in pepper, we performed analysis of the organization of syntenic sequence blocks, repeat distribution, and localization of CMS-associated genes in six additional CMS mitochondrial genomes from other species. In all of the cases, CMS genes were located at the edge of considerably long CMS-specific sequences between syntenic blocks and close to intermediate-sized repeat sequences or on the repeat sequence itself (e.g., CMS-S in maize). None of the CMS genes originated by the fusion of sequences that exist on predominant subgenomic molecules of male-fertile lines nor by small insertions or deletions on pre-existing sequences. These findings fit well with the notion that subgenomic structure containing CMS genes might originate from multiple rearrangements mediated by microhomology-mediated recombination or NHEJ to create novel DNA sequence regions and copy number increases due to recombination via adjacent repeat sequences. Proximity of pre-existing low copy-number CMS genes to intermediate-sized repeat sequences might be the prerequisite to ensure amplification of these sequences required for the induction of CMS as discussed by Davila et al. [7]. The close localization of CMS genes to syntenic sequence blocks might be due to the need for sequence elements required for transcription of a chimeric *orf*. In Arabidopsis, the majority of chimeric *orfs* were shown not to be transcribed [51]. This implies that utilization of promoters in conserved regions might be requisite for the transcription of *orfs*. These common features of genomic environment of CMS-associated genes can be clues to understand the evolution of CMS as well as provide a strategy to screen for unknown CMS-gene candidates by comparative genomics approaches.

## Conclusions

The complete mitochondrial genome sequences of pepper were obtained in CMS and male-fertile lines. A large

portion of the intergenic sequences in the pepper lines could not be aligned with the mitochondrial genome of *Nicotiana tabacum*, which is a member of the same family (Solanaceae), whereas sequences and clustering patterns of genes were largely conserved. In the comparison between mitochondrial genomes of CMS and male-fertile pepper lines, however, most genome sequences could be aligned although syntenic sequences were divided into eighteen sequence blocks that were generated by rearrangements in intergenic regions. The CMS candidate genes *orf507* and *ψatp6-2* were located on the edges of CMS-specific sequence segments that were between syntenic sequence blocks. The presence of many repeat sequences and connection of sequence segments overlapped each other by a few nucleotides implied that extensive rearrangements by homologous recombination and/or NHEJ might be involved in evolution and substoichiometric shift of this region. Extended investigation using CMS-associated genes identified in other species revealed that these genes are specifically localized near edges of CMS-specific DNA regions and intermediate or large-sized repeat sequences indicating the evolution of CMS-associated genes might involve the common mechanism.

## Additional files

**Additional file 1:** Sequencing and contig assembly results.

**Additional file 2:** Alignment of mitochondrial genomes of pepper and tobacco.

**Additional file 3:** Distribution of sequence blocks syntenic between FS4401 and tobacco or Jeju and tobacco (>2 kb; >95%) on FS4401, Jeju, and tobacco mitochondrial genomes.

**Additional file 4:** Localization of syntenic sequences blocks (>2 kb; > 95%) and size of gap or overlapping sequences between blocks on FS4401 mtDNA.

**Additional file 5:** Localization of syntenic sequence blocks (>2 kb; > 95%) and size of gap or overlapping sequences between blocks on Jeju mtDNA.

**Additional file 6:** mtDNA ORFs with polymorphism or unique to one pepper line.

**Additional file 7:** Repeated sequences around *orf507* and *ψatp6-2*.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

YDJ and YC performed research (material preparation, sequence analysis). DHK helped to draft the manuscript. BDK and BCK participated in its design and coordination of research and helped to draft the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This research was supported by the Golden Seed Project, Ministry of Agriculture, Food and Rural Affairs (MAFRA), Ministry of Oceans and Fisheries (MOF), Rural Development Administration (RDA) and Korea Forest Service (KFS) and a grant (Project No. 710001-07) from the Vegetable Breeding Research Center through the R&D Convergence Center Support Program, Ministry of Agriculture, Food and Rural Affairs (MAFRA) Republic of Korea.

## Author details

<sup>1</sup>Department of Plant Science, Plant Genomics and Breeding Institute, and Institute of Green BioScience and Technology, Seoul National University, Seoul 151-921, South Korea. <sup>2</sup>Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongseup 580-185, South Korea.

<sup>3</sup>Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA.

Received: 11 March 2014 Accepted: 20 June 2014

Published: 4 July 2014

## References

1. Andre C, Levy A, Walbot V: Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet* 1992, **8**:128–132.
2. Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K: Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci U S A* 2000, **97**:6960–6966.
3. Palmer JD, Herbon LA: Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol* 1988, **28**:87–97.
4. Palmer JD: Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet* 1990, **6**:115–120.
5. Wolfe KH, Li WH, Sharp PM: Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A* 1987, **84**:9054–9058.
6. Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S: Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 2001, **158**:851–864.
7. Davila JJ, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedje V, Xu YZ, Weigel D, Mackenzie SA: Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. *BMC Biol* 2011, **9**:64.
8. Small I, Suffolk R, Leaver CJ: Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* 1989, **58**:69–76.
9. Hanson MR, Bentolila S: Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 2004, **16**(Suppl):S154–S169.
10. Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA: Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc Natl Acad Sci U S A* 2003, **100**:5968–5973.
11. Zaegel V, Guermann B, Le Ret M, Andres C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P: The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in Arabidopsis. *Plant Cell* 2006, **18**:3548–3563.
12. Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T: Male sterility-inducing mitochondrial genomes: how do they differ? *Crit Rev in Plant Sci* 2011, **30**:378–400.
13. Ashutosh, Kumar P, Dinesh Kumar V, Sharma PC, Prakash S, Bhat SR: A novel *orf108* co-transcribed with the *atpA* gene is associated with cytoplasmic male sterility in *Brassica juncea* carrying *Moricandia arvensis* cytoplasm. *Plant Cell Physiol* 2008, **49**:284–289.
14. Kim DH, Kang JG, Kim BD: Isolation and characterization of the cytoplasmic male sterility-associated *orf456* gene of chili pepper (*Capsicum annuum* L.). *Plant Mol Biol* 2007, **63**:519–532.
15. Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG: 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.
16. Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ: Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 2007, **177**:1173–1192.
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**(2):996–1017.
18. Chen J, Guan R, Chang S, Du T, Zhang H, Xing H: Substoichiometrically different mitotypes coexist in mitochondrial genomes of *Brassica napus* L. *PLoS ONE* 2011, **6**:e17662.

19. Handa H: The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res* 2003, **31**:5907–5916.
20. 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**:209.
21. Liu H, Cui P, Zhan K, Lin Q, Zhuo G, Guo X, Ding F, Yang W, Liu D, Hu S, Yu J, Zhang A: Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. *BMC Genomics* 2011, **12**:163.
22. Park JY, Lee YP, Lee J, Choi BS, Kim S, Yang TJ: Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in radish (*Raphanus sativus* L.) containing DCGMS cytoplasm. *Theor Appl Genet* 2013, **128**:1763–1774.
23. Satoh M, Kubo T, Nishizawa S, Estiati A, Itchoda N, Mikami T: The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs. *Mol Genet Genomics* 2004, **272**:247–256.
24. Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T: A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genomics* 2012, **13**:352.
25. Peterson PA: Cytoplasmically inherited male sterility in *Capsicum*. *Amer Nat* 1958, **92**:111–119.
26. Kim DH, Kim BD: The organization of mitochondrial *atp6* gene region in male fertile and CMS lines of pepper (*Capsicum annuum* L.). *Curr Genet* 2006, **49**:59–67.
27. Gulyas G, Shin Y, Kim H, Lee JS, Hirata Y: Altered transcript reveals an *orf507* sterility-related gene in chili pepper (*Capsicum annuum* L.). *Plant Mol Bio Rep* 2010, **28**:605–612.
28. Li J, Pandeya D, Jo YD, Liu WY, Kang BC: Reduced activity of ATP synthase in mitochondria causes cytoplasmic male sterility in chili pepper. *Planta* 2013, **237**(4):1097–1109.
29. Millar AH, Sweetlove LJ, Giege P, Leaver CJ: Analysis of the *Arabidopsis* mitochondrial proteome. *Plant Physiol* 2001, **127**:1711–1727.
30. Kim DH: Isolation of Cytoplasmic Male Sterility (CMS)-Associated Gene and Development of CMS-Specific SCAR Markers in Chili Pepper, PhD thesis. Seoul, Republic of Korea: Seoul National University, Plant Science Department; 2004.
31. Yoo EY, Kim S, Kim JY, Kim BD: Construction and characterization of a bacterial artificial chromosome library from chili pepper. *Mol Cell* 2001, **12**(1):117–120.
32. Kubo N, Arimura S: Discovery of the *rpl10* gene in diverse plant mitochondrial genomes and its probable replacement by the nuclear gene for chloroplast *RPL10* in two lineages of angiosperms. *DNA Res* 2010, **17**(1):1–9.
33. Jo YD, Park J, Kim J, Song W, Hur CG, Lee YH, Kang BC: Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Rep* 2011, **30**:217–229.
34. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 1986, **5**:2043–2049.
35. Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M: The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Genet Genomics* 2005, **272**:603–615.
36. Lelandais C, Gutierrez S, Mathieu C, Vedel F, Remacle C, Marechal-Drouard L, Brennicke A, Binder S, Chetrit P: A promoter element active in run-off transcription controls the expression of two cistrons of *nad* and *rps* genes in *Nicotiana sylvestris* mitochondria. *Nucleic Acids Res* 1996, **24**:4798–4804.
37. Gutierrez S, Lelandais C, Paepe RD, Vedel F, Chetrit P: A mitochondrial sub-stoichiometric *orf87-nad3-nad1* exonA co-transcription unit present in solanaceae was amplified in the genus *Nicotiana*. *Curr Genet* 1997, **31**:55–62.
38. Singer BS, Gold L, Gauss P, Doherty DH: Determination of the amount of homology required for recombination in bacteriophage T4. *Cell* 1982, **31**(1):25–33.
39. Watt VM, Ingles CJ, Urdea MS, Rutter WJ: Homology requirements for recombination in *Escherichia coli*. *Proc Natl Acad Sci U S A* 1985, **82**(14):4768–4772.
40. Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA: Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. *Plant Cell* 2007, **19**:1251–1264.
41. Feng X, Kaur AP, Mackenzie SA, Dweikat IM: Substoichiometric shifting in the fertility reversion of cytoplasmic male sterile pearl millet. *Theor Appl Genet* 2009, **118**(7):1361–1370.
42. Newton KJ, Knudsen C, Gabay-Laughnan S, Laughnan JR: An abnormal growth mutant in maize has a defective mitochondrial cytochrome oxidase gene. *Plant Cell* 1990, **2**(2):107–113.
43. Hunt MD, Newton KJ: The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in *Zea mays* mitochondria. *EMBO J* 1991, **10**(5):1045–1052.
44. Marechal A, Brisson N: Recombination and the maintenance of plant organelle genome stability. *New Phytologist* 2010, **186**(2):299–317.
45. Marechal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N: Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A* 2009, **106**(34):14693–14698.
46. Stohr BA, Kreuzer KN: Coordination of DNA ends during double-strand-break repair in bacteriophage T4. *Genetics* 2002, **162**:1019–1030.
47. Jo YD, Jeong HJ, Kang BC: Development of a CMS specific marker based on chloroplast-derived mitochondrial sequence in pepper. *Plant Biotechnol Rep* 2009, **3**:309–315.
48. Jo YD, Jeong HJ, Kang BC: Origin of the *Capsicum* CMS cytoplasm revealed by cytoplasmic DNA derived marker analysis. *Sci Horti* 2011, **131**:74–81.
49. Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA: Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* 1998, **10**:1163–1180.
50. Kim S, Lim H, Park S, Cho KH, Sung SK, Oh DG, Kim KT: Identification of a novel mitochondrial genome type and development of molecular markers for cytoplasm classification in radish (*Raphanus sativus* L.). *Theor Appl Genet* 2007, **115**:1137–1145.
51. Giege P, Konthur Z, Walter G, Brennicke A: An ordered *Arabidopsis thaliana* mitochondrial cDNA library on high-density filters allows rapid systematic analysis of plant gene expression: a pilot study. *Plant J* 1998, **15**:721–726.

doi:10.1186/1471-2164-15-561

Cite this article as: Jo et al.: Extensive structural variations between mitochondrial genomes of CMS and normal peppers (*Capsicum annuum* L.) revealed by complete nucleotide sequencing. *BMC Genomics* 2014 **15**:561.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

