Mol Breeding (2021) 41: 61
https://doi.org/10.1007/s11032-021-01252-x

Mitochondrial genes are involved in the fertility transformation of the thermosensitive male-sterile line YS3038 in wheat

Yucui Han · Yujie Gao · Hao Zhou · Xiaoguang Zhai · Qin Ding · Lingjian Ma

Received: 26 October 2020 / Accepted: 5 September 2021 / Published online: 30 September 2021

Abstract Heterosis can improve the stress resistance, quality, and yield of crops, and the male sterility of wheat can be utilized to accelerate the breeding process of hybrid. To determine whether mitochondrial genes are involved in the fertility of K-type cytoplasmic male-sterile (CMS) line and the YS-type thermosensitive male-sterile (TMS) line in wheat, we sequenced and assembled the mitochondrial genomes of K519A, 519B, and YS3038 by next-generation sequencing (NGS). The non-synonymous mutations were analyzed, and the first-generation sequencing was conducted to verify the non-synonymous mutation sites. Furthermore, the expression patterns of genes with non-synonymous mutations were analyzed. Finally, the candidate genes were silenced by barley stripe mosaic virus-induced gene silencing (BSMV-VIGS) to test the functions of the candidate genes. The results revealed that the mitochondrial genomes of K519A, 519B, and YS3038 were 420,543, 433,560, and 452,567 bp in length, respectively. Besides, 33, 31, and 37 protein-coding genes were identified in K519A, 519B, and YS3038, respectively. There were 14 protein-coding genes and 83 open reading frame (ORF) sequences that differed between K519A and 519B and 10 protein-coding genes and 122 ORF sequences that differed between K519A and YS3038. At the binucleate stage, seven genes (*nad6*, *ORF256*, *ORF216*, *ORF138*, *atp6*, *nad3*, and *cox1*) were downregulated in K519A compared with 519B, and 10 genes (*nad6*, *atp6*, *cox3*, *atp8*, *nad3*, *cox1*, *rps3*, *ORF216*, *ORF138*, and *ORF224*) were downregulated in YS3038 compared with K519A. Besides, six genes (*nad6*, *ORF138*, *cox3*, *cox1*, *rps3*, and *ORF224*) were downregulated under fertile conditions relative to sterile conditions in YS3038. Gene silencing analysis showed that the silencing of *cox1* significantly reduced the seed setting rate of YS3038, indicating that the *cox1* gene may be involved in the fertility transformation of YS3038.

Keywords Wheat · Mitochondrial genome · Male sterility · qPCR · Gene silencing

Yucui Han, Yujie Gao, Hao Zhou, and Xiaoguang Zhai contributed equally to this work

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-021-01252-x.

Y. Han · Y. Gao · H. Zhou · X. Zhai · L. Ma (✉)
College of Agronomy, Northwest A&F University, Xianyang 712100, Yangling, Shaanxi, China
e-mail: malingjian@nwafu.edu.cn

Y. Han
College of Agronomy and Biotechnology, Hebei Normal University of Science and Technology, Qinhuangdao 066004, Hebei, China

Q. Ding (✉)
College of Horticulture, Northwest A&F University, Xianyang 712100, Yangling, Shaanxi, China
e-mail: dingqin@nwafu.edu.cn

Introduction

Wheat (*Triticum aestivum* L.) is a widely cultivated gramineous crop in the world, and its yield is of great significance to global food security (Kovacs et al. 2011). With the continuous development of the global economy, increasing attention has been paid to the nutritional value and edible quality of wheat food. It remains a great challenge to quickly and effectively breed new wheat varieties with high yield and high quality. Utilization of heterosis has gradually become an important way to improve the yield and quality in the breeding of new crop varieties (Luo et al. 2013a; Hochholdinger and Baldauf 2018). To date, heterosis has been used to generate hybrid varieties in maize (Li et al. 2017; Ding et al. 2012), rape (Shen et al. 2005), and sunflower (Aslam et al. 2010), which has brought about great economic and social benefits. Since wheat is a self-pollinated crop, the utilization of male-sterile lines cannot only reduce the production cost of hybrid seeds but also improve the purity of hybrid seeds (Ryan et al. 2013; Rajaram 2001). Cytoplasmic male sterility (CMS) is controlled by both nuclear genes and cytoplasmic genes, and the restorer genes in the nucleus can restore fertility (Chen and Liu 2014; Cui et al. 1996). In addition, the male sterility can be changed upon variations in the external environment. According to their response to environmental conditions, male-sterile lines can be divided into three types: photosensitive male-sterile lines (PMS), thermosensitive male-sterile (TMS) lines, and photothermosensitive male-sterile lines (PTMS) (Chen and Liu 2014).

The mitochondria and chloroplasts, which are semi-autonomous and possess a complete system for transmission and expression of genetic information, are necessary organelles for energy production (Fernie et al. 2004). Plant mitochondrial genomes encode many proteins participating in multiple functions, such as the respiratory chain and oxidative phosphorylation pathways, tRNA and rRNA synthesis, and DNA repair and transcription (Dingenen et al. 2016). Compared with chloroplast DNA, mitochondrial DNA encodes a variety of enzymes in respiratory metabolism, and more than 90% of the enzymes for ATP synthesis are produced within the mitochondria (Nieminen 2003). Mitochondrial genomes are relatively complex in structure with an irregular arrangement of molecules (Kitazaki and Kubo 2010). The variations of mitochondrial genomes are mainly ascribed to the number and length of repeat sequences (Palmer et al. 2000; Feagin et al. 1991; Ward et al. 1981). Active repetitive sequences may be involved in the process of plant mitochondrial DNA replication and repair (Marechal and Brisson 2010). In addition, the abundant repetitive sequences of sub-genomic structures tend to mediate frequent intramolecular recombination, which is the driving force for the evolution of plant mitochondrial genomes (Small et al. 1989; Kmiec et al. 2006). Although the number of encoding genes in mitochondrial genome differs greatly among different species, the number of major functional encoding genes is highly similar. Previous studies have shown that the mitochondrial genomes of higher plants generally encode 22 subunits of oxidative phosphorylation, 6–11 subunits of ribosomal protein, 17–22 tRNA genes, three rRNA genes, and many open reading frames (ORFs) with unknown functions (Itoh et al. 2002; Hirokazu 2003).

With the progress of research on plant CMS, more and more studies have revealed the close association of mitochondrial genes with CMS in plants (Belliard et al. 1979). For example, it was found that cytoplasmic abnormality can cause disharmony between the mitochondria and nuclear genes, which will further affect the development of gametophytes (Hanson and Bentolila 2004; Pruitt and Hanson 1991). The mutation of some genes may also hinder the normal development of mitochondria and then affect the growth and development of pollen, resulting in a sterile phenotype (Fujii and Toriyama 2008; Laser and Lersten 1972). In the mitochondrial genome of plants, about 60 conserved genes encode approximately 30 proteins, forming some complex subunits of respiratory chain metabolism: complex I genes (nad1, nad2, nad3, nad4, nad4l, and nad5), complex II genes (sdh4 and sdh3), complex III genes (cob), complex IV genes (cox1, cox2, and cox3), complex V genes (atp1, atp4, atp6, atp8, and atp9), cytochrome C synthesis bases (ccmC, ccmB, ccmFC, and ccmFN), and some ORFs with unknown functions (Kubo and Newton 2008). The mitochondrial genes *atp6* and *cox2* are involved in important biological processes and are related to CMS in wheat, rice, and other species (Yi et al. 2002; Kong et al. 2006; Howad and Kempken

1997; Landgren et al. 1996). It was found that CMS line in rice contains the chimeric *atp6* gene (*urf-rmc*) and normal *atp6* gene, but there is no *urf-rmc* in fertile cytoplasm. The introduction of the restorer gene changed the transcription of the *urf-rmc* gene but not that of the *atp6* gene, indicating that the chimeric gene is involved in the CMS (Kadowaki et al. 1990). In addition, it was found that the occurrence of CMS is often related to the ORF of mitochondrial chimera, which is formed into a unique protein to interfere with the function of mitochondria and the development of pollen (Schnable and Wise 1998). Mohammed Sabar et al. (Sabar et al. (n.d.)) found that the expression of *ORF522* would lead to abnormal energy production and metabolism in the sterile line, resulting in pollen sterility. Some studies have proved that changes in the encoded proteins may interfere with the normal respiratory chain reaction, reduce the supply and production of energy, and affect the fertility of pollen (Warmke and Lee 1978; Dieterich et al. 2003).

The CMS line K519A and TMS line YS3038 bred in our laboratory are completely sterile, and YS3038 has a high seed setting rate under fertile conditions. Although both K519A and YS3038 have highly promising application prospects, the mechanism of their fertility remains unclear. In this study, the mitochondrial genomes of K519A, its homomaintainer line 519B, and YS3038 were studied to analyze the effect of mitochondrial genes on the fertility of K519A and YS3038.

Materials and methods

Material planting and cultivation

The wheat CMS line K519A, its homomaintainer line 519B, and TMS line YS3038 from our laboratory (College of Agronomy, Northwest A&F University, Yangling, China) were used in this study. Among them, K519A is a K-type CMS line with the cytoplasm of *Ae. Kotschyi*, and the nucleus of *Triticum aestivum* L. 519B, with the cytoplasm and nucleus of *T. aestivum* L., is the homomaintainer line of K519A. YS3038, a YS-type TMS line with the cytoplasm of *Ae. Kotschyi* and the nucleus of *T. aestivum* L., is sterile at below 18 °C (with a seed setting rate of 0%) but fertile at above 20 °C (with a seed setting rate of over 99%) at the microspore development stage. YS3038 line was unexpectedly discovered in the breeding process of K519A. Hence, it is believed that YS3038 and K519A have similar cytoplasm and nucleus, which remains to be verified. The materials were planted in the experimental farm of Northwest A&F University (108° 4′E, 34° 16′N) in October 2018, with a row spacing of 25 cm and a seed spacing of 7–8 cm. The field management was carried out according to local practice. Young leaves were taken in March 2019 and stored in a –80 °C refrigerator for DNA-seq analysis and the following genome assembly. The seed setting rate of K519A and YS3038 was 0%, and that of 519B was 100%.

The plants used for qPCR and gene silencing were cultivated in an artificial constant temperature incubator. K519A, 519B, and YS3038 were all cultivated under 17/13 °C until just before meiosis. Then, K519A, 519B, and fertile YS3038 plants were cultivated under 24/20 °C after the initiation of meiosis.

The genome of Chinese Spring with the cytoplasm and nucleus of *Triticum aestivum* L. was used as the reference genome. Ks3 is a K-type CMS line with the cytoplasm of *Ae. Kotschyi* and the nucleus of *Triticum aestivum* L., and Yumai *T. aestivum* (Km3) is the maintainer line of Ks3. *Aegilops speltoides* is of the cytoplasm and nucleus of *Ae. Kotschyi*. The sequences of Chinese Spring (KJ614396.1), Ks3 (GU985444), Km3 (EU534409), and *Ae. speltoides* (NC022666) were obtained from the genome database at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/).

Total RNA and DNA extraction and mitochondrial genome sequencing

Total RNA was extracted from anthers with the TaKaRa MiniBEST Plant RNA Extraction Kit (Takara). Then, the RNA concentration was determined by a NanoDrop2000 Spectrophotometer (Thermo Fisher Scientific). The integrity of RNA was detected by 1.5% agarose gel electrophoresis, and the total RNA was used for subsequent experiments.

The next-generation sequencing (NGS) technology with the Illumina HiSeq-XTM Ten sequencing platform and paired-end 150 bp sequencing strategy was used to obtain the whole genomic DNA sequences. By using the whole genomic comparison software Bowtie2 (Langmead 2012), the mitochondrial genome reads were captured from the total

reads. The advantage of this approach is that it could screen and filter mismatches or multiple matches of reads and obtain pure mitochondrial reads through subsequent quality control. A total of 4G clean data were obtained after removal of the low-quality reads and connector sequences, which were then used for further assembly of the mitochondrial genome.

Genome assembly and annotation

The mitochondrial genome was assembled with the method of Hahn et al. (Hahn et al. 2013). The accuracy of MITObim developed by Hahn et al. could reach more than 99.5%. Firstly, the clean data after quality control were preliminarily spliced with the SPAdes software (Bankevich et al. 2012), with the setting of default parameters except that the cut off parameter was not selected. The scaffolds were then spliced. By using the mitochondrial genome of Chinese Spring (KJ614396.1) as a reference, the DNA and amino acid sequences of the protein-coding genes were aligned by BLASTn and Exonerate, respectively. The threshold *e* value was set as $1e^{-10}$, and the protein similarity threshold was set as 70%. The scaffolds matched to genes were selected, and the scaffolds were sorted by coverage rate, and those not related to the target genome were deleted. PRICE and MITObim were used to merge and splice the collected scaffolds and reduce the number of scaffolds as much as possible with 50 iterations. The Bowtie 2 software was employed to compare the original clean reads with the obtained scaffolds and pick out the matched reads, and then the SPAdes software was used for re-splicing. Then, it was checked whether there was an obvious circle diagram. If there was an obvious circle diagram, the circular genome was extracted; otherwise, the above steps were repeated.

Organelle genome annotation consisted of three parts: protein-coding gene annotation, RNA annotation, and structure annotation. The protein-coding genes were annotated mainly using the Ugene ORF finder tool. First, the standard codon table was selected to predict ORFs, and then the predicted ORFs were compared with the nr database using the BLASTp program, followed by annotation of the functions. For genes containing introns, the Exonerate software was used to determine intron boundaries and length by comparing them with amino acid sequences of genes from proximal species.

The mitochondrial sequence was submitted to the tRNAscan-SE website (http://lowelab.ucsc.edu/tRNAscan-SE/) for tRNA annotation. The sequence was submitted to the RNAmmer 1.2 server website (http://www.cbs.dtu.dk/services/RNAmmer/) for rRNA prediction, and homologous sequence alignment was performed to determine the boundary range. After the sequence annotation, the sequence was edited by Sequin to generate a file that could be submitted to the GenBank database. The edited GenBank file was submitted to OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) to draw the annotation map.

Gene comparison analysis and verification of non-synonymous mutation sites

The sequences of protein-coding genes were compared among 519A, 519B, and YS3038 by the MEGA-X software. In order to verify the accuracy of the results of NGS, assembly, and splicing, the non-synonymous mutation sites in protein-coding genes were selected for the first-generation sequencing verification. According to the NGS results, Primer Premier 6.0 software was used to design the corresponding primers. The 20 μL PCR reaction mixture contained 10 μL mix (Takara Primer STAR Max DNA Polymerase), 0.5 μL forward primer, 0.5 μL reverse primer, 0.4 μL DNA (100 ng/μL), and 8.6 μL ddH2O. Then, PCR amplification was carried out, and the products were sent to Sangon Biotech Company for first-generation sequencing. Finally, the results of first-generation sequencing were aligned to those of NGS.

Analysis of mitochondrial genome collinearity

In addition to the identification of site mutations in each gene, the comparison of mitochondrial genome structures could reflect the overall differences in the mitochondria between different materials. The consistency in sequence among different genomes could well reflect the common origin of the genomes. The Mauve software (Darling et al. 2004) was used for collinearity analysis with default parameters, which could perform traditional multiple alignments of conserved regions for the identification of chromosome translocation, chromosome inversion, and indels.

Quantitative real-time PCR (qPCR) analysis

The gene-specific primers were designed using the Premier 6.0 software. The cDNA was synthesized

with random primers using a RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). The TB GreenTM Premix Ex TaqTM II (Tli RNaseH Plus) Kit (Takara Biological Engineering) was used for the qPCR analysis. The qPCR was performed on the Applied Biosystems 7300 Real-Time PCR System (Life Technologies U.S). The wheat *Actin* gene (GenBank, AB181991.1) was used as the reference gene. The expression of target genes was calculated by the $2^{-\Delta\Delta CT}$ method. The qPCR was performed with three biological replicates and three technical replicates.

Functional verification of candidate genes via *the BSMV-VIGS method*

Construction of γ-cox1 vector

Primer Premier 6 software was used to design the primers for about 200 bp product, and the *PacI* (TTA ATTAA) and *NotI* (GCGGCCGC) were selected for the forward and reverse primers, respectively. After the digestion of the plasmids of γ-PDS and T-gene by *PacI* and *NotI*, the vector and gene fragments were collected and purified. The vector and the gene fragments were connected, and the correct clone was used for subsequent experiments.

Linearization and in vitro transcription of γ-cox1 vector

The plasmids of α, β, γ, γ-PDS, and γ-cox1 were extracted, respectively. The α and γ plasmids were digested with *MluI*; the β plasmid was digested with *Spe*I; the γ-PDS and γ-gene plasmids were digested with *BssHII*. Ribo m^7G Cap Analog (Promega), RiboMAX Large Scale RNA Production System-T7 (Promega), and Ribolock RNase Inhibitor (Thermo) were used for in vitro transcription.

Creation of transfection mixture and virus infection of seedlings

The α, β, and γ mixture was used as the negative control combination, while the α, β, and γ-PDS mixture was used as the positive control combination, and the α, β, and γ-*cox1* mixture was used as the treatment combination. A total of 21 µl of mixture was prepared for each combination, including 7 µl in vitro transcription product for each vector, 40 µl sterilized 1% DEPC water, and 200 µl GK-Pbuffer, which could be used to infect five individuals.

When the flag leaves were fully unfolded, the seedlings were infected. Sterile rubber gloves were used for the infection. The mixture was taken to the fingers to gently rub the penultimate leaf and the flag leaf. The infected wheat seedlings were placed in a 24–26 °C incubator for dark cultivation, and the light was restored after 24 h.

Detection of silencing efficiency and phenotypes

After virus infection, when the microspores develop to the binucleate stage, 4–5 spikelets were collected from the top of the ear for RNA extraction and qPCR analysis. After the wheat plants grew to the mature stage, the seed setting rate of each spikelet was calculated as follows: Seed setting rate per ear = number of grains per ear / (effective spikelet number × 2) × 100%.

Results

Genome composition and characteristics

The NGS technology and biological information software were used to assemble a complete mitochondrial genome. A total of 29,000,000 paired-end reads (150 bp) were obtained by sequencing and then used for mitochondrial genome assembly. The Q30 quality scores of three samples were all greater than 92%, and the length of the mitochondrial genome was 420–450 kb, indicating a sequencing coverage of 40×. The mitochondrial genomes of K-type CMS line K519A, its homomaintainer line 519B, and YS3038 were found to have a typical circular structure. The total length of the mitochondrial genome of K519A was 420,543 bp (Fig. S1), with a G/C content of 44.14%; that of 519B was 433,560 bp, with a G/C content of 44.14% as well (Fig. S2); and that of YS3038 was 452,567 bp, with a G/C content of 44.35% (Fig. S3). The GC content of the mitochondrial genomes was generally between 43 and 45% (Liu et al. 2011), indicating the reliability of the results in the present study.

The encoded genes were very similar between the K519A and 519B genomes. The encoded genes were

generally divided into nine categories: tRNA genes, rRNA genes, respiratory chain complex I genes, cytochrome c biogenesis genes, complex III genes, complex IV genes, ATP synthase genes, ribosomal proteins genes, and others. The mitochondrial genome of K519A contained 24 tRNA genes, three rRNA genes, 32 known protein-coding genes, and three other genes (Table 1); besides, there were multiple copies for tRNA, including three copies for *tRNA-Trp*, *tRNA-Ser*, and *tRNA-Met*, and two copies for *tRNA-Cys*, *tRNA-Pro*, *tRNA-Lys*, and *tRNA-Asp*. The mitochondrial genome of 519B included 25 tRNA genes, three rRNA genes, 31 known protein-coding genes, and two other genes (Table 2). The mitochondrial genome of YS3038 harbored 26 tRNA genes, nine rRNA genes, and 37 known protein-coding genes

Table 1 Gene structure of mitochondrial genome in K519A

Category for genes	Gene names				
tRNAs	tRNA-Glu	tRNA-Trp(3)	tRNA-Cys(2)	tRNA-Gly*	tRNA-Pro(2)
	tRNA-Gln	tRNA-Lys(2)	tRNA-His	tRNA-Ser(3)	tRNA-Met(3)
	tRNA-Phe	tRNA-Asn	tRNA-Asp(2)	tRNA-Ala*	
rRNAs	rrn5	rrn18	rrn26		
Complex I (NADH dehydrogenase)	nad2	nad3	nad4***	nad4L	nad5*
	nad6	nad7****	nad9		
Cytochrome c biogenesis	ccmB	ccmFN	ccmFC*	ccmC	
Complex III (ubiquinol cytochrome c reductase)	cob				
Complex IV(cytochrome c oxidase)	cox1	cox2*	cox3		
ATP synthase	atp1	atp4	atp6	atp8	atp9
	atpA-like[1]				
Ribosomal proteins	rps1	rps2	rps3**	rps4	rps7
	rps12	rps13	rps19	rpl5	rpl16
Others	mttB	matR	ndhB-like[2]		

*, the number of introns; (2) two copies; (3) three copies

Table 2 Gene structure of mitochondrial genome in 519B

Category for genes	Gene names				
tRNAs	tRNA-Glu	tRNA-Met(3)	tRNA-Cys(2)	tRNA-Trp(3)	tRNA-Pro(2)
	tRNA-Gln	tRNA-Lys(2)	tRNA-His	tRNA-Ser(3)	tRNA-Asn
	tRNA-Phe	tRNA-Gly	tRNA-Asp(2)	tRNA-Ala	tRNA-Leu
rRNAs	rrn5	rrn18	rrn26		
Complex I (NADH dehydrogenase)	nad2	nad3	nad4***	nad4L	nad5*
	nad6	nad7****	nad9		
Cytochrome c biogenesis	ccmB	ccmFN	ccmFC*	ccmC	
Complex III (ubiquinol cytochrome c reductase)	cob				
Complex IV(cytochrome c oxidase)	cox1	cox2*	cox3		
ATP synthase	atp1	atp4	atp6	atp8	atp9
Ribosomal proteins	rps1	rps2	rps3**	rps4	rps7
	rps12	rps13	rps19	rpl5	rpl16
Others	mttB	matR			

*, the number of introns; (2) two copies; (3) three copies

(Table 3). The compositions of wheat mitochondrial genomes in this study were basically consistent with previous research results (Yasunari et al. 2005; Wang et al. 2015). The *tRNA-Leu* gene was only present in the 519B mitochondrial genome. *tRNA-Tyr* and *tRNA-Ile* were specifically present in YS3038, but *tRNA-Gly* was only found in K519A and 519B. Among the protein-encoding genes, the *nad7* gene harbored the most introns (four), followed by the *nad4* gene (three introns). The *rrnS-3* gene and the *nad1* gene were unique to YS3038 and not found in K519A and 519B.

Alignment and analysis of protein-encoding genes

The protein-encoding genes were compared between K519A and 519B. As a result, the *atp6* and *rps13* genes were only present in 519B, while the *atpA-like* gene was unique to K519A. A total of 36 mutation sites were found in 12 genes besides *rrn18*. In addition, 25 non-synonymous mutation sites occurred on 11 genes besides *rrn18*, accounting for 69.44% of the total mutation sites. There were 16 gaps and 45 mutation sites for the *rrn18* gene in K519A relative to 519B. The *ccmC*, *atp4*, and *rps3* genes had the most non-synonymous mutation sites, with six, five, and four mutation sites, respectively (Table 4). Similarly, Liu et al. (Liu et al. 2011) compared the mitochondrial genomes of sterile line Ks3 and maintainer line Km3 and found 20 non-synonymous mutation sites occurring on nine genes.

A comparison of the mitochondrial protein-encoding genes between YS3038 and K519A revealed that the *atp6* gene was only present in YS3038, and the *atpA-like* gene was specific to K519A. There were a lot of different bases on *nad2*, *nad9*, *cob*, *rpl16*, and *matR* between YS3038 and K519A. Except for those genes, a total of 28 mutation sites were found on seven genes. Among them, 15 non-synonymous mutation sites occurred on five genes, accounting for 53.57% of the total mutations. The *atp4* gene had the most mutation sites, with six non-synonymous mutation sites and three synonymous mutation sites (Table 5). These non-synonymous mutation sites between K519A and YS3038 were fewer than those between K519A and 519B, which was also in line with the characteristics of the kinship among the three lines.

Alignment and analysis of open reading frames

A lot of ORFs were found in K519A, 519B, and YS3038 besides the annotated genes, which were designated according to the length of the amino acid. 519B had more ORFs (147) than 519A (138) and YS3038 (126). There were 83 ORF sequences that differed between K519A and 519B (Table S1),

Table 3 Gene structure of mitochondrial genome in YS3038

Category for genes	Gene names				
tRNAs	tRNA-Glu	tRNA-Met(4)	tRNA-Ala	tRNA-Cys	tRNA-Pro(2)
	tRNA-Gln(3)	tRNA-Lys(3)	tRNA-Tyr	tRNA-Asp(2)	tRNA-Ser(3)
	tRNA-Phe	tRNA-Asn	tRNA-lle	tRNA-His	tRNA-Trp
rRNAs	rrnS-3	rrn18(3)	rrn26-p	rrn26(2)	rrn5(3)
Complex I (NADH dehydrogenase)	nad1****	nad2****	nad3***	nad4L	nad4***
	nad5****	nad6	nad7****	nad9	
Cytochrome c biogenesis	ccmB	ccmFN	ccmFC*	ccmC	
Complex III (ubiquinol cytochrome c reductase)	cob				
Complex IV(cytochrome c oxidase)	cox1	cox2*	cox3		
ATP synthase	atp1	atp4	atp6(2)	atp8(2)	atp9
Ribosomal proteins	rps1	rps2	rps3	rps4	rps7
	rps12	rps13	rps19	rpl2-p	rpl5
	rpl16				
Others	mttB	matR			

*, the number of introns; (2) two copies; (3) three copies; (4) four copies

Table 4 Alignment analysis of protein-coding genes of K519A and 519B

Genes	Position and nucleic acid (K519A, 519B)	Position and amino acid (K519A, 519B)	Mutation type	SNP type
rrn18	**	**	-	-
nad3	185C-T	62Phe-Leu	N	2
nad4	536G-T		S	2
	803A-C		S	2
	821A-G		S	2
	1289 T-G		S	2
nad6	709G-A	237Asp-Asn	N	2
	730A-C	244Lys-Gln	N	2
	734 T-A	245Met-Lys	N	1
	742A-T		S	1
ccmFN	306A-C		S	2
	1103 T-A	368Gln-Leu	N	1
ccmC	178G-A	60Ala-Thr	N	2
	188 T-C	63Met-Thr	N	2
	190A-G	64Ser-Gly	N	2
	208G-A	70Ala-Met	N	2
	209C-T		S	2
	262C-T	88Arg-Cys	N	2
	278G-A	93Gly-Asp	N	2
cox3	157C-A	53Leu-Ile	N	2
atp4	32C-G	11Thr-Arg	N	1
	46 T-G	16Ser-Ala	N	2
	56G-C	19Arg-Pro	N	1
	102G-A		S	2
	128G-T	43Arg-Ile	N	2
	137A-T	46Gln-Leu	N	1
	183 T-C		S	2
atp6	A no		-	-
atp8	149A-G	50His-Arg	N	2
	167G-A		S	2
atpA-like	B no		-	0
rps1	144C-A		S	2
	508 T-C	170Ile-Thr	N	2
rps2	521A-C	174Tyr-Ser	N	2
rps3	256A-C	86Lys-Gln	N	2
	505 T-G	169Phe-Asp	N	2
	506 T-A	169Phe-Asp	N	1
	1333C-A	445Gln-Lys	N	2
rps4	146 T-G	49Leu-Arg	N	2
rps13	A no		-	-

**, abundant difference. *S*, synonymous; *N*, non-synonymous

among which 37 ORFs had 98 base mutations. Non-synonymous mutations accounted for 68.37% of the total mutations. There were 124 ORF sequences that differed between K519A and YS3038 (Table S2), with 56 base mutations occurring in 28 ORFs, and the most differential ORF in sequence was ORF247, which had 15 base mutations. Non-synonymous mutations accounted for 67.86% of the total base

Table 5 Alignment analysis of protein-coding genes of K519A and YS3038

Genes	Position and nucleic acid (K519A, YS3038)	Position and amino acid (K519A, YS3038)	Mutation type	SNP type
nad2	**	**	-	-
nad4	1475C-A		S	2
nad6	483A-G		S	2
	537 T-A		S	1
	630C-G	210Ser-Arg	N	1
	631C-A	211His-Asn	N	2
nad9	**	**	-	-
cob	**	**	-	-
cox1	1456 T-A	486Cys-Ser	N	1
	1551A-T		S	1
	1575G-A		S	2
atp4	32C-G	11Thr-Arg	N	1
	46 T-G	16Ser-Ala	N	2
	56G-C	19Arg-Pro	N	1
	102G-A		S	2
	128G-T	43Arg-Ile	N	2
	137A-T	46His-Leu	N	1
	183 T-C		S	2
	250G-C	84Glu-Pro	N	1
	251A-C		S	2
atp6	A no		-	-
atp8	39C-T		S	2
	149A-G	50His-Arg	N	2
	167G-A	56Ile-Ser	N	2
atpA-like	YS no		-	-
rps2	837A-T		S	1
	838C-A	280Gln-Lys	N	2
	844C-A	282Arg-Thr	N	2
	845G-C	282Arg-Thr	N	1
	961C-A	321His-Asn	N	2
	1002G-T		S	2
	1062C-A		S	2
rps3	1333C-A		S	2
rpl16	**	**	-	-
matR	**	**	-	-

**, abundant difference; *S*, synonymous; *N*, non-synonymous; *1*, transversion; *2*, transition; *A no*, no gene in K519A; *YS no*, no gene in YS3038

mutations, and base transversion accounted for 12.5%.

Verification of non-synonymous mutation sites by first-generation sequencing

In order to verify the accuracy of the base mutations of the protein-encoding genes in K519A, 519B, and YS3038, the non-synonymous mutation sites from 10 selected genes and ORFs were sequenced by first-generation sequencing (FGS). The corresponding primers are listed in Table 6. The FGS results (Supplementary data) revealed that the mutation sites of the *cox3*, *nad3*, *ORF138*, and *ORF256* were basically consistent with the results of NGS. In addition, there were also some differences

Table 6 Primers for the first-generation sequencing and qPCR

Genes	Forward primers	Reverse primers
cox3	AGAGCCTCCTTCTTTACCACTT	ATGATGAGCCCAAGTTACGG
nad3	CGAATCCACATCTGAGCGTCTT	GGTTACTCCCGATCCGAAGCA
rps3	CAGATCCAAGTCGGTTCAGTGA	ACTTCCGCAGGAGCATAATCG
atp8	GGATGCACAAGGGTACAACTG	CTTGCTTGCTTCTTCGAATCTC
atp6	ATCGGCTCGACGCAGGATT	TGGTCTTCTCGGCTGGAATCTA
cox1	TCACCATAGGAGGGCTCACTG	GGAGGGCTTTGTACCAACCATT
ORF138	TCAGGCAAGAAAGCAAAGAAGA	TTCCAACAAAGAGGAGAAGACA
ORF256	TATGGTTCGATGGCTCTTCTCC	CTCCCTCATTCTCCCGCAGAT
ORF216	CCTCCACACCAATCACGAGTT	TCCAAGAGCCGAACGAGAATG
ORF224	TCTCTGGAACTTGTTCGATCCT	CCTTCTCCTTCACCGTCATCAT
cox1-VIGS	CCTTAATTAAGCCATTCTGGAGGAGCAGTT	ATAAGAATGCGGCCGCCGCCA GTACCGGAAGTGATA
cox1-qPCR	CTTCGGTCATCCAGAGGTGT	TCAACGTCTAAGCCCACAGT

between the results of FGS and NGS. For example, the *rps3* gene had a difference of one base at the mutation site in FGS compared with NGS, which did not cause any amino acid change. It was found that there was no non-synonymous mutation in the NGS on *cox1* and *atp8* between K519A and YS3038. In the NGS results, *ORF224* was not found in K519A and 519B, but it was detected in K519A and 519B in the FGS results. Besides, compared with 519B, there was a non-synonymous mutation site in K519A and YS3038. The *atp6* and *ORF216* genes were not found in K519A by NGS, but FGS revealed that K519A contained these two genes. By integrating the FGS and NGS results, a total of 33 protein-coding genes and 140 ORFs in K519A, 31 protein-coding genes and 148 ORFs in 519B, and 37 protein-coding genes and 126 ORFs in YS3038 were found. There were 14 non-synonymous mutation protein-coding genes between K519A and 519B and 10 between YS3038 and K519A.

Comparison of non-synonymous mutation protein-coding genes with those from other species

A total of 12 genes with non-synonymous mutations between K519A and 519B and between YS3038 and K519A (*nad3*, *nad6*, *ccmFN*, *ccmC*, *cox3*, *atp4*, *atp6*, *atp8*, *rps1*, *rps2*, *rps3*, and *rps4*) were selected. Then, the genome of Chinese Spring wheat was used as a reference, and these 12 genes from the CMS line Ks3, its maintainer line Km3, *Ae. speltoides*, K519A, 519B, and YS3038 were compared with the reference genome. Compared with those in the Chinese Spring, eight genes were coincident, and four genes (*ccmFN*, *cox3*, *rps2*, *rps4*) exhibited non-synonymous mutations in *Ae. speltoides*. In Km3, a total of 10 genes were coincident, and only the *atp8* and *ccmFN* genes showed non-synonymous mutations relative to Chinese Spring. In Ks3, only four genes were coincident, and seven genes (*ccmFN*, *cox3*, *nad3*, *rps1*, *rps2*, *rps3*, *rps4*) had non-synonymous mutations. And the *nad6* shows insertion and mismatch of fragments. In K519A, only three genes were coincident, and seven genes (*ccmFN*, *cox3*, *nad3*, *rps1*, *rps2*, *rps3*, *rps4*) had non-synonymous mutations, with the mismatching, insertion, and deletion of some fragments on *atp4* and *atp8*. In 519B, six genes were coincident; four genes (*ccmC*, *cox3*, *rps2*, *rps4*) had non-synonymous mutations; and a 200 bp fragment was inserted in *nad6*, and an about 200 bp fragment was deleted in *atp8*. In YS3038, only two genes were coincident, and nine genes (*atp4*, *atp6*, *atp8*, *cox3*, *nad3*, *rps1*, *rps2*, *rps4*, *nad6*) had non-synonymous mutations, with a fragment deletion occurring on *rps3* (Table S3).

Collinearity analysis of mitochondrial genomes

The collinearity in mitochondrial genome between species can be used to analyze their evolutionary relationship. The collinear alignment of mitochondrial

genomes of K519A and 519B was performed, with that of 519B being used as the reference genome. The results showed that there were many translocation events and 12 inversion events, and some fragments had almost no similarity to each other (Fig. 1a). Similarly, the genome of K519A was used as the reference to compare the collinearity of the YS3038. As a result, there were 19 inversion events and many translocation events. The results indicated that the probability of genomic rearrangement in YS3038 relative to K519A was greater than that in K519A relative to 519B (Fig. 1b). By using the Chinese Spring as a reference, the collinearity in mitochondrial genomes of the six species was analyzed (Fig. 1c). As a result, Km3 and YS3038 had the highest collinearity with the Chinese Spring, suggesting that they had the lowest probability of genomic rearrangement. The mitochondrial genome of *Ae. speltoides* had the lowest collinearity with that of the Chinese Spring, which may be related to the characteristics of the species itself. It can also be concluded that Km3 had the most similar genomic composition to YS3038. Besides, Chinese Spring and Km3 had the highest collinearity with K519A.

qPCR analysis of non-synonymous mutation genes

In order to further explore the effect of non-synonymous mutation genes in the mitochondria, we performed qPCR analysis on 11 genes to assess their expression patterns. The corresponding primers are listed in Table 6. Compared with those in 519B, the expression levels of *nad6*, *ORF256*, *ORF216*, *ORF138*, *atp6*, *nad3*, and *cox1* genes were significantly downregulated in K519A at the binucleate stage, while there was no significant difference in the expression of *cox3*, *atp8*, *rps3*, and *ORF224* genes. Compared with those in K519A, the *nad6*, *atp6*, *cox3*, *atp8*, *nad3*, *cox1*, *rps3*, *ORF216*, *ORF138*, and *ORF224* genes were downregulated in YS3038 under sterile conditions, but the expression of *ORF256* showed no significant difference. In YS3038, the expression of *nad6*, *ORF138*, *cox3*, *cox1*, *rps3*, and *ORF224* was downregulated under fertile conditions compared with under sterile conditions. However, the expression of *ORF256*, *ORF216*, *atp6*, *atp8*, and *nad3* was not significantly different under different conditions (Fig. 2).

Functional analysis of candidate genes by BSMV-VIGS method

In order to further investigate the relationship between mitochondrial genes and fertility, the *cox1* and *nad6* genes were selected to perform a gene silencing analysis for functional verification by combining the results from previous studies and the qPCR results. First, primers containing 200 bp sequence and restriction enzyme restriction sites were designed (Table 6). Then, the recombinant vectors of γ and *cox1* were constructed. The γ empty vector was used as the negative control, and the *PDS* gene was used as the positive control. The YS3038 seedlings at the stage of microspore meiosis were infected. Ten individuals were infected with each combination and cultivated in an incubator after infection. After approximately 10 days of infection with γ-*PDS*, the positive control group showed obvious whitening of leaves (Fig. 3a). The seed setting rate of *cox1*-silenced plants (3%) was significantly lower than that of the negative control (92%) (Fig. 3b) (Table 7), indicating that *cox1* is likely involved in the fertility transformation of YS3038.

In order to test the efficiency of gene silencing, the γ vector was used as the negative control, and the spikelets of infected individuals at the binucleate stage were sampled. Then, the corresponding primers for qPCR were designed. The expression level of the *cox1* gene in the gene-silenced individuals with the cox1-VIGS primer (Table 6) was 0.22 on average, exhibiting a significant downregulation compared with the negative control. The relative expression level of the fragment used for gene silencing was 15,000 in the silenced plants with cox1-qPCR primer (Table 6), which was extremely significantly upregulated compared with the negative control, indicating that the vector carrying the segment of the gene replicated rapidly in the plants (Fig. 4). All the results indicated that the *cox1* gene was successfully silenced in the treated plants.

Discussion

Assembly of the mitochondrial genome

The traditional mitochondrial genome assembly process is as follows. Firstly, mitochondrial organelles

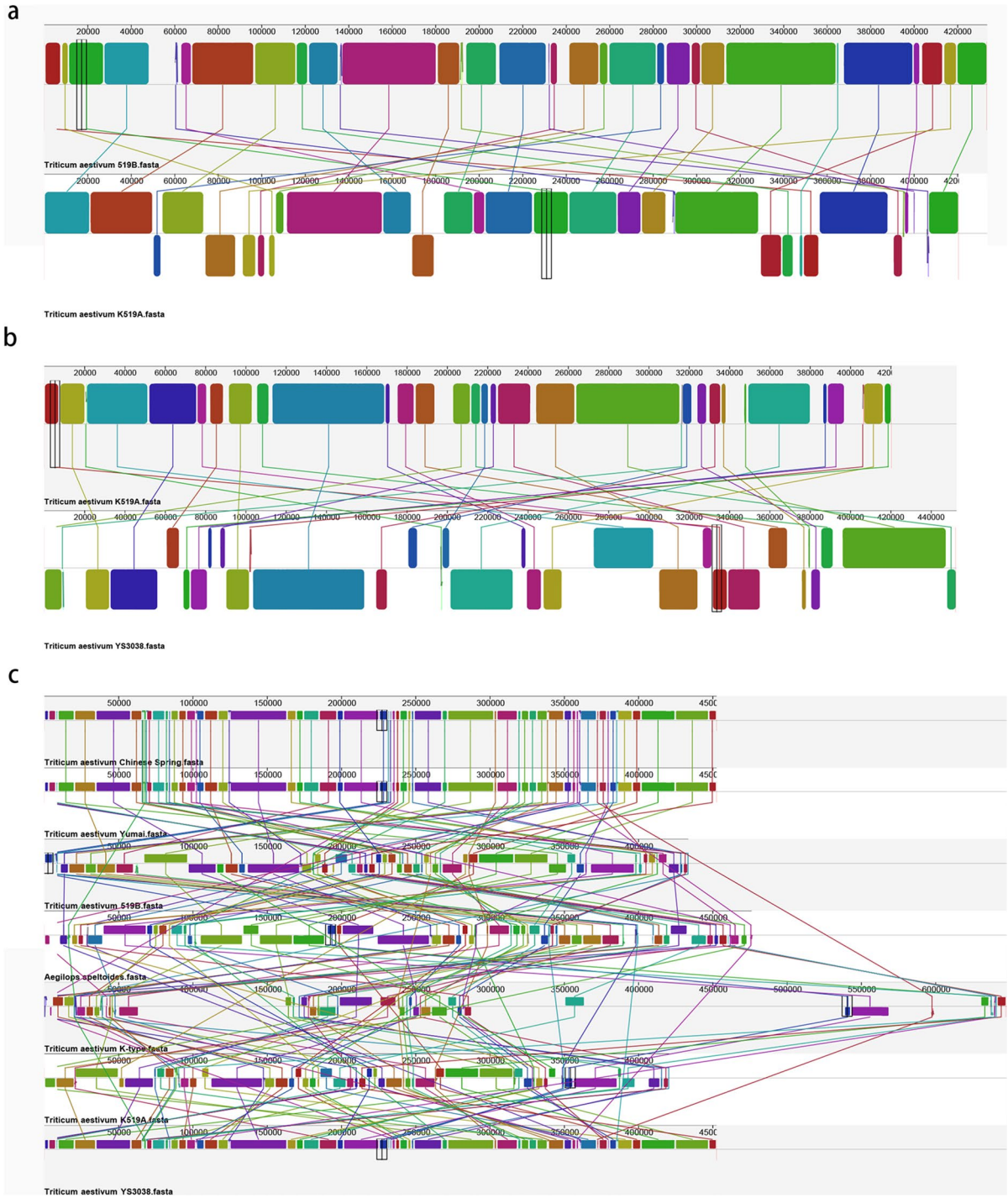

Fig. 1 Collinearity analysis of mitochondrial genomes. Each chromosome is oriented horizontally, and homologous blocks are shown as identically colored regions linked across genomes. Blocks below the midline are homologous blocks that are inverted. Gaps between blocks indicate lineage-specific sequences. **a** K519A and 519B. **b** YS3038 and K519A. **c** Among seven species (Chinese Spring, Km3, 519B, *Ae. Kotschyi*, Ks3, K519A, YS3038)

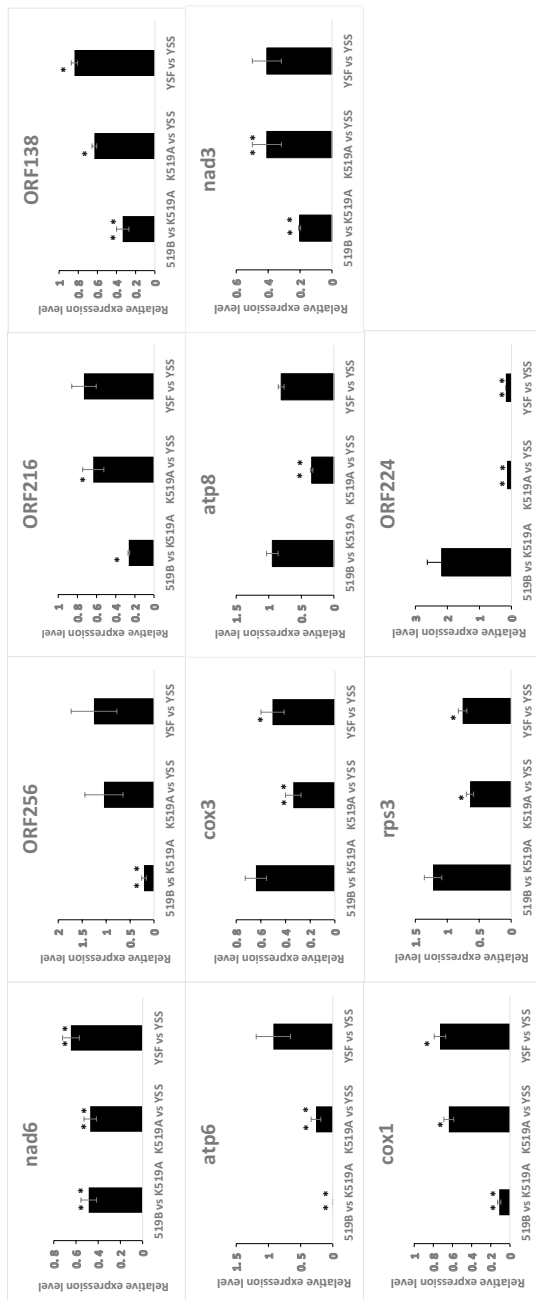

Fig. 2 Relative expression levels of non-synonymous mutant genes. YSF, YS3038 under fertile conditions. YSS, YS3038 under sterile conditions. The abscissa represents the comparable group. The ordinate represents the expression level of related genes

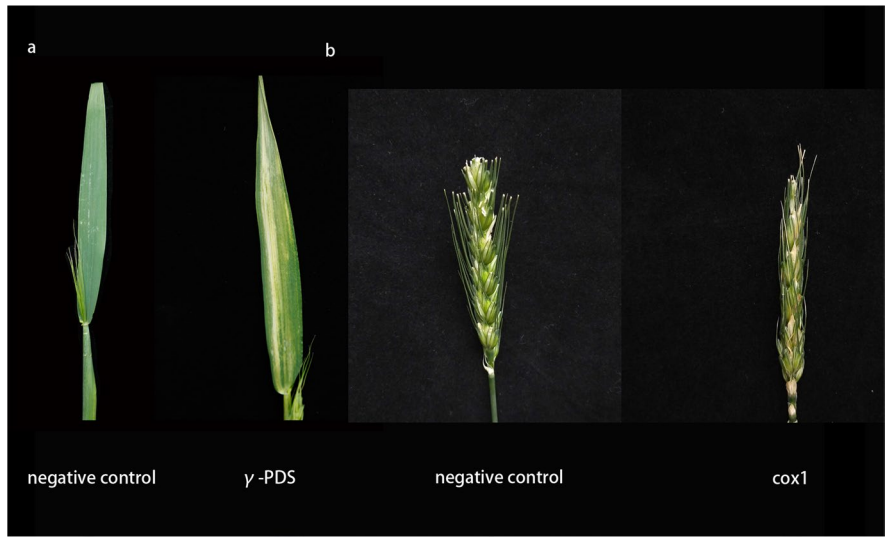

Fig. 3 Phenotypic identification of gene-silenced plants. **a** Leaf observation of negative control and γ-PDS. **b** Ear observation of negative control and γ-cox1

and DNA in the organelles are extracted, and then a BAC library is constructed. Finally, primer walking and FGS (Sanger) sequencing methods are used to obtain original sequences for genome assembly (Liu et al. 2011; Sugiyama et al. 2005; Noyszewski et al. 2014). The advantage of this assembly process is to avoid the contamination of nuclear DNA, but the assembly is complicated and time-consuming and requires expensive kits for physical separation and purification, as well as it has certain requirements for the quality and quantity of samples. Hence, there are still many challenges. Besides, the cost of first-generation sequencing is high and the throughput is low. Thus, it is difficult to perform sequencing of large genomes.

A baiting and iterative mapping approach can be used to reconstruct mitochondrial genomes directly from genomic NGS reads (Christoph et al. 2013). The advantage of this method is that it does not require the extraction of mitochondrial organelles and mitochondrial DNA and can directly extract mitochondrial genome reads from the whole-genome reads. Combined with the NGS method, the process of mitochondrial genome assembly can be simplified, and the sequencing efficiency can be improved. This method has been widely applied to the assembly

Table 7 Setting rate of *cox1*-silenced plants

cox1-silenced plants			Negative control plants			
Grain number	Effective spikelet number	Seed setting percentage (%)	Grain number	Effective spikelet number	Seed setting percentage (%)	Statistical *P*
1	8	6%	12	6	100%	5.2342E-19
0	7	0	13	7	93%	
1	9	6%	11	6	92%	
0	8	0	11	7	79%	
0	6	0	12	7	86%	
1	6	8%	11	6	92%	
1	9	6%	13	7	93%	
1	8	6%	12	6	100%	
0	7	0	13	7	93%	
0	6	0	9	5	90%	
Average		3%			92%	

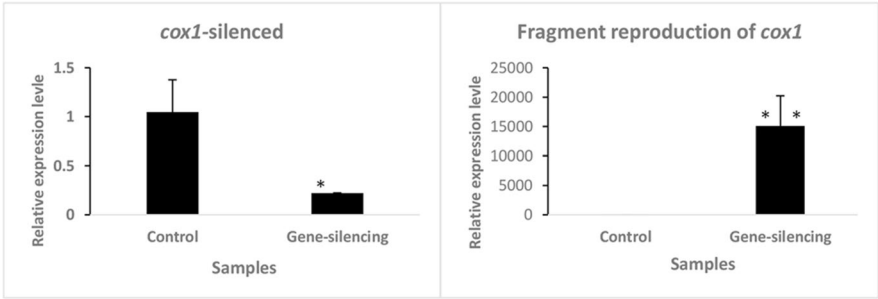

Fig. 4 Relative expression levels of gene-silenced plants. Values were calculated by the $2^{-\Delta CT\Delta CT}$ method with three biological replicates and three technical replicates, and the bar represents SE. Significant differences in the expression level were assessed by Student's *t* test (*$P<0.05$; **$P<0.01$)

of mitochondrial genomes (Bachmann et al. 2016; Groenenberg et al. 2012; Williams et al. 2014). However, the NGS method also has certain shortcomings. During the sequencing process, Illumina sequencing may have different degrees of base misreading or non-recognition, and sequencing errors may occur at both ends of reads. Besides, sequence contamination caused by the addition of adapter primers during the sequencing process will also affect the accuracy of genome assembly. Therefore, it is necessary to pre-process the original data before assembling, filter the low-quality data, and remove pollution and joint sequences (Meyer and Kircher 2010; Sun et al. 2012). The main source of sequencing errors is the replacement of bases (Aird et al. 2011). In addition, there are a few sequencing omissions in the NGS results, which are within the normal range of deviations in the NGS results. The FGS technology has the advantage of high accuracy of readings and can well handle repetitive sequences and multimeric sequences. Hence, it is necessary to use FGS for the verification of the key sites.

In this study, six protein-coding genes and four ORFs with non-synonymous mutation sites were selected and verified by FGS. It was found that the non-synonymous mutation sites of three genes and two ORFs were consistent with those of NGS. The non-synonymous mutation sites of one gene and one ORF in NGS did not exist in FGS results, and the sequences of one gene and two ORFs were not detected in the NGS results of some materials but were detected in the FGS results. These results indicate that there are some unavoidable defects in mitochondrial assembly using the baiting and iterative mapping approach based on whole-genome sequencing, and the assembly method needs to be further improved. The use of FGS to verify non-synonymous mutation sites can improve the accuracy of non-synonymous mutation sites, and the combination of NGS and FGS is an important supplementation to the current mitochondrial assembly methods.

Deletion and functional abnormality of coding genes

The mitochondrial CMS-associated genes were reported by comparing the mitochondrial genomes and gene expression between CMS and normal lines. Lin et al. (Lin et al. 2014) studied the CMSI and CMSII in tobacco (*Nicotiana sylvestris*) caused by the deletion of a mitochondrial coding gene *nad7* and found that the respiration of microspores and tapetum cells was inhibited by 60%, and the growth of leaves was inhibited as well. Luo et al. (Luo et al. 2013b) revealed that there was a *WA352* mutant gene in the mitochondria of a rice CMS line. The accumulation of *WA352* protein in the tapetum during microspore development inhibited the function of *cox1*, which was located in the mitochondria in peroxide metabolism and induced premature degradation of tapetum, leading to pollen abortion. Liao et al. (Liao et al. 2016) found that the CMS line had significantly lower expression of *cox1* than the maintainer line. They speculated that the *cox1* gene plays a very important role in the energy metabolism of pollen in *Hibiscus cannabinus* L., which is consistent with our qPCR results. Our qPCR results showed that the *cox1* gene was downregulated at the binucleate stage of fertility transition, and *cox1* was functionally related to

electron transfer of the mitochondrial respiratory chain. Pollen development is an energy consumption process, which is regulated by the expression of mitochondrial genes. When electron transfer is blocked, insufficient energy supply may cause pollen sterility. It can be speculated that the abortion of YS3038 may be attributed to insufficient protein synthesis, damage of the respiratory chain, and lack of energy production after the decrease in *cox1* gene expression.

The F_0 part of ATPase is the membrane region of enzyme proteins, which has the function of transmembrane proton transport and includes four subunits: *ORF25*, *ORFb*, *atp6*, and *atp9*. Wang et al. (Wang et al. 2010) found that *cox1*, *cox2*, *atp6*, *atp9*, *atpA*, and some other genes were different between upland cotton CMS-D2 and its maintainer line by using the RFLP method, particularly the *atpA* gene. In our study, non-synonymous mutations were found on *atp6* and *cox1* between K519A and 519B, and on *cox1* between K519A and YS3038 as well. In addition, *atp6* was differentially expressed between K519A and 519B as well as between K519A and YS3038, but the expression level was not significantly different under fertile and sterile conditions in YS3038. The *cox1* gene was differentially expressed between K519A and 519B and between K519A and YS3038, as well as between sterile conditions and fertile conditions for YS3038. The *atp6* gene was compared between the CMS and male fertile line in pepper, and the results revealed that the two copies of the *atp6* gene in CMS were one intact gene and one pseudogene. Southern blot analysis showed that there were differences in mRNA band patterns between the CMS and fertile line, indicating that the *atp6* gene is one of the candidate genes for CMS in pepper (Dong and Kim 2006). Zhao et al. (Zhao et al. 2009) found non-synonymous mutation sites on the *atp6* gene between the CMS line and maintainer line in tobacco. Then, it was revealed that the *atp6* gene fragment can be cut by the enzyme only in CMS plants, proving that the non-synonymous mutations of the *atp6* gene are related to CMS of tobacco. The deletion of genes may lead to functional defects of ATPase, which may hinder the normal synthesis of energy in plants. Many studies have shown that the deficiency in ATPase function will cause a decrease in oxidative phosphorylation function, resulting in a lower ATP content in sterile anthers than in fertile anthers (Wang and Zhou 1986; Chen and Liang 1991). This may be one of the important reasons for sterility.

Male sterility is a complex trait that involves multiple genes. In this study, we analyzed the protein-coding genes of mitochondrial genome in two sterile lines and one maintainer line. The results showed that there were some mutations in some gene sequences between different materials, and the expression levels of mutant genes were different between the sterile line and fertile line. We found that the expression of most mitochondrial genes in sterile lines was downregulated compared with those in fertile lines, indicating that the downregulation of these mitochondrial genes is involved in sterility. Besides, BSMV-VIGS is a method to achieve transient gene silencing in the current generation of plants. In this study, *cox1* of YS3038 was silenced by the BSMV-VIGS method. Silencing of the *cox1* gene led to a significant decrease in the expression level of *cox1* and seed setting rate of YS3038 under fertile conditions. These results prove that the downregulation of mitochondrial gene *cox1* is involved in sterility. Therefore, the relationship between the *cox1* gene and the fertility transformation of YS3038 was further confirmed.

Mitochondrial gene recombination and CMS

The structure of the mitochondrial genome is complex and variable, and the coding sequence of mitochondrial DNA is conserved, while the non-coding sequence is more active during evolution. Many researchers believe that CMS is related to mitochondrial DNA mutations, which often produce ORFs that can be expressed. DNA rearrangement, insertion, and deletion are the results of mutations in mitochondrial sequences. These characteristics contribute to easy occurrence of homologous recombination, integration of foreign DNA sequence, and production of chimeric mitochondrial genes during evolution. Chimeric genes formed through intramolecular or intermolecular recombination of the mitochondria are important for CMS. If proper promoters are available for the chimeric genes, they will generally be co-transcribed with normal genes in the mitochondrial genome, producing abnormal transcripts and translating them into a series of abnormal products (Bonhomme et al. 1992; Grelon et al. 1994), which will interfere with the normal function of the mitochondria at the critical

stage of pollen development and finally results in CMS (Stahl et al. 1994).

Chimeric genes in the mitochondrial genome can affect the CMS via gene-mediated network regulation, and most studies have been focused on rice CMS. Among these chimeric genes, *ORF79* related to CMS has been identified in many studies. The BT-CMS mitochondrial genome contained two duplicated copies of the *atp6* gene which encodes the F_0 part of the subunit 6 of ATPase. A special sequence located downstream of the *atp6* gene was named as *ORF79*, whose co-transcription with *atp6* was found to affect the fertility (Hiromori Akagi 1994). The *Rf1* gene, a PPR (pentatricopeptide-repeat) gene that corresponds to the 2.0 kb *atp6-ORF79* RNA consisting of 1.5 kb sequence from atp6 and 0.5 kb sequence from *ORF79*, made *ORF79* untranslated, resulting in the restoration of fertility. Therefore, it was proved that the *ORF79* chimeric gene is involved in BT-CMS in rice (Kazama et al. 2011). In addition, Wang et al. (Wang 2006) carried out a functional test of the sterile gene in CMS-BoroII rice. Through genetic transformation, the *ORF79* gene was transferred into normal rice plants, and as a result, the transgenic plants were found to be sterile. The hypothesis for the CMS is that pollen abortion is caused by the lack of energy during anther development. Therefore, the reason for pollen abortion caused by sterile protein may be that it interferes with the electron transport chain. Wang et al. (Wang et al. 2013) found that *ORF79* interacts with mitochondrial electron transport chain complex subunit III in Honglian-CMS rice. Furthermore, compared with the maintainer line, the sterile line had a lower ATP concentration while a higher level of reactive oxygen species. Therefore, it is believed that the *ORF79* protein interacts with the mitochondrial electron transport chain complex III, which reduces the activity of complex III. This defect causes ATP insufficiency, which eventually leads to pollen sterility. A recent study showed that the *WA352* gene in rice WA-CMS has no homology to any known mitochondrial genes. It is formed by the partial sequence of two known mitochondrial ORFs and inhibits the function of *cox1* and leads to pollen abortion (Luo et al. 2013a).

In previous studies of mitochondrial genomes, some ORFs found in sterile lines were also not detected in the maintainer lines. These differential ORFs in sequences are important candidates for studying the mechanism of sterility (Sun et al. 2012; Aird et al. 2011). The expression products of these CMS-related ORFs are mostly toxic proteins, which affect the normal physiological activity of some enzymes on the respiratory chain and cause programmed cell death to result in CMS. Previous studies have shown that most of the ORF proteins are related to CMS, which have a significant toxic effect on *E. coli* (Wang 2006). These toxic proteins affect the normal functions of the mitochondria in many ways. Some ORF proteins polymerize in the form of oligomers and then form a protein aggregate across the mitochondrial inner membrane. Some ORF proteins can be co-translated with relevant genes, changing the structure of protein hydrolysates so that the related genes cannot specifically bind to other subunits and affecting the normal expression of mitochondrial genes. Some ORF proteins may compete with some subunits on the mitochondrial inner membrane, affecting their normal functions. Electron transfer and oxidative phosphorylation of the mitochondrial inner membrane are directly related to the rate of ATP synthesis (Vedel 1999). When the normal function of mitochondria is affected by toxic proteins, plants cannot normally produce the indispensable ATP for life activities, and the cells will have a lack of energy, which will interfere with the development of microspores and ultimately affect the fertility of plants (Bergman et al. 2000).

Male sterility caused by RNA editing

RNA editing will cause a certain degree of change in gene expression and make the transcribed product polymorphic (Ichinose and Sugita 2017). If the genetic information is changed at the mRNA level, the final sequence of the transcript will be different from the sequence originally encoded by the gene, and it is the same case for the final translated amino acid sequence. Inappropriate editing of mitochondrial RNA may cause mitochondrial dysfunction, resulting in failure or insufficient energy supply to further causing CMS (Rurek et al. 2001). To some extent, RNA editing can also alter the water retention and hydrophobicity of proteins translated from amino acids. Zhang (Zhang et al. 2013) studied the RNA editing of mitochondrial *atp6* and *cox1* genes in CMS96 and the maintainer line in Chinese cabbage (*Brassica rapa*). The results showed that two new stop codons

and two new ORFs were produced by gene editing in the *atp6* gene of CMS96, and the editing of the *cox1* gene increased the hydrophobicity of proteins. Moreover, there were fewer editing sites of RNA in the maintainer line than in CMS96. Yang et al. (Yang et al. 2007) carried out the RNA editing of the *atp9* gene in stem mustard (*Brassica juncea* var. tumida) and found that there was a four-base variation on *atp9* between the sterile line and fertile line, which may lead to a decrease in the hydrophobicity of the *atp9* protein; besides, the dysfunction of specific mitochondrial genes caused by RNA editing may also be a factor for the occurrence of CMS.

In wheat male-sterile plants, RNA editing was performed in the *atp9* gene. After base C was changed to base U in the *atp9* gene, a stop codon was generated, and then the edited and unedited *atp9* gene were separately transformed into tobacco. As a result, the tobacco plants containing the edited gene were fertile, while those carrying the unedited gene appeared to be sterile (Hernould et al. 1993). In addition, Szklarczyk et al. (Szklarczyk et al. 2000) studied the transcription process of mitochondrial *atp9* and speculated that the conversion of C to U in the second glutamine triplet of the *atp9* may be related to fertility restoration in petaloid carrot (*Daucus carota* L. var. sativa Hoffm.). Rurek et al. (Rurek et al. 2001) conducted RNA editing on the *nad3* gene in sterile line and maintainer line in carrot (*Daucus carota* L. var. sativa Hoffm.) and found that the sterile line had C-U conversions in three bases compared with the maintainer line, which may be an important factor for CMS. Furthermore, the expression products of *nad3* after RNA editing showed polymorphisms. At present, most studies have proved that CMS is closely associated with mitochondrial RNA editing, but the mechanism is complicated and needs further research.

Conclusion

In this study, we compared the mitochondrial genomes of K519A and 519B and found 14 protein-coding genes and 83 ORF sequences different between the two genomes. There were 10 protein-coding genes and 122 ORF sequences that differed between YS3038 and K519A. The qRT-PCR results showed that the sterile line K519A had significantly lower expression of the *cox1* gene than the maintainer line, and the relative expression level of *cox1* gene in YS3038 was lower than that in K519A under sterile conditions. Gene silencing analysis showed that the silencing of *cox1* significantly reduced the fertility of plants. We speculated that the fertility of the TMS line YS3038 is probably related to the *cox1* gene.

Abbreviations CMS: Cytoplasmic male sterility; FGS: First-generation sequencing; NGS: Next-generation sequencing; TMS: Thermosensitive male sterilityAcknowledgements We thank Biomarker Technologies (Beijing, China) for helping with transcriptome sequencing and technical assistance.

Author contribution LM and QD designed and supervised the study and wrote the paper. YH planned and performed the experiments, analyzed the data, and wrote the manuscript. YG performed the experiments and analyzed the data. HZ and XZ performed the experiments. All the authors revised and approved the final manuscript.

Funding This word was supported by the National Key Research and Development Program of China (2016YFD0101602).

Data availability The sequences of mitochondrial genomes have been deposited in the GenBank database at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/genbank/) and can be accessed by the accession number SUB8115657.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol 12(2):18

Aslam S, Khan SM, Saleem M, Qureshi AS, Khan A, Islam M, Khan SM (2010) Heterosis for the improvement of oil quality in sunflower (Helianthus annuus L.). Pak J Bot 42(2):1003–1008

Bachmann L, Fromm B, Luciana P, Goeger W (2016) The mitochondrial genome of the egg-laying flatworm

Aglaiogyrodactylus forficulatus (*Platyhelminthes: Monogenoidea*). Parasites Vectors 9(1):285

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477

Belliard G, Vedel F, Pelletier G (1979) Mitochondrial recombination in cytoplasmic hybrids of Nicotiana tabacum by protoplast fusion. Nature 281(5730):401–403

Bergman P, Edqvist J, Farbos I, Glimelius K (2000) Male-sterile tobacco displays abnormal mitochondrial *atp1* transcript accumulation and reduced floral ATP/ADP ratio. Plant Mol Biol 42(3):531–544

Bonhomme S, Budar FO, Lancelin D, Small I, Defrance MC, Pelletier G (1992) Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in Brassica cybrids. Mol Gen Genet 235(2–3):340–348

Chen XF, Liang CY (1991) Relation of energy metabolism in HPGMR abortive anthers and accumulation of H_2O_2 to male sterility. Plant Physiol Commun 217(2):161–173

Chen L, Liu Y (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65(1):579–606

Christoph H, Lutz B, Bastien C (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res 13:e129–e129

Cui X, Wise RP, Schnable PS (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272(5266):1334–1336

Darling AE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14(7):1394–1403

Dieterich JH, Braun HP, Schmitz UK (2003) Alloplasmic male sterility in *Brassica napus* (CMS 'Tournefortii-Stiewe') is associated with a special gene arrangement around a novel *atp9* gene. Mol Genet Genomics 269(6):723–731

Ding D, Wang Y, Han M, Fu Z, Li W, Liu Z, Hu Y, Tang J (2012) MicroRNA transcriptomic analysis of heterosis during maize seed germination. PLoS ONE 7(6):1–9

Dong HK, Kim BD (2006) The organization of mitochondrial atp6 gene region in male fertile and CMS lines of pepper (Capsicum annuum L.). Curr Genet 49(1):59–67

Feagin JE, Gardner MJ, Williamson DH, Wilson RJM (1991) The putative mitochondrial genome of plasmodium falciparum. J Eukaryotic Microbiol 38(3):243–245

Fernie AR, Carrari F, Sweetlove LJ (2004) Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. Curr Opin Plant Biol 7(3):254–261

Fujii S, Toriyama K (2008) Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. Plant Cell Physiol 49(10):1484–1494

Grelon M, Budar F, Bonhomme S, Pelletier G (1994) Ogura cytoplasmic male-sterility (CMS)-associated *orf138* is translated into a mitochondrial membrane polypeptide in male-sterile *Brassica cybrids*. Mol Gen Genet 243(5):540

Groenenberg DS, Pirovano W, Gittenberger E, Schilthuizen M (2012) The complete mitogenome of Cylindrus obtusus (*Helicidae, Ariantinae*) using Illumina next generation sequencing. BMC Genomics 13(1):114

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Res 41(13):129

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:154–169

Hernould M, Suharsono S, Litvak S, Araya A, Mouras A (1993) Male-sterility induction in transgenic tobacco plants with an unedited *atp9* mitochondrial gene from wheat. P Natl Acad Sci USA 90(6):2370–2374

Hirokazu H (2003) 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 31(20):5907–5916

Hiromori Akagi MS (1994) Chou Shinjyo, Hiroaki Shimada, Tatsuhito Fujimura: A unique sequence located downstream from the rice mitochondrialatp6 may cause male sterility. Curr Genet 25:52

Hochholdinger F, Baldauf JA (2018) Heterosis in plants. Curr Biol 28(18):1089–1092

Howad W, Kempken F (1997) Cell type-specific loss of *atp6* RNA editing in cytoplasmic male sterile Sorghum bicolor. P Natl Acad Sci USA 94(20):11090–11095

Ichinose M, Sugita M (2017) RNA editing and its molecular mechanism in plant organelles. Genes 8(1):1–5

Itoh T, Martin W, Nei M (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. P Natl Acad Sci USA 99(20):12944–12948

Kadowaki KI, Suzuki T, Kazama S (1990) A chimeric gene containing the 5' portion of *atp6* is associated with cytoplasmic male-sterility of rice. Mol Genet Genomics 224(1):10–16

Kazama T, Nakamura T, Watanabe M, Sugita M, Toriyama K (2011) Suppression mechanism of mitochondrial *ORF79* accumulation by *Rf1* protein in BT-type cytoplasmic male sterile rice. Plant J 67(4):619–628

Kitazaki K, Kubo T (2010) Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. J Bot 2010:1–12

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Curr Genet 50(3):149–159

Kong J, Tan YP, Chen ZY, Li SQ, Zhu YG (2006) Study on the editing sites in transcripts of functional genes of HL-cytoplasmic male sterility rice mitochondria during microgametogenesis. J Wuhan Bot Res 24(2):95–99

Kovacs B, Mathijs E, Freibauer A, Brunori G, Damianova Z, Faroult E, Gomis JG, Obrien L, Treyer S (2011) Sustainable food consumption and production in a resource-constrained world. EuroChoices 10(2):38–43

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion 8(1):5–14

Landgren M, Zetterstrand M, Sundberg E, Glimelius K (1996) Alloplasmic male-sterile Brassica lines containing B. tournefortii mitochondria express an ORF 3' of the atp6 gene and a 32 kDa protein. Plant Mol Biol 32(5):879–890

Langmead B (2012) Salzberg S L : Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359

Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot Rev 38(3):425–454

Li H, Yang Q, Gao L, Ming Z, Ni Z, Zhang Y (2017) Identification of heterosis-associated stable QTLs for ear-weight-related traits in an elite maize hybrid zhengdan 958 by design III. Front Plant Sci 8:561

Liao XF, Diao Y, Qiu AH, Zhao YH, Zhou JB, Chen P, Zhou RY (2016) Cloning and expression analysis of *cox1* in cytoplasmic male sterility line UG93A and its maintainer line UG93B of kenaf. J China Agr U 21(3):36–45

Lin C, Zhang C, Zhao H, Xing S, Wang Y, Liu X, Yuan C, Zhao L, Dong Y (2014) Sequencing of the chloroplast genomes of cytoplasmic male-sterile and male-fertile lines of soybean and identification of polymorphic markers. Plant Sci 229:208–214

Liu HT, Cui P, Zhan KH, Lin Q, Zhuo GY, Guo XL, Ding F, Yang WL, Liu DC, Hu SN, Yu J, Zhang AM (2011) Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. BMC Genomics 12(1):163

Luo D, Xu H, Liu Z, Guo J, Liu YG (2013a) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet 45(5):573

Luo D, Hong X, Liu Z, Guo J, Liu Y-G (2013b) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet 45(5):573–577

Marechal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. New Phytol 186(2):299–317

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc 2010(6):5448

Nieminen AL (2003) Apoptosis and necrosis in health and disease: role of mitochondria. Int Rev Cytol 224:29–55

Noyszewski AK, Ghavami F, Alnemer LM, Soltani A, Gu YQ, Huo NX, Meinhardt S, Kianian PM, Kianian SF (2014) Accelerated evolution of the mitochondrial genome in an alloplasmic line of durum wheat. BMC Genomics 15(1):67

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. P Natl Acad Sci USA 97(13):6960–6966

Pruitt KD, Hanson MR (1991) Splicing of the Petunia cytochrome oxidase subunit II intron. Curr Genet 19(3):191–197

Rajaram S (2001) Prospects and promise of wheat breeding in the 21st century. Euphytica 119(1–2):3–15

Rurek M, Szklarczyk M, Adamczyk N, Michalik B, Augustyniak H (2001) Differences in editing of mitochondrial *nad3* transcripts from CMS and fertile carrots. Acta Biochim Pol 48(3):711–717

Ryan W, Delphine F, Reif JC, Melissa G, Takashi O, Viktor K, Peter L (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot 64(18):5411–5428

Sabar M, Gagliardi D, Balk J, Leaver CJ (n.d.) ORFB is a subunit of F1F(O)-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. Embo Rep 4(4):381–386

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3(5):175–180

Shen JX, Fu TD, Yang GS, Ma CZ, Tu JX (2005) Genetic analysis of rapeseed self-incompatibility lines reveals significant heterosis of different patterns for yield and oil content traits. Plant Breeding 124(2):111–116

Small I, Suffolk RA, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell 58(1):69–76

Stahl R, Sun S, Lhomme Y, Ketela T, Brown GG (1994) RNA editing of transcripts of a chimeric mitochondrial gene associated with cytoplasmic male-sterility in Brassica. Nucleic Acids Res 22(11):2109–2113

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomic 272(6):603–15

Sun XD, Zhou SF, Meng FL, Liu SQ (2012) De novo assembly and characterization of the garlic (Allium sativum) bud transcriptome by Illumina sequencing. Plant Cell Rep 31(10):1823–1828

Szklarczyk M, Oczkowski M, Augustyniak H, BoRner T, Linke B, Michalik B (2000) Organisation and expression of mitochondrial *atp9* genes from CMS and fertile carrots. Theor Appl Genet 100(2):263–270

Van Dingenen J, Blomme J, Gonzalez N, Inze D (2016) Plants grow with a little help from their organelle friends. J Exp Bot 67(22):6267–6281

Vedel F (1999) Lalanne é, Sabar M, Chétrit P, Paepe RD: The mitochondrial respiratory chain and ATP synthase complexes: composition, structure and mutational studies. Plant Physiol Bioch 37(9):629–643

Wang X, Zhou R (1986) The relationship between the ATP content in anthers of maize and sorghum and cytoplasmic male-sterility. Acta Agron Sin 12(3):178–180

Wang Z, Zou YJ, Li X, Zhang QY, Chen L, Wu H, Su DH, Chen YL, Guo JX, Luo D (2006) 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 18(3):676–687

Wang F, Feng C, O'Connell MA, Stewart JM, Zhang J (2010) RFLP analysis of mitochondrial DNA in two cytoplasmic male sterility systems (CMS-D2 and CMS-D8) of cotton. Euphytica 172(1):93–99

Wang K, Gao F, Ji Y, Liu Y, Dan Z, Yang P, Zhu Y, Li S (2013) *ORFH79* impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. New Phytol 198(2):408–418

Wang XQ, Yang WL, Pu MY, Liu DC, Sun JZ, Zhang WD, Zhang AM, Gao QR (2015) Comparative analysis of mitochondrial genes from wheat K- and V-type cytoplasmic male sterility lines. Mol Plant Breeding 13(8):1673–1682

Ward BL, Anderson RS, Bendich AJ (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell 25(3):793–803

Warmke HE, Lee SLJ (1978) Pollen abortion in T cytoplasmic male-sterile corn (*Zea mays*): a suggested mechanism. Ence 200(4341):561–563

Williams ST, Foster PG, Littlewood D (2014) The complete mitochondrial genome of a turbinid vetigastropod from MiSeq Illumina sequencing of genomic DNA and steps towards a resolved gastropod phylogeny. Gene 533(1):38–47

Yang J, Zhang M, Yu J (2007) Alterations of RNA editing for the mitochondrial atp9 gene in a new orf220-type cytoplasmic male-sterile line of Stem Mustard (Brassica juncea var. tumida). J Integr Plant Biol 49(5):672–677

Yasunari O, Yukiko Y, Koji M, Akira K, Toru T, Takashi S, Naohiko M, Shuhei N, Chiharu N, Naoki M (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res 33(19):6235–6250

Yi P, Wang L, Sun QP, Zhu YG (2002) Study on the editing sites in the transcript of *atp6* gene of HL-rice mitochondria. Prog Biochem Biophys 29(5):729–733

Zhang D, Li Z, Yu S, Zhang F, Yu Y, Zhao X, Wang W (2013) Study on RNA editing in Chinese Cabbage CMS96 and maintainer line. Acta Agriculturae Boreali Sinica 28(6):42–52

Zhao T, Zhu TY, Liu QY, Zhang ML (2009) SNP in tobacco mitochondrial gene *atp6* and its correlation with CMS. Acta Agron Sin 35(9):1655–1661