Assembly and comparative analysis of chloroplast genome of wheat K-CMS line and maintainer line

Yucui Han

Northwest Agriculture and Forestry University

Yujie Gao

Northwest Agriculture and Forestry University

Xiaoguang Zhai

Northwest Agriculture and Forestry University

Hao Zhou

Northwest Agriculture and Forestry University

Qin Ding

Northwest Agriculture and Forestry University

Lingjian Ma (✉ malingjian@nwsuaf.edu.cn)

Northwest Agriculture and Forestry University

Research article

Keywords: Chloroplast, Cytoplasmic male sterility, Genome, Non-synonymous mutation, SSR

Posted Date: March 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-15231/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The sterile line is vital for the heterosis utilization of wheat (*Triticum aestivum* L.), and the cytoplasmic male sterility (CMS) line has more practical significance in the utilization of heterosis. To reveal the sterile mechanism of wheat K-CMS line K519A from the perspective of the chloroplast genome, the chloroplast genomes of the K519A and its same nuclear genotype maintainer line 519B were sequenced using second-generation high-throughput technology and assembled. Then the chloroplast genomic contents, SSR sequences, long repeat fragments, and qPCR were analyzed. The results showed that the total length of K519A and 519B were 136,996 bp and 136,235 bp, respectively. Both chloroplast genomes encode 126 genes, including 89 protein-coding genes, 8 rRNA genes, and 39 tRNA genes. There were 186 and 188 SSRs in K519A and 519B, respectively. And a developed SSR maker, which from *atpF*, can be used to distinguish the cytoplasmic type of *Aegilops kotschy* and *Triticum aestivum*, respectively. There were 50 and 52 long repeat fragments, which come from the gene sequences, in K519A and 519B, respectively. A total of 107 mutations between K519A and 519B, which from 45 genes, were identified, of which, 16 genes (*matK*, *rps16*, *rpoB*, *rpoC2*, *atpI*, *ndhK*, *atpB*, *rbcL*, *psbB*, *psbH*, *rpl14*, *ndhH*, *ndhF*, *rpl32*, *ccsA*, *ndhA*) contained non-synonymous mutations. Further, the qPCR analysis was performed, and the gene expression levels of selected six genes for K519A compared to 519B were mostly downregulated at the binucleate and trinucleate stages of pollen development, thus, these non-synonymous mutant genes might affect the fertility of K519A.

1. Introduction

Wheat (*Triticum aestivum* L.) is the world's most widely grown food crop, feeding nearly half of the world's population [1, 2]. It is also an important raw material for industry. For example, wheat gluten can be used as a natural binder in the manufacture of paper [3]. And dry gluten extracted from wheat can be used as an additive for bread processing [4]. In recent years, with the acceleration of the global urbanization process, the planting area of wheat has been gradually declining, which may impact the food problem of human beings. Improving yield per unit area and stress resistance of wheat are very effective strategies to alleviate the food problem. The utilization of heterosis is an effective way to improve wheat yield and quality, and it plays a very important role in the breeding of crops [5, 6].

Wheat is a monoecious, self-pollinated crop. Therefore, male sterile lines are essential for utilizing of heterosis. Male sterility in plants refers to the development of gynoecium is normal during the growth process, but fertilization and seed setting after receiving pollen are abnormal, which caused by the abnormal development of stamens. The abnormal development of stamens means that the pollens are abortive or anther tissue structure is abnormal, which eventually leads to the decline or loss of pollen vitality. The causes of male sterility in plants are various. According to the origin of male sterile genes, male sterile lines can be divided into three types: genic male sterility (GMS), cytoplasmic male sterility (CMS) and nucleus-cytoplasmic male sterility. At first, Kihara [7] transferred the nucleus of common wheat into *Ae. Caudata* and cultivated the first nucleus-cytoplasmic male sterile line in 1951. Wilson and Ross [8] introduced the common wheat nucleus into the *T. timopheevi*. Finally, they successfully selected and obtained T-type CMS wheat. After this experiment, the maintainer and restorer lines were also selected,

these studies were important for the utilization of wheat heterosis [9]. After years of unremitting efforts scientists, there are many different cytoplasmic sterile lines that have been cultivated in China [10]. The K-type wheat sterile line has the *Aegilops kotschy* cytoplasm, S-type has *Aegilops crassa* cytoplasm, Q-type has wild oat cytoplasm, and V type has *Aegilops variabilis* cytoplasm, and they all have the nuclear of common wheat.

In general, the genetic systems of higher plants mainly include the nuclear genome and the cytoplasmic genome. The cytoplasmic genome mainly includes the mitochondrial and chloroplast genomes. Both mitochondria and chloroplast inheritance belong to the semi-autonomous genetic system, which is capable of autonomous DNA replication, transcription and translation [11]. Besides, many physiological metabolisms of plant cells depend on the combined action of nuclear, mitochondria, and chloroplast [12, 13]. At present, scholars from all over the world have conducted a lot of research on the physiological and biochemical mechanisms of cytoplasmic male sterility in wheat [14, 15].

Chloroplasts are important organelles for photosynthesis in higher plants and are important organelles with independent genetic information in cells. The chloroplast genome, also known as chloroplast DNA, refers to the sum of the genetic information contained in the chloroplast [16]. Most chloroplast genomes are 120–250 kb in size and have a closed circular double-stranded structure [17]. They are typically four-segment structures consisting of two inverted repeats (IRs), a short single copy region (SSC) and a long single copy region (LSC). The sequence of the IRs' regions is the same but in the opposite direction. The structures of the chloroplast genomes are relatively conservative and have fewer changes during the evolution of species. Therefore, the chloroplast genome can provide reliable and accurate information for biodiversity research and provide a lot of basic information for comparative evolution research [18, 19]. Some special structural compositions in chloroplasts were also considered to be related to the fertility of plants [20]. Li et al. [21] used SDS-PAGE and two-dimensional electrophoresis to compare the differences of peptides on chloroplast thylakoid membranes between CMS lines and their maintainers in beet, corn, and sorghum. There were significant differences in the size, number, and distribution of peptide plaques between two lines, proving that the composition of chloroplast thylakoid polypeptides might affect cytoplasmic male sterility. Li et al. [22] studied the chloroplast ultrastructure of the CMS line and its maintainer in rape, the number of stacks of grana in the thylakoid of the sterile line was significantly reduced. And the lamellar structure between granas became thinner or even fractured so that the lamellar system was disordered. Chen et al. [23] found that after the digestion of chloroplast genome DNA of sorghum male sterile line and maintainer line with HindIII endonuclease, the male sterile line produced a 3.7 kb specific fragment, with a 165 bp lack compared to the maintainer line (3.8 kb). The deletion fragment was from the *rpoC2* gene, which encodes the β subunit of the RNA polymerase. The amino acids of the deletion fragment will participate in the formation of an alpha helix. It was found that identified 1501 differentially expressed transcripts in leaves and anthers at different developmental stages were most of these DETs being localized in plastid and mitochondrion in *Brassica napus* L. male sterile line induced by the chemical hybridization agent compared to the wild *Brassica napus* L. [24].

K-type 1B/1R CMS line is a heterozygote with *Aegilops kotschy* cytoplasm and 1B/1R *Triticum aestivum* L. nucleus [25]. On this basis, K-type non-1B/1R CMS lines were cultivated. It has the pollen of complete

abortion, stable sterility, wide recovery source, good agronomic traits, so it has great application value. There have been some researches about K-CMS in physiology [26], cytology [27], protein expression [28] and gene function analysis [29]. For instance, Yang et al. [29] cloned and analyzed the functions of MADS-box genes, TaAG-A and TaAG-B, from a wheat K-CMS line. They found that TaAG-A and TaAG-B had relatively high expression levels in spikes. In the maintainer line, both genes showed downregulation during the uninucleate to the trinucleate stage. Then the TaAG-A and TaAG-B were silenced in fertile wheat lines using the VIGS method, resulting in green and yellow striped leaves, emaciated spikes, and decreased seed set rates. Zhang et al. [28] found a total of 41 protein spots differentially expressed through proteomic analysis of pollen in K-CMS line and maintainer, among which, all seven identified proteins were downregulated in CMS line, but there were no significant differences in the maintainer line at the uninucleate stage. The caffeic acid O-methyltransferase was downregulated in CMS and the other five proteins were upregulated in CMS at the binucleate stage, except glutamine synthetase isoform GS1a. However, genome sequence analysis of chloroplast of K-CMS in wheat has not been published.

K519A is a non-1B/1R K-CMS line, and 519B has the same nuclear genotype maintainer line of K519A. Their nuclear genomes are derived from the same paternal species, but their cytoplasm genomes are from different maternal species. Thus, the differences in fertility come from the differences between cytoplasmic genomes. To understand the sterile mechanism of K519A, the chloroplast genomes of K519A and 519B were sequenced, assembled, annotated and compared in this study. Then the non-synonymous mutant sites of coding genes between K519A and 519B were verified, and the expressed patterns were analyzed. Besides, SSR, long repeat sequences, and phylogeny were analyzed. This study can provide an important reference value for the sterile mechanism study of K519A and its genetic relationship.

2. Materials And Methods

2.1 Material planting

The test materials are K-type male sterile line K519A (non1B/1R *Triticum aestivum* L. with *Aegilops kotschy* cytoplasm) and its same nuclear genotype maintainer line 519B (*Triticum aestivum* L.). They are provided by our laboratory (Northwest A&F University, Yangling, Shaanxi, China). After pollens matured, the opening angle of the glumes of K519A become larger than 519B, the anthers stretch out from the glume (Fig. 1- A1, B1). And the anther is wizened, short, and crooked, and the pollens can't spread out from the anther (Fig. 1- A2). After pollens matured, the anthers of 519B are plump and cracked, and amount of fertile pollens spread out from the anthers (Fig. 1- B2). The test materials were planted in the experimental field of the Northwest A&F University (108°4'E, 34°16'N) on October 3, 2018. The two materials were planted in 4 rows (25 cm between rows), respectively, with 15 seeds in 1 m. Planting with general field management measures. Young leaves were stored in -80°C refrigerator in March 2019.

2.2 CTAB extraction of DNA and chloroplast genome sequencing

The extraction of genomic DNA from young leaves of K519A and 519B wheat was by the CTAB method. The whole-genome DNA sequences were obtained using second-generation high-throughput technology, with the Illumina HiSeq XTM Ten sequencing platform and paired-end 150 bp sequencing strategy. Firstly, the 10G original sequencing data for each sample was obtained, then low-quality reads and connector sequences are removed. And the clean data was used for further genome assembly of the chloroplast.

2.3 Genome assembly and annotation

The chloroplast assembly was carried out by the assembly method of Hahn et al. [30]. The assembly steps are as follows: Firstly, the clean data is spliced, using the SPAdes [31] software, setting default parameters except that the cut off parameter is not selected. The clean data are spliced into scaffolds. Using the published wheat chloroplast DNA sequences and protein-coding gene sequences as a reference, then sequences are aligned using Blastn and Exonerate, respectively. The evaluate threshold of DNA alignment is $1e^{-10}$, and the protein similarity threshold is 70%, respectively. Using PRICE and MITObim to splice the collected fragmented sequences to extend scaffolds. The number of iterations generally is 50. The original sequencing reads are aligned to scaffolds using bowtie2, then the matched paired reads are picked out and spliced by SPAdes. Look at the path to see if a distinct circle diagram can be formed. If so, the circle genome is extracted, otherwise, the above 3–5 steps are repeated until a circular genome is formed.

The organelle genome annotation is mainly divided into three parts: protein-coding gene annotation, RNA annotation, and structural annotation. Three methods are commonly used for protein-coding gene prediction. The DNA sequences or the protein-coding sequences directly align in NCBI to confirm the genes. Submitting sequences to the online service annotation tools DOGMA (<http://dogma.ccbb.utexas.edu/>) and CpGAVAS [32], predicting the ORFs, and then annotating genes in the nr database. In combination with the three methods and choose the most accurate annotation results.

Chloroplast tRNA annotations are performed using tRNAscan-SE online site (<http://lowelab.ucsc.edu/tRNAscan-SE/>). The rRNA annotations are performed using the RNAmmer 1.2 Server (<http://www.cbs.dtu.dk/services/RNAmmer/>), in combination with homologous alignments, and the boundaries correcting. After the sequence annotation is completed, the structure note is edited by Sequin to generate a submission file that can be submitted to the GenBank database. Submit the file to OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) to draw the annotated map.

2.4 Codon preference analysis of chloroplast gene

A codon is a link between genetic information from DNA and protein, and every three adjacent codons encode an amino acid. In all organisms in nature, there are only 20 amino acids in total, however, each amino acid corresponds to at least one or up to six codons. Synonymous codons mean that different codons can encode the same amino acid [33]. The analysis of codon preference is performed using the CHIPS program in the EMBOSS software package [34].

2.5 Simple sequence repeat (SSR) and long repeat sequences analysis

The Microsatellite identification tool (MiSa) [35] is used for finding simple sequence repeat (SSR) in the genome. We used this software to detect SSR with eight times for minimal repeat number of mononucleotides, five times for dinucleotides, and at least three times for three or more bases.

The polymorphism SSR markers at the same position in the same gene of K519A and 519B were selected. Primer Premier 6.0 software was used to design SSR primers to verify the polymorphism between K519A and 519B.

Using REPuter online software [36] to find long repeat sequences. All options: forward, reverse, complement, and palindromic are selected for "Match Direction". Set the parameter "Hamming distance" to 3, "maximum computed repeats" to 50, and "minimal repeat size" to 15.

2.6 Gene alignment analysis and phylogenetic analysis

We used DNAMAN software to align the homology of each coding gene in K519A and 519B, then we selected all non-synonymous mutant sites to verify the accuracy of second-generation sequencing results. According to the company's sequencing sequence, Primer Premier 6.0 software is used to design primers of non-synonymous mutant sites, then PCR products were sent to the company for first-generation sequencing (FGS). The DNAMAN software was used to align the gene sequences and amino acids of K519A and 519B.

The genome-wide alignment of chloroplast genomes A, B and other 100 chloroplasts of near-source species were performed using the HomBlocks (<https://github.com/fenghen360/HomBlocks>) software with the trimAl trim method. Using the NJ method to build the evolutionary tree with the 1000 bootstrap value.

2.7 RNA extraction and qPCR

The extraction of RNA from anthers using the TaKaRa miniBEST Plant RNA Extraction Kit (Takara NO.9769). The DNase I reaction solution was used to effectively remove genomic DNA during RNA extraction so that the extracted RNA generally does not contain genomic DNA. Then, agarose gel electrophoresis was used for quality testing. The total RNAs were reverse-transcribed into cDNA using PrimeScript RT Reagent Kit (Takara). The TB Green™ Premix Ex Taq™ II (Tli RNaseH Plus) Kit (Takara Biological Engineering) was used for the qPCR to study gene expression patterns with three biological repeats. The target gene-specific primers were designed using Premier6.0 software. qPCR was performed with Q7 Real-Time PCR instrument produced by Life Technologies, USA. Using Actin as the reference gene and the expressions of target genes were calculated by the $2^{-\Delta\Delta CT}$ method.

2.8 Data Availability Statement

All data of chloroplast genomes generated in this study have been deposited in NCBI GenBank, with accession number MN605257 (K519A) and MN605258 (519B).

3. Results

3.1 Genome content and characteristics

Like the chloroplast genomes of other higher plants, the two chloroplast genomes in this experiment contained two inverted repeats, IRA and IRB, which divided the entire genome into four parts, and the remaining regions were the large single-copy region (LSC) and the small single-copy region (SSC). The total length of the chloroplast genome of K519A was 136,996 bp, and the lengths of the four parts were 81,136 bp (LSC), 12,776 bp (SSC), and 21,542 bp (IRA and IRB) (Fig. 2). The total length of the 519B was 136,235 bp, and the lengths of the four parts were 80,335 bp (LSC), 12,796 bp (SSC), and 21,552 bp (IRA and IRB) (Fig. 3). The GC contents of genomes were 38.27% (K519A) and 38.28% (519B) respectively. The structure of the two chloroplast genomes was different, such as the expansion and contraction of the boundaries of the four parts, and the sequence insertion of some regions.

Both chloroplast genomes encoded 126 genes, including 89 protein-coding genes, 8 rRNA genes, and 39 tRNA genes. Chloroplast genes were mainly divided into three categories: photosynthesis genes (45), self-replication genes (59), and unknown function biosynthesis genes (8) (Table 1). Of these, the protein-coding genes *rps12*, *rps19*, *rpl2*, *ycf2*, *ndhB*, *rps7*, *ycf1*, *rps5*, and *ndhH* had two copies, and the *rpl23* gene had three copies. In the tRNA genes, *trnA*-UGC, *trnH*-GUG, *trnI*-GAU, *trnN*-GUU, *trnV*-GAC, and *trnR*-ACG contain two copies respectively. The *trnM*-CAU gene had three different copies between two chloroplast genomes, respectively. Besides, *trnK*-UUU, *rps16*, *trnG*-UCC, *atpF*, *trnL*-UAA, *trnV*-UAC, *petB*, *petD*, *rpl16*, *rpl2*, *trnI*-GAU, *trnA*-UGC, *ndhA*, *ndhB* genes contained an intron respectively, *ycf3* gene contained two introns, and *rps12* gene had a trans-splicing situation.

The regions of exons and introns for each gene were predicted, and the codon preferences were analyzed based on the sequence of exons. The codon preferences of K519A and 519B were different. And the codon preferences of genes in the K519 and 519B were shown in Table 2 and Table 3, respectively. Among them, both Met and Trp were encoded by only one codon. Ile and stop codons were encoded by three. Ala, Gly, Pro, Thr, and Val were encoded by four. Arg, Leu, and Ser were encoded by six kinds.

3.2 Discovery of SSRs

The SSR markers of chloroplasts have been applied for the studies of population genetics, evolution processes and phylogeny [37, 38]. In this study, we found 186 SSR repeat motifs in K519A (Table 4) and 188 in 519B (Table 5). Of SSRs, the mononucleotide repeats were the most abundant compared to polynucleotide repeats, and the number of A/T repeat motifs was significantly higher than the number of C/G, which was consistent with the results of six legume plastid genomes [39]. In K519A, the proportion of single nucleotide repeat motifs was 68.28% and was 67.02% in 519B. The SSR number composed of dinucleotide was 8, of which, 6 AT/TA repeat motifs and 2 TC in K519A. The SSR number composed of dinucleotide was 9, of which, 7 AT/TA repeat motifs and 2 TC in 519B. Among the trinucleotides, AAC/TTC most frequently appeared. And the SSR number of the remaining repeat motifs type appeared only once.

After analysis, it was found that there was one SSR marker with the polymorphic difference between K519A and 519B from the five coding genes, respectively. Thus, PCR was performed after primers were designed for the five sites (Table 6), and the bands of SSR markers from the *atpF* gene were polymorphic between K519A (200 bp) and 519B (197 bp) according to the polyacrylamide gel electrophoresis (Fig. 4).

3.3 Discovery of Long repeat sequences

Besides to SSR analysis, we also analyzed long repeats of the chloroplast genome. Long repeat sequences are special DNA sequences that occur repeatedly in the genome and usually occupy a large proportion in the genome. In general, the proportion of repeat sequences is positively related to the size of the genome. For example, the Arabidopsis genome is 120 Mb, and the repeat sequences account for 10%. The wheat genome is 17,000 Mb, and the repeat sequences account for 81% [40, 41]. Most repeat sequences exist in the intergenic region, and a small number of repetitive sequences exist in the gene coding region [42]. Repetitive fragments have important molecular significance for plant evolution research [43]. In the K519A, a total of 50 long repeats were found, and the sizes were 20–286 bp. Of the 50 repeats, the positive repeats, palindrome repeats, and inverted replicates were 37, 11, and 2, respectively. The longest repeat was the forward repeat with 286 bp in length (Table 7). In 519B, a total of 52 long repeats were found, the sizes of which were also 20–286 bp, which was consistent with K519A, of which, positive repeats, palindrome repeats, and inverted repeats were 39, 10, and 3, respectively (Table 8).

3.4 Gene alignment analysis between K519A and 519B

A comparative analysis of the K519A and 519B revealed that 107 SNPs were identified in the protein-coding sequence in 45 genes (Table 9). Among them, 75 were synonymous mutations, 32 were non-synonymous mutations, the proportion of non-synonymous mutations was 29.9% of the total mutations. Of non-synonymous mutations, the RNA polymerase-related genes are the most, and six genes occurred mutations. And followed by the ATP synthase-related genes and the NADH dehydrogenase-related genes, with five mutations occurred, respectively. Among the 45 genes that had mutated, 29 gene sequence differences were synonymous mutations that the encoded amino acid sequences were identical. The base conversion occurred 73 times and the transversion occurred 34 times. Among the 126 genes, the gene with the largest difference between K519A and 519B sequence was the *rpoC2* gene, and there were 12 SNPs with 5 amino acid differences. The number of tRNA and rRNA genes was completely consistent between the K519A and 519B chloroplast genomes, respectively.

3.5 The verification of non-synonymous mutant sites by first-generation sequencing

To verify the accuracy of second-generation sequencing results, 32 non-synonymous mutant sites from 16 genes, based on the results of the above alignment analysis, were identified via FGS technology. These genes included *matK*, *rps16*, *rpoB*, *rpoC2*, *atpI*, *ndhK*, *atpB*, *rbcL*, *psbB*, *psbH*, *rpl14*, *ndhH*, *ndhF*, *rpl32*, *ccsA*, and *ndhA*. The corresponding primers were listed in Table 10. Then, we designed primers for all non-synonymous mutant sites and performed FGS analysis to compare the sequences of the FGS with those of the second-generation sequencing, and it was concluded that the amino acids of 16 genes were all consistent under two methods, excepting some mutated bases. Some comparison results were listed in Fig. 5

3.6 Analysis of qPCR

To further study the effect of non-synonymous mutant genes in the chloroplasts, we performed a qPCR analysis to assess the expression patterns of genes. According to the existing research results, we selected six non-synonymous mutant genes (*atpB*, *ccsA*, *matK*, *rbcL*, *rpoC2*, and *ndhH*) for analysis of gene expression levels of K519A compared with 519B. The corresponding primers of qPCR were listed in Table 11.

It can be seen from the figures that the expression levels of all genes were downregulated at the binucleate and trinucleate stages, except that the *rbcL* gene was downregulated only at the trinucleate stage. And the *rpoC2* gene was upregulated at the late uninucleate stage (Fig. 6).

3.7 Phylogenetic analysis of K519A and 519B

The cluster analysis and evolutionary tree of the chloroplast genome of K519A, 519B, and other 100 species that have near relatives were performed. The *Bromus vulgaris* was used as the outgroup. It can be concluded that the chloroplasts of K519A and 519B were from two independent taxa, the chloroplast of 519B was closely related to the *Triticum timopheevii*, *Aegilops spelta* L., and other *Triticum* species. However, K519A was closely related to the *Aegilops* genus and other *Aegilops* species (Fig. 7), which was consistent with the characteristic that K519A's cytoplasm was derived from *Aegilops kotschyi*. This is consistent with the characteristics of the materials themselves, the K519A has the cytoplasm of *Aegilops* genus, 519B has a common wheat cytoplasm.

4. Discussion

4.1 Plant chloroplast genomics research and phylogenetic analysis

The chloroplast genome sequencing technology has developed from the FGS to the current third-generation sequencing [44–46]. Currently, most of the chloroplast genome sequences on NCBI are performed by second-generation sequencing technology [47, 48]. Due to the fast improvement of sequencing technology, results in that the sequencing costs continue to decrease, so chloroplast genome data is largely supplemented. In the past few decades, chloroplasts studies have made a breakthrough. More than 2400 plant chloroplast genomes have been published on the NCBI database. Of which, the chloroplast genome of tobacco is the first genome to be sequenced in higher plants [45]. The chloroplast genome encodes 100–120 genes, including 70–88 protein structural genes, 30–32 tRNA genes, and 4 rRNA genes. According to the functional classification of chloroplast genes, they can be divided into three categories: the first are genes encoding photoreactive structural proteins, such as *psaA*, *psbB*, *petD*, etc. The second are the coding genes of energy metabolism-related enzymes, such as *atpA*, *ndhB*, *rbcL*, and other genes. The third are the encoding genes of the transcriptional translation system, mainly including the ribosome large subunit encoding *rpl* family, the ribosome small subunit encoding *rps* family, the transport RNA encoding *trn* family, and transport RNA polymerase encoding *rpoA* family and so on. Besides, there are some genes whose functions have not yet been determined, such as the *ycf* family [49]. Researches have shown that many genes of plant chloroplast are involved in some important biological processes. For instance, Zhong et al.

[50] demonstrated that chloroplast heat shock protein HSP21 is involved in the plastid-encoded RNA polymerase (PEP)-dependent transcription in *Arabidopsis thaliana*. Yu et al. [51] found that the downregulated expression of ribosomal protein S1 (RPS1) in *rps1* mutants negatively modulates the expression of heat-responsive gene *HsfA2* and its target genes in *Arabidopsis thaliana*.

Our study has completed the sequencing of the chloroplast genome of the K-type CMS K519A and maintainer 519B. This is similar to the chloroplast genome structure of most plants, which is a typical four-segment structure. The chloroplast genome of angiosperms belongs to the maternal inheritance and is generally not affected by genetic recombination. The chloroplast genome is much smaller than the nuclear genome and there are more copies in normal cells, so the chloroplast genome is easy to obtain. And the structure and coding genes of the chloroplast genome are relatively conserved [52, 53]. In recent years, with the continuous completion of plant chloroplast genome sequencing, phylogenetic researches based on chloroplast genomes or genes have been greatly promoted. Kallersjo et al. [54] utilized the *rbcL* chloroplast gene to analyze the phylogenetic relationship of 2538 species, covering the prokaryotic cyanobacteria to higher flowering plants, using parsimony jackknife analysis method. Bruneau et al. [55] analyzed the evolutionary relationship of 223 legume materials by analyzing the introns of the chloroplast *trnI* gene. Kim [56] sequenced the chloroplast genome of *Panax schinseng* Nees and analyzed the evolutionary relationship about vascular plants. Liu [57] analyzed and reported the chloroplast genome of *Morella rubra* and the evolution patterns of Fagales, proved that Myricaceae is a sister to Juglandaceae.

When analyzing evolutionary relationships, the sequencing depth and coverage, annotation methods, gene losses, and pseudogenizations possibly affect the results [58]. It is not difficult to see from the successful application of next-generation sequencing technology in the acquisition of chloroplast genomic data in recent years that the emergence of this new technology will greatly promote the development of chloroplast phylogenetic analysis, and the phylogenetic research of plants will have a bright prospect [59, 60].

4.2 SSR molecular marker

Molecular marker technology is to detect the polymorphism of DNA sequences based on PCR, which can directly compare the mutations of DNA sequences. With the continuous improvement of molecular biology technologies, it is becoming more and more mature and is applied to many research directions [61, 62]. Among the molecular markers, the SSR marker has high variability, codominant inheritance, easy operation and easy analysis [63]. It is widely used in genetic mapping construction, gene mapping, genetic diversity analysis and variety identification of various crops. Such as, Stachel et al. [3] successfully analyzed the genetic diversity of 60 wheat cultivated varieties originating from three agro-ecological zones using microsatellite markers. Salameh et al. [64] used a molecular marker to transfer the resistant QTL (*Fhb1*) on the 5A chromosome into nine European winter wheat lines to develop new scab-resistant lines. Besides, Shim et al. [65] sequenced the whole chloroplast genome of millet and compared it with other reported chloroplast genomes, and obtained 125 SSR loci and 34 indels changes, which can provide effective information for the phylogeny. In our study, SSR polymorphic analysis between the K519A and 519B on the chloroplast genes was performed. Among identified 5 SSR polymorphic markers, the SSR marker from *atpF*

was deeply demonstrated by polyacrylamide gel electrophoresis. Thus, this SSR marker can be used for the chloroplast and cytoplasmic type identification for the *Aegilops kotschyi* and *Triticum aestivum* L.

4.3 Analysis of chloroplast gene related to cytoplasmic male sterility

Some nuclear male sterility genes have been identified. For example, Xing et al. [66] cloned the gene TaAPT2 related to thermosensitive sterility in wheat, and the gene expression levels of the young anthers, which under fertility conversion stage, were significantly reduced under the sterile condition by Northern analysis, indicating that the TaAPT2 gene may be related to the fertility. Xia et al. [67] confirmed that the male-sterile ms2 mutant in wheat was caused by an insertion of terminal-repeat retrotransposons in miniature (TRIM) element in the promoter region of Ms2 gene. In addition to this, some studies have shown that the deletion, duplication or mutation of plant chloroplast DNA may cause the wrong transmission of genetic information between chloroplast, mitochondrial and nuclear, leading to male sterility [68, 69].

Some studies have found that some chloroplast genes in plants may affect the fertility of plants. In this study, it was found that some genes had non-synonymous mutant sites between K519A and 519B, and these genes might be related to male sterility. For example, the matK gene is located in the intron of the chloroplast trnK gene. Jia et al. [70] extracted DNA of abortive flower buds from radish male sterile line BT-18 and normal flower buds using DD-PCR technology to study the differential expression of RNA between abortive and normal buds in Radish. And it was found that the matK gene appeared three times in different primers of PCR amplification, and the amplification bands were enhanced in abortive buds, which showed upregulated expression, indicating that the abortion of radish buds had a great relationship with matK gene. Ou et al. [71] studied the intron sequences of ORFs and rps16 genes in chloroplasts of different CMS lines in rice. And they found that in the gametophyte sterility line, the rps16 gene intron contains a GMTGAG sequence and a unique G at position 595, which can be used as a molecular marker to distinguish sporophytic sterility and gametophyte sterility.

The RNA polymerase is a key enzyme controlling chloroplast genomic transcription and gene expression. It is encoded by the copy genes rpoA, rpoB, rpoC1 and rpoC2 genes [72]. The α subunit and β subunit of RNA polymerase are encoded by rpoA and rpoB genes respectively and the β' and β'' subunit by rpoC1 and rpoC2 genes. Chen et al. [73] found a new chloroplast DNA deletion in the sorghum CMS line, which occurs in the middle of the rpoC2 gene, which may be associated with cytoplasmic male sterility in sorghum. We infer that the non-synonymous mutations of rpoB and rpoC2 genes may affect the production of β and β'' subunits of RNA polymerase.

In K519A, the expression levels of atpB, ccsA, matK, rbcL, rpoC2, and ndhH genes were mostly downregulated at the binucleate and trinucleate stages. Previous studies found that the key periods of pollen abortion in non-1B/1R and 1B/1R wheat male-sterile lines were binucleate and trinucleate stages [74]. Therefore, the result of this study was coincident with the previous research[26].

4.4 Chloroplast genome fragment transfer phenomenon

In the chloroplast of higher plants, a phenomenon is that the chloroplast gene can be transferred to the nucleus, which is called chloroplast genome fragment transfer. The phenomenon has been found in tobacco [64]. However, the genes in the nucleus are not transferred to the chloroplast [66]. Besides, the chloroplast gene can also be transferred to mitochondria, and the genes that are transferred to mitochondria often evolve into non-functional pseudogenes [75]. We speculate that the combination of different types of cytoplasm and nucleus may destroy the intrinsic balance between chloroplast, nucleus, and mitochondria, which may lead to the formation of cytoplasmic male sterility.

5. Conclusion

In this study, we obtained and analyzed two chloroplast genomes of CMS line K519A and maintainer line 519B in wheat, which encoded 126 genes, respectively. Based on a comparative analysis of K519A and 519B, a total of 107 mutations were found in 45 genes. 16 genes (*matK*, *rps16*, *rpoB*, *rpoC2*, *atpl*, *ndhK*, *atpB*, *rbcL*, *psbB*, *psbH*, *rpl14*, *ndhH*, *ndhF*, *rpl32*, *ccsA*, *ndhA*) contained non-synonymous mutations. The *atpB*, *ccsA*, *matK*, *rbcL*, *rpoC2*, and *ndhH* genes were mostly downregulated at the binucleate and trinucleate stages. The developed SSR maker can be used to distinguish the cytoplasmic type of *Aegilops kotschy* and *Triticum aestivum*.

Abbreviations

CMS Cytoplasmic male sterility

GMS Genic male sterility

FGS First-generation sequencing

Declarations

Author contributions

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

Funding

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

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] S. Asseng, I. Foster, N.C. Turner, The impact of temperature variability on wheat yields, *Global Change Biol.* 17 (2011) 997-1012.
- [2] S.S. Hussain, A. Rivandi, Molecular breeding for drought tolerance in plants: wheat perspective, *Proc. Pakistan Acad. Sci.* 44 (2007) 35-62.
- [3] M. Stachel, T. Lele, H. Grausgruber, J. Vollmann, Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use, *Theor. Appl. Genet.* 100 (2000) 242-248.
- [4] L. Day, M.A. Augustin, I.L. Batey, C.W. Wrigley, Wheat-gluten uses and industry needs, *Trends Food Sci. Tech.* 17 (2006) 82-90.
- [5] W. Li, S. Liu, The relationship among yield structure heterosis of strong heterosis hybrid wheat, *Acta Triticea* 22 (2002) 1-6.
- [6] L.m. Wu, Z. Ni, Z. Wang, Z. Lin, Q. Sun, Relationship between differential expression patterns of multigene families and heterosis in a wheat diallel crosses, *Acta Genetica Sinica* 28 (2001) 256-266.
- [7] H. Kihara, Substitution of nucleus and its effects on genome manifestations, *Cytologia* 16 (1951) 177-193.
- [8] J.A. Wilson, W.M. Ross, Male sterility interaction of the *Triticum aestivum* nucleus and *Triticum timopheevi* cytoplasm, *Wheat Information Serv.* 14 (1962) 14-29.
- [9] J.A. Wilson, W.M. Ross, Crossbreeding in wheat, *Triticum aestivum* L. frequency of the pollen-restoring character in hybrid wheats having *Aegilops ovata* Cytoplasm1, *Crop Sci.* 1 (1961) 191-193.
- [10] G. Zhang, T. Yang, A preliminary on the Male Sterile Lines of Wheat with *Ae. ventricosa*, *Ae. kotschy* and *Ae. vauiabilis* Cytoplasms, *Crop J.* 15 (1989) 1-10.
- [11] S.R. Downie, R.K. Jansen, A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions, *Syst. Bot.* 40 (2015) 336-351.
- [12] E. Sundberg, J.G. Slagter, I. Fridborg, S.P. Cleary, C. Robinson, G. Coupland, *ALBINO3*, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria, *Plant Cell* 9 (1997) 717-730.
- [13] J.D. Woodson, J. Chory, Coordination of gene expression between organellar and nuclear genomes, *Nat. Rev. Genet.* 9 (2008) 383-395.
- [14] P. Bergman, J. Edqvist, I. Farbos, K. Glimelius, Male-sterile tobacco displays abnormal mitochondrial *atp1* transcript accumulation and reduced floral ATP/ADP ratio, *Plant Mol. Biol.* 42 (2000) 531-544.

- [15] R. Mittler, Oxidative stress, antioxidants and stress tolerance, Trends Plant Sci. 7 (2002) 405-410.
- [16] S. Wicke, G.M. Schneeweiss, C.W. dePamphilis, K.F. Müller, D. Quandt, The evolution of the plastid chromosome in land plants: gene content, gene order, gene function, Plant Mol. Biol. 76 (2011) 273-297.
- [17] R.K. Jansen, L.A. Raubeson, J.L. Boore, C.W. Depamphilis, L. Cui, Methods for obtaining and analyzing whole chloroplast genome sequences, Method Enzymol. 395 (2005) 348-384.
- [18] A. Sajjad, W. Muhammad, K. Abdul Latif, K. Muhammad Aaqil, K. Sang-Mo, I. Qari Muhammad, S. Raheem, B. Saqib, Y. Byung-Wook, L. In-Jung, The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species, Front. Plant Sci. 8 (2017) 304.
- [19] H. Schaefer, C. Heibl, S.S. Renner, Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (*Cucurbitaceae*) and numerous oversea dispersal events, Proc. Biol. Sci. 276 (2009) 843-851.
- [20] M. Hu, Z. Chen, B. Wang, Comparison on leaf ultrastructure in cytoplasmic male sterile line f or tuber mustard (*Brassica juncea* var. *tumida*) J. Zhejiang Agr. Univ. 27 (2001) 535-540.
- [21] J. Li, J. Li, Chloroplast thylakoid membrane peptides and cytoplasmic male sterility, J. Genet. genomics 13 (1986) 430-436.
- [22] J. Li, W. Gao, Studies on ultrastructure of chloroplast in cytoplasmic male sterile lines and their maintainer lines of rape, J. Genet. Genomics 10 (1983) 280-283.
- [23] Z. Chen, S. Muthukrishnan, G.H. Liang, K.F. Schertz, G.E. Hart, A chloroplast DNA deletion located in RNA polymerase gene *rpoC2* in CMS lines of sorghum, Mol. Gen. Genet. 236 (1993) 251-259.
- [24] Z. Li, Y. Cheng, J. Cui, P. Zhang, H. Zhao, S. Hu, Comparative transcriptome analysis reveals carbohydrate and lipid metabolism blocks in *Brassica napus* L. male sterility induced by the chemical hybridization agent monosulfuron ester sodium, BMC Genomics 16 (2015) 206.
- [25] B. He, Y. Ning, S. Liu, Y. Feng, A preliminary studies on the male sterile system of 1B/1R wheat varieties with *Ae.Kotschy* cytoplasm, Acta Univ. Septentrionali Occident Agric. 15 (1987) 107-109.
- [26] H. Gong, L. Ma, B. He, C. Fan, H. Wu, W. Li, Changes of RNase and protein content in the fertility sensitive period of wheat male sterile lines, J. Triticeae 28 (2008) 37-40.
- [27] Y. Yao, G. Zhang, H. Liu, J. Wang, Correlation between the inner wall of pollen grains of K-type wheat and ATPase activity and male sterility, Acta Bot. 22 (2002) 333-337.
- [28] F. Zhang, G. Li, Q. Ding, Z. Wang, X. Ma, H. Zhang, Z. Zhang, F. Jin, L. Ma, Proteome analysis of pollen in the K-type male sterility line 732A and its maintainer 732B in wheat (*Triticum aestivum* L.) by two-dimensional gel electrophoresis, Acta Physiol. Plant. 38 (2016) 84.

- [29] W. Yang, L. Xueyuan, L. Juan, P. Mingyu, M.A. A., L. Dongcheng, S. Jiazhu, Z. Kehui, H. Lixiong, Z. Aimin, Cloning and functional analysis of MADS-box genes, *TaAG-A* and *TaAG-B*, from a wheat K-type cytoplasmic male sterile line, *Front. Plant Sci.* 8 (2017) 1081-.
- [30] C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach, *Nucleic Acids Res.* 41 (2013) e129-e129.
- [31] B. Anton, N. Sergey, A. Dmitry, A. Gurevich, D. Mikhail, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455-477.
- [32] C. Liu, L. Shi, Y. Zhu, H. Chen, J. Zhang, X. Lin, X. Guan, CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences., *BMC Genomics* 13 (2012) 715-715.
- [33] Rédei, G. P., Encyclopedia of genetics, genomics, proteomics and informatics, New York: Springer 1 (2008) 1102.
- [34] H. Sakai, T. Washio, R. Saito, A. Shinagawa, M. Tomita, Correlation between sequence conservation of the 5' untranslated region and codon usage bias in *Mus musculus* genes., *Gene* 276 (2001) 101-105.
- [35] T. Thiel, W. Michalek, R. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* 106 (2003) 411-422.
- [36] S. Kurtz, C. Schleiermacher, REPuter: fast computation of maximal repeats in complete genomes, *Bioinformatics* 15 (1999) 426-427.
- [37] S.M. Chung, J.E. Staub, J.-F. Chen, Molecular phylogeny of Cucumis species as revealed by consensus chloroplast SSR marker length and sequence variation, *Genome* 49 (2006) 219-229.
- [38] L.D.N. Vieira, F. Helisson, R. Marcelo, F.H.P.d. Freitas, C.R.L. Alves, d.S.E. Maltempi, d.O.P. Fábio, N.R. Onofre, G.M. Pedro, M.U. Gertrud, The complete chloroplast genome sequence of *Podocarpus Lambertii*: genome structure, evolutionary aspects, gene content and SSR detection, *Plos One* 9 (2014) e90618.
- [39] Z. Cai, M. Guisinger, H.-G. Kim, E. Ruck, J.C. Blazier, V. McMurtry, J.V. Kuehl, J. Boore, R.K. Jansen, Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions, *J. Mol. Evol.* 67 (2008) 696-704.
- [40] T.A.G. Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* 408 (2000) 796-815.
- [41] T.I.W.G.S.C. (IWGSC), A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome, *Science* 345 (2014) 1251788.

- [42] M.A. Biscotti, A. Canapa, M. Forconi, E. Olmo, M. Barucca, Transcription of tandemly repetitive DNA: functional roles, *Chromosome Res.* 23 (2015) 463-477.
- [43] T. Cavalier Smith, Chloroplast evolution: secondary symbiogenesis and multiple losses, *Curr. Biol.* 12 (2002) R62-R64.
- [44] H. Shimda, M. Sugiuro, Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes, *Nucleic Acids Res.* 19 (1991) 983-995.
- [45] M.O. Kazuo Shinozaki, Minoru Tanaka, Tatsuya Wakasugi, Nobuaki Hayashida, Tohru Matsubayashi, N Zaita, J Chunwongse, Junichi Obokata, Kazuko Yamaguchishinozaki, C Ohto, Keita Torazawa, B Y Meng, Mamoru Sugita, Hiroshi Deno, T Kamogashira, Kyoji Yamada, Jun Kusuda, F Takaiwa, A Kato, N Tohdoh, Hiroaki Shimada, Masahiro Sugiura, The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression, *Plant Mol. Biol. Rep.* 5 (1986) 2043-2049.
- [46] Y. Zhao, J. Yin, H. Guo, Y. Zhang, W. Xiao, C. Sun, J. Wu, X. Qu, J. Yu, X. Wang, The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*, *Front. Plant Sci.* 5 (2015) 696.
- [47] C. Alkan, S. Sajjadian, E.E. Eichler, Limitations of next-generation genome sequence assembly, *Nat. Methods* 8 (2010) 61-65.
- [48] H. Huang, C. Shi, Y. Liu, S. Mao, L. Gao, Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships, *BMC Evol. Biol.* 14 (2014) 151.
- [49] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y. Chen, Z. Chen, Genome sequencing in microfabricated high-density picolitre reactors, *Nature* 437 (2005) 376-380.
- [50] L. Zhong, W. Zhou, H. Wang, S. Ding, Q. Lu, X. Wen, L. Peng, L. Zhang, C. Lu, Chloroplast small heat shock protein *HSP21* interacts with plastid nucleoid protein *pTAC5* and is essential for chloroplast development in *Arabidopsis* under heat stress, *Plant Cell* 25 (2013) 2925-2943.
- [51] H. Yu, X. Yang, S. Chen, Y. Wang, J. Li, Q. Shen, X. Liu, F. Guo, L. Qu, Downregulation of chloroplast *RPS1* negatively modulates nuclear heat-responsive expression of *HsfA2* and its target genes in *Arabidopsis*, *Plos Genet.* 8 (2012) e1002669.
- [52] M. Ueda, M. Fujimoto, S.-i. Arimura, J. Murata, N. Tsutsumi, K.-i. Kadowaki, Loss of the *rp/32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*, *Gene* 402 (2007) 0-56.
- [53] R.C. Haberle, H.M. Fourcade, J.L. Boore, R.K. Jansen, Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes, *J. Mol. Evol.* 66 (2008) 350-

- [54] M. Källersjö, J.S. Farris, M.W. Chase, B. Bremer, M.F. Fay, C.J. Humphries, G. Petersen, O. Seberg, K.r. Bremer, Simultaneous parsimony jackknife analysis of 2538 *rbcl* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants, *Plant Syst. Evol.* 213 (1998) 259-287.
- [55] A. Bruneau, F. Forest, P.S. Herendeen, B.B. Klitgaard, G.P. Lewis, Phylogenetic relationships in the caesalpinioideae (*leguminosae*) as inferred from chloroplast *trnL* intron sequences, *Syst. Bot.* 26 (2001) 487-514.
- [56] K.-J. Kim, H.-L. Lee, Complete chloroplast genome wequences from Korean Ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 Vascular plants, *DNA Res.* 11 (2004) 247-261.
- [57] L.X. Liu, L. Rui, W.J.R. P., L. Xian, L. Pan, C.K. M., F. Cheng-Xin, The complete chloroplast genome of chinese bayberry (*Morella rubra*, *Myricaceae*): implications for understanding the evolution of Fagales, *Front. Plant Sci.* 8 (2017) 968-968.
- [58] J. Wu, B. Liu, F. Cheng, N. Ramchiary, S.R. Choi, Y.P. Lim, X.-W. Wang, Sequencing of Chloroplast Genome Using Whole Cellular DNA and Solexa Sequencing Technology, *Front. Plant Sci.* 3 (2012) 243.
- [59] P.M. Hollingsworth, S.W. Graham, D.P. Little, D. Steinke, Choosing and using a plant DNA barcode, *Plos One* 6 (2011) e19254.
- [60] M.S. Lima, L.C. Woods, M.W. Cartwright, D.R. Smith, The (in)complete organelle genome: exploring the use and non-use of available technologies for characterizing mitochondrial and plastid chromosomes, *Mol. Ecol. Resour.* 16 (2016) 1279-1286.
- [61] J. Korlach, K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J.L. Wegener, S.W. Turner, Chapter 20 – Real-Time DNA sequencing from single polymerase molecules, *Science* 323 (2009) 133-138.
- [62] C.P. Middleton, N. Senerchia, N. Stein, E.D. Akhunov, B. Keller, T. Wicker, B. Kilian, Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe, *Plos One* 9 (2014) e85761.
- [63] I. Braslavsky, B. Hebert, E.P. Kartalov, S.R. Quake, Sequence information can be obtained from single DNA molecules, *P. Natl. Acad. Sci. USA.* 100 (2003) 3960-3964.
- [64] A. Salameh, M. Buerstmayr, B. Steiner, A. Neumayer, M. Lemmens, H. Buerstmayr, Effects of introgression of two QTL for fusarium head blight resistance from Asian spring wheat by marker-assisted backcrossing into European winter wheat on fusarium head blight resistance, yield and quality traits, *Mol. Breeding* 28 (2011) 485-494.

- [65] D. Shim, S. Raveendar, J. Lee, G. Lee, N. Ro, Y. Jeon, G. Cho, H. Lee, K. Ma, J. Chung, The complete chloroplast genome of *Capsicum Frutescens* (*Solanaceae*), *Appl. Plant Sci.* 4 (2016) 1600002.
- [66] Q. Xing, Z. Ru, J. Li, C. Zhou, D. Jin, Y. Sun, B. Wang, Cloning a second form of adenine phosphoribosyl transferase gene (*TaAPT2*) from wheat and analysis of its association with thermo-sensitive genic male sterility (TGMS), *Plant Sci.* 169 (2005) 37-45.
- [67] C. Xia, L. Zhang, C. Zou, Y. Gu, J. Duan, G. Zhao, J. Wu, Y. Liu, X. Fang, L. Gao, A TRIM insertion in the promoter of *Ms2* causes male sterility in wheat, *Nat. Commun.* 8 (2017) 15407.
- [68] K. Chen, V.G. Meyer, Mutation in chloroplast DNA coding for the large subunit of fraction 1 protein correlated with male sterility in cotton, *J. Hered.* 70 (1979) 431-433.
- [69] V. Kohn, Christopher, Kielkowska, Agnieszka, Havey, Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasms of onion, *Genome* 56 (2013) 737-742.
- [70] J. Jia, L.G. Zhang, mRNA differential display and EST sequence analysis of aborted bud and normal bud in radish (*Raphanus sativus*), *J. Nucl. Agr. Sci.* 22 (2008) 426-431.
- [71] L.-J. OU, G.-W. Huang, W.-J. LI, G.P. Kang, J.-L. Chen, S.Luan, L.-B. Chen, Chloroplast DNA polymorphism in different types of cytoplasmic male sterile rice, *Biol. Plantarum* 53 (2009) 593-596.
- [72] M.C. Little, R.B. Hallick, Chloroplast *rpoA*, *rpoB*, and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription, *J. Biol. Chem.* 263 (1988) 14302-14307.
- [73] Z. Chen, K.F. Schertz, J.E. Mullet, A. Dubell, G.E. Hart, Characterization and expression of *rpoC2* in CMS and fertile lines of sorghum, *Plant Mol. Biol.* 28 (1995) 799-809.
- [74] L. Ma, Y. Hu, X. Song, B. He, The comparative investigation of 1B/1R and non_1B/1R K Type wheat male sterile lines, *Acta Agriculturae Boreali-Sinica* 21 (2006) 5-8.
- [75] O. Kenji, Y. Katsuyuki, O. Eiji, N. Yasukazu, T. Miho, N. Naoko, A. Kinya, O. Kanji, Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: the absence of chloroplast-like tRNAs, *Nucleic Acids Res.* 20 (1992) 3773-3777.

Tables

Table 1 Gene contents of K519A and 519B chloroplast genomes

Category for genes	Group of genes	Gene names
Genes for photosynthesis	ATP synthase	atpA▯atpB▯atpE▯atpF*▯atpH▯atpI
	Cytochrome b/f complex	petA▯petB*▯petD*▯petG▯petL▯petN
	NADH dehydrogenase	ndhA*▯ndhB*▯ndhC▯ndhD▯ndhE▯ndhF▯ndhG▯ndhH▯ndhI▯ndhJ▯ndhK
	Photosystem I	psaA▯psaB▯psaC▯psaI▯psaJ
	Photosystem II	psbA▯psbB▯psbC▯psbD▯psbE▯psbF▯psbH▯psbI▯psbJ▯psbK▯psbL▯psbM▯psbN▯psbT▯psbZ
	Large subunit of rubisco	rbcL
	ATP-dependent protease subunits P gene	clpP**
	Proteins of small ribosomal (SSU)	rps2▯rps3▯rps4▯rps7▯rps8▯rps11▯rps12#▯rps14▯rps15▯rps16▯rps18▯rps19
	Proteins of large ribosomal (LSU)	rpl2*▯rpl14▯rpl16▯rpl20▯rpl22▯rpl23▯rpl32▯rpl33▯rpl36
Expression related gene	Ribosomal RNAs	rrn4.5▯rrn5▯rrn16▯rrn23
	RNA polymerase	rpoA▯rpoB▯rpoC1▯rpoC2
	Transfer RNAs	trnA-UGC*▯trnC-GCA▯trnD-GUC▯trnE-UUC▯trnF-GAA▯trnM-CAU▯trnG-GCC▯trnG-

UCC*trnH-GUGtrnI-CAUtrnI-GAU*trnK-UUU*trnL-CAAtrnL-UAA*trnL-UAGtrnM-CAUtrnN-GUUtrnP-UGGtrnQ-UUGtrnR-UCUtrnR-ACGtrnS-GCUtrnS-GGAtrnS-UGAtrnT-GGUtrnT-UGUtrnV-GACtrnV-UAC*trnW-CCAtrnY-GUA

Other genes	Maturase	matK
	Envelope membrane protein	cemA
	C-type cytochrome synthesis gene	ccsA
	Translation initiation factor	infA
	Conserved	ycf1ycf2ycf3**ycf4
	unknown open reading frame	

* Genes containing an intron. ** Genes containing two introns. # Trans splicing genes.

Table 2 Codon preference analysis of K519A chloroplast genes

Amino acid name	Total number of amino acids	Total number of amino acids
Ala(A)	832	GCU (37.26%), GCC (21.63%), GCA (25.96%), GCG (15.14%)
Arg(R)	1250	CGU (16.88%), CGC (10.08%), CGA (14.16%), CGG (7.84%), AGA (31.68%), AGG (19.36%)
Asn(N)	1043	AAU (70.85%), AAC (29.15%)
Asp(D)	595	GAU (74.29%), GAC (25.71%)
Cys(C)	468	UGU (53.21%), UGC (46.79%)
Gln(Q)	603	CAA (77.78%), CAG (22.22%)
Glu(E)	862	GAA (73.20%), GAG (26.80%)
Gly(G)	1166	GGU (29.07%), GGC (14.84%), GGA (35.59%), GGG (20.50%)
His(H)	562	CAU (78.83%), CAC (21.17%)
Ile(I)	1697	AUU (42.78%), AUC (25.34%), AUA (31.88%)
Leu(L)	1651	UUA (22.71%), UUG (26.83%), CUU (21.93%), CUC (10.60%), CUA (10.30%), CUG (7.63%)
Lys(K)	1148	AAA (66.90%), AAG (33.10%)
Met(M)	378	AUG (100.00%)
Phe(F)	1259	UUU (57.11%), UUC (42.89%)
Pro(p)	838	CCU (28.28%), CCC (24.70%), CCA (36.28%), CCG (10.74%)
Ser(S)	1866	UCU (21.01%), UCC (23.26%), UCA (6.91%), UCG (10.77%), AGU (19.13%), AGC (18.92%)
Thr(T)	1178	ACU (34.38%), ACC (28.61%), ACA (19.44%), ACG (17.57%)
Trp(W)	271	UGG (100.00%)
Tyr(Y)	851	UAU (64.98%), UAC (35.02%)
Val(V)	942	GUU (35.35%), GUC (15.61%), GUA (33.01%), GUG (16.03%)
Stop_Codon	709	UAG (22.99%), UGA (18.62%), UAA (58.39%)

Table 3 Codon preference analysis of 519B chloroplast genes

Amino acid name	Total number of amino acids	Total number of amino acids
Ala(A)	837	GCU (37.40%), GCC (21.74%), GCA (26.16%), GCG (14.70%)
Arg(R)	1250	CGU (16.80%), CGC (10.16%), CGA (14.56%), CGG (8.08%), AGA (31.04%), AGG (19.36%)
Asn(N)	1033	AAU (70.86%), AAC (29.14%)
Asp(D)	595	GAU (74.29%), GAC (25.71%)
Cys(C)	471	UGU (53.08%), UGC (46.92%)
Gln(Q)	601	CAA (77.54%), CAG (22.46%)
Glu(E)	861	GAA (73.17%), GAG (26.83%)
Gly(G)	1158	GGU (28.67%), GGC (14.94%), GGA (35.84%), GGG (20.55%)
His(H)	560	CAU (78.57%), CAC (21.43%)
Ile(I)	1692	AUU (42.43%), AUC (25.24%), AUA (32.33%)
Leu(L)	1675	UUA (23.28%), UUG (26.63%), CUU (21.61%) CUC (10.69%), CUA (10.15%), CUG (7.64%)
Lys(K)	1153	AAA (67.13%), AAG (32.87%)
Met(M)	380	AUG (100.00%)
Phe(F)	1261	UUU (56.78%), UUC (43.22%)
Pro(p)	848	CCU (28.42%), CCC (24.17%), CCA (37.62%), CCG (9.79%)
Ser(S)	1860	UCU (21.08%), UCC (23.33%), UCA (7.10%), UCG (10.54%), AGU (19.25%), AGC (18.71%)
Thr(T)	1177	ACU (34.41%), ACC (28.21%), ACA (19.46%), ACG (17.93%)
Trp(W)	268	UGG (100.00%)
Tyr(Y)	843	UAU (65.36%), UAC (34.64%)
Val(V)	948	GUU (35.55%), GUC (15.30%), GUA (33.44%), GUG (15.72%)
Stop_Codon	710	UAG (23.80%), UGA (18.73%), UAA (57.46%)

Table 4 Characterization of simple sequence repeats discovered in the K519A chloroplast genome

Microsatellite sequences		Number of repeats																							Total
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23			
A		-	-	-	-	-	37	10	5	3	2	3	-	-	1	-	-	-	-	-	-	1	62		
C		-	-	-	-	-	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4		
G		-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3		
T		-	-	-	-	-	25	18	9	5	-	1	-	-	-	-	-	-	-	-	-	-	58		
AT		-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3		
TA		-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3		
TC		-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2		
AAC		7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7		
AAG		3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3		
AAT		1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2		
AGA		5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5		
AGC		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
AGT		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
ATA		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
CTT		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
GAA		2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2		
GAT		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
GCA		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
GTT		2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2		
TAA		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
TAT		2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3		
TCT		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
TGC		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		
TTC		7	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8		
TTG		1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1		

AACG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAGA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AGAA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TCCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TCGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TTCA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TTCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
CCATA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Total																					186

Table 5 Characterization of simple sequence repeats discovered in the 519B chloroplast genome

Microsatellite sequences		Number of repeats																Total
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
	A	-	-	-	-	-	40	13	3	3	3	-	-	-	-	1	1	64
	C	-	-	-	-	-	3	-	1	-	-	-	-	-	-	-	-	4
	G	-	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	2
	T	-	-	-	-	-	27	13	11	2	2	1	-	-	-	-	-	56
	AT	-	-	3	1	-	-	-	-	-	-	-	-	-	-	-	-	4
	TA	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	3
	TC	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	AAC	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
	AAG	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
	AAT	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	AGA	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
	AGC	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	AGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	ATA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	ATT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	CTT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	GAA	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	GAT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	GCA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	GTT	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
	TAA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	TAT	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
	TCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	TGC	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
	TTC	6	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7

TTG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AACG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AAGA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AATA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
AGAA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TCCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TCGT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TTCA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TTCT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
ATAGA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
CCATA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
TTTAT	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Total																	188

Table 6 Primers for the simple sequence repeats analysis.

Genes	Forward primers	Reverse primers
atpF1	TGGAAGAACCATAGCATTTCGC	ACATAGGTCGTCGATTTCGGCAT
atpF2	TATCTAGGCGGAAGAGTCCTCT	AGTTGCTCAATGGTCCCTCTAC
rps16	TTTCTCCTCATACGGCTCGA	TCCCGAAGTAATGTCTAAACCC
trnI-GAU	TTGGTCTGTGAAGATGCGTTGT	GCTCTCAGCCACATGGATAGTT
trnK-UUU	AGTGCAATATGGTCAGAACAGG	AATAGATTAGTGCCGGATGCG

Table 7 Features and positions of long repeat fragments in the K519A genes

Repeat length	Starting position	Mapped gene	Match direction	Repeat length	Starting position	Mapped gene	Distance	Calculated e-value of this repeat
286	56620		F	286	134618	rpl2-rpl23	0	3.41E-163
77	0		F	77	136919	rps19	0	2.31E-37
40	66331	rps18	F	40	66373	rps18	-2	3.07E-11
38	38456	psaB	F	38	40680	psaA	-2	4.42E-10
30	87709		F	30	130393	ndhB	0	4.58E-09
30	130393	ndhB	P	30	130393	ndhB	0	4.58E-09
36	43404	ycf3	F	36	90709		-2	6.34E-09
36	43404	ycf3	P	36	127387		-2	6.34E-09
33	66338	rps18	F	33	66380	rps18	-1	7.08E-09
29	7614	trnS-GCU	P	29	44922	trnS-GGA	0	1.83E-08
35	76836	infA	F	35	76854	infA	-2	2.39E-08
35	76845	infA	F	35	76863	infA	-2	2.39E-08
37	66351	rps18	F	37	66393	rps18	-3	5.86E-08
33	27173	rpoC2	F	33	27224	rpoC2	-2	3.40E-07
29	27177	rpoC2	F	29	27228	rpoC2	-1	1.59E-06
34	27251	rpoC2	F	34	27272	rpoC2	-3	2.89E-06
34	66373	rps18	F	34	66394	rps18	-3	2.89E-06
25	80343	rps3	F	25	80388	rps3	0	4.69E-06
31	27376	rpoC2	F	31	27442	rpoC2	-2	4.79E-06
31	66389	rps18	F	31	66410	rps18	-2	4.79E-06
33	27256	rpoC2	F	33	27277	rpoC2	-3	1.05E-05
30	12940	trnG-UCC	F	30	36397		-2	1.79E-05
30	16817	psbM	P	30	16817	psbM	-2	1.79E-05
30	66787	rps18	P	30	66787	rps18	-2	1.79E-05
32	27132	rpoC2	F	32	27231	rpoC2	-3	3.83E-05
32	27309	rpoC2	F	32	27384	rpoC2	-3	3.83E-05

32	27313	rpoC2	F	32	27388	rpoC2	-3	3.83E-05
29	11348	trnS-UGA	P	29	44925	trnS-GGA	-2	6.69E-05
26	27138	rpoC2	F	26	27237	rpoC2	-1	9.14E-05
31	27114	rpoC2	F	31	27384	rpoC2	-3	1.39E-04
31	27301	rpoC2	F	31	27442	rpoC2	-3	1.39E-04
28	36250	trnR-UCU	R	28	36250	trnR-UCU	-2	2.49E-04
22	16821	psbM	P	22	16821	psbM	0	3.00E-04
25	7615	trnS-GCU	F	25	11352	trnS-UGA	-1	3.52E-04
25	11352	trnS-UGA	P	25	44925	trnS-GGA	-1	3.52E-04
30	66320	rps18	F	30	66425	rps18	-3	5.02E-04
30	66408	rps18	F	30	66429	rps18	-3	5.02E-04
27	27160	rpoC2	F	27	27355	rpoC2	-2	9.26E-04
27	66373	rps18	F	27	66415	rps18	-2	9.26E-04
21	7619	trnS-GCU	F	21	11356	trnS-UGA	0	1.20E-03
21	11356	trnS-UGA	P	21	44925	trnS-GGA	0	1.20E-03
21	14855	trnT-GGU	P	21	46177	trnT-UGU	0	1.20E-03
21	27124	rpoC2	F	21	27319	rpoC2	0	1.20E-03
21	113039	ndhA	R	21	113039	ndhA	0	1.20E-03
29	27092	rpoC2	F	29	27146	rpoC2	-3	1.81E-03
29	38447	psaB	F	29	40671	psaA	-3	1.81E-03
26	27319	rpoC2	F	26	27394	rpoC2	-2	3.43E-03
26	77132	infA	F	26	77150	infA	-2	3.43E-03
26	107460	ccsA	P	26	107460	ccsA	-2	3.43E-03
20	7684	trnS-GCU	F	20	11421	trnS-UGA	0	4.80E-03

F forward. **R** reverse. **P** palindromic. **C** complementary.

Table 8 Features and positions of long repeat fragments in the 519B genes

Repeat length	Starting position	Mapped gene	Match direction	Repeat length	Starting position	Mapped gene	Distance	Calculated e-value of this repeat
286	56605		F	286	133861	rpl2-rpl23	-1	2.90E-160
263	56628		F	263	133884	rpl2-rpl23	0	2.38E-149
77	0		F	77	136158	rps19	0	2.29E-37
40	65528	rps18	F	40	65570	rps18	-2	3.03E-11
38	38432	psaB	F	38	40656	psaA	-2	4.37E-10
30	86907		F	30	129632	ndhB	0	4.53E-09
30	129632	ndhB	P	30	129632	ndhB	0	4.53E-09
36	43403	ycf3	F	36	89907		-2	6.27E-09
36	43403	ycf3	P	36	126626		-2	6.27E-09
33	65535	rps18	F	33	65577	rps18	-1	7.00E-09
29	7613	trnS-GCU	P	29	44931	trnS-GGA	0	1.81E-08
35	76031	infA	F	35	76049	infA	-2	2.37E-08
35	76040	infA	F	35	76058	infA	-2	2.37E-08
37	65548	rps18	F	37	65590	rps18	-3	5.80E-08
33	27181	rpoC2	F	33	27232	rpoC2	-2	3.36E-07
29	27185	rpoC2	F	29	27236	rpoC2	-1	1.58E-06
34	27259	rpoC2	F	34	27280	rpoC2	-3	2.86E-06
34	27321	rpoC2	F	34	27396	rpoC2	-3	2.86E-06
34	65570	rps18	F	34	65591	rps18	-3	2.86E-06
25	79542	rps3	F	25	79587	rps3	0	4.64E-06
31	27384	rpoC2	F	31	27450	rpoC2	-2	4.74E-06
31	65586	rps18	F	31	65607	rps18	-2	4.74E-06
33	27264	rpoC2	F	33	27285	rpoC2	-3	1.04E-05
30	12893	trnG-UCC	F	30	36373		-2	1.77E-05
30	16815	psbM	P	30	16815	psbM	-2	1.77E-05
30	65984	rps18	P	30	65984	rps18	-2	1.77E-05

32	27140	rpoC2	F	32	27239	rpoC2	-3	3.79E-05
32	27287	rpoC2	F	32	27341	rpoC2	-3	3.79E-05
32	27317	rpoC2	F	32	27392	rpoC2	-3	3.79E-05
29	11303	trnS-UGA	P	29	44934	trnS-GGA	-2	6.62E-05
26	27146	rpoC2	F	26	27245	rpoC2	-1	9.04E-05
31	27309	rpoC2	F	31	27450	rpoC2	-3	1.37E-04
28	27327	rpoC2	F	28	27402	rpoC2	-2	2.46E-04
28	36225	trnR-UCU	R	28	36225	trnR-UCU	-2	2.46E-04
22	16819	psbM	P	22	16819	psbM	0	2.97E-04
22	78893	rpl16	R	22	78893	rpl16	0	2.97E-04
25	7614	trnS-GCU	F	25	11307	trnS-UGA	-1	3.48E-04
25	11307	trnS-UGA	P	25	44934	trnS-GGA	-1	3.48E-04
30	65517	rps18	F	30	65622	rps18	-3	4.96E-04
30	65605	rps18	F	30	65626	rps18	-3	4.96E-04
27	27168	rpoC2	F	27	27363	rpoC2	-2	9.15E-04
27	65570	rps18	F	27	65612	rps18	-2	9.15E-04
21	7618	trnS-GCU	F	21	11311	trnS-UGA	0	1.19E-03
21	11311	trnS-UGA	P	21	44934	trnS-GGA	0	1.19E-03
21	14864	trnT-GGU	P	21	46185	trnT-UGU	0	1.19E-03
21	27132	rpoC2	F	21	27327	rpoC2	0	1.19E-03
21	112256	ndhA	R	21	112256	ndhA	0	1.19E-03
24	27259	rpoC2	F	24	27334	rpoC2	-1	1.34E-03
29	27100	rpoC2	F	29	27154	rpoC2	-3	1.79E-03
29	38423	psaB	F	29	40647	psaA	-3	1.79E-03
26	76327	infA	F	26	76345	infA	-2	3.39E-03
20	7683	trnS-GCU	F	20	11376	trnS-UGA	0	4.75E-03

F forward. **R** reverse. **P** palindromic. **C** complementary.

Table 9 Alignment analysis of coding genes of K519A and 519B

Genes	K519A -519B	K519A -519B	Mutation type	SNP type
	nucleic acid	amino acid		
matK	75T-A	489Ser-Cys	N	1
	932 T-A	203Tyr-Phe	N	1
	976A-G		S	2
	1003C-T		S	2
	1117T-A	141Leu-Phe	N	1
	1440G-C	34His-Asp	N	1
	1486A-G		S	2
rps16	12 C-T	60Ala-Thr	N	2
	13 T-G	59Arg-Ser	N	1
	70 T-C		S	2
psbD	738 C-A		S	1
psbC	147G-T		S	1
psbZ	126C-T		S	2
rpoB	438 C-G		S	1
	459 G-A		S	2
	1245 C-A		S	1
	1366 A-G	456Lys-Glu	N	2
	1899 C-T		S	2
	1908 G-A		S	2
	2628 G-A		S	2
	2937 C-G		S	1
rpoC1	712A-C		S	1
	783T-A		S	1
	1036C-A		S	1

	1683C-T		S	2
rpoC2	1002 G-A		S	2
	1695 A-T		S	1
	2083T-C	695Ser-Pro	N	2
	2125C-A	709Leu-Ile	N	1
	2223 C-T		S	2
	2780T-C	927Phe-Ser	N	2
	2984G-A	995Gly-Asp	N	2
	3564T-C		S	2
	3789G-A		S	2
	4074 A-G		S	2
	4334A-T	1445Lys-Ile	N	1
	4413 T-C		S	2
rps2	552G-A		S	2
atpI	225 T-C		S	2
	477C-A	159Ser-Arg	N	1
	555G-A		S	2
atpF	741A-G		S	2
atpA	792T-C		S	2
	1164G-T		S	1
rps14	37C-T		S	2
psaB	562T-C		S	2
	1051G-C		S	1
	1266T-G		S	1
	1750T-C		S	2
ycf3	1008C-T		S	2
rps4	210 G-T		S	1
ndhJ	427C-T		S	2
ndhK	195 T-C	162Ile-Val	N	2

	583G-A		S	2
ndhC	88A-G		S	2
	280 A-T		S	1
atpE	361T-C		S	2
atpB	9 T-G	497Lys-Gln	N	1
	283G-A		S	2
	847T-C		S	2
	1177G-A		S	2
	1334G-T	55Ala-Asp	N	1
	1442C-T	19Ser-Asn	N	2
	1463 A-G	12Val-Ala	N	2
rbcL	40C-A	14Gln-Lys	N	1
	285A-G	95Asn-Ser	N	2
	1429T-A		S	1
	1431A-G	477*-Lys	N	2
cemA	642T-C		S	2
petA	606C-T		S	2
	666G-A		S	2
psbJ	73A-G		S	2
rps18	163C-T		S	2
clpP	343C-T		S	2
	442T-C		S	2
psbB	1188T-C		S	2
	1523T-C	508Val-Ala	N	2
psbH	133A-G	45Ile-Val	N	2
petB	1002A-C		S	1
petD	1046T-C		S	2
	1109C-T		S	2
rpoA	396G-A		S	2

	802T-C		S	2
rpl14	47T-C	109Glu-Gly	N	2
rpl16	175C-T		S	2
rps3	124A-G		S	2
	631A-C		S	1
rpl2	1008A-C		S	1
ndhB	751G-A		S	2
ndhH	390T-C	265Ile-Val	N	2
	664A-C	173Asp-Glu	N	1
ndhF	94C-T		S	2
	169G-T		S	1
	232G-A		S	2
	457G-A		S	2
	1453C-T		S	2
	1560T-C	221Ile-Val	N	2
rpl32	172G-A	58Asp-Asn	N	2
	186G-T	62Leu-Phe	N	1
ccsA	78T-G	26Phe-Leu	N	1
	553T-C	185Phe-Leu	N	2
ndhD	1264C-T		S	2
ndhE	43T-C		S	2
ndhI	312A-G		S	2
ndhA	646T-C		S	2
	988G-A		S	2
	1046G-C	15Ser-Trp	N	1

* Stop codon. **S** Synonymous. **N** Non-synonymous. **1** transversion. **2** transition.

Table 10 Primers for the first-generation sequencing

Genes	Forward primers	Reverse primers
matK	TCAACCCTTTCCTGTTTCTT	CGTCAACAATACTTTGTCTACC
rps16	GCTTATGTTGGATTGGCACGAT	CCCGAAGTAATGTCTAAACCCA
rpoB	CGGATTGGCTCTTGGTCGTT	CGATTTCGCATCATTGTGCTCAA
rpoC2	GGAATCCTAGAAGACGAATACG	GACCTGTTAGTGTTCTAAGTTC
atpI	TCCCTGGGTTCCCTTTATTGG	AGAGCTTGAATACCGCTTGTA
ndhK	GCAATTCTTGCTCATAAGTTCC	GGAATGTTTCAGTACGGATTTC
atpB	GGACCCATATCTACTGCTGTGT	GGGCAACGAAATCAAGTCCTG
rbcL	CATACACAGGGTGTACGCATTA	AGATTGAGCCGAGTTTAATTGC
psbB	TGGTGGCGAACTTAATGGAGTA	TTGGAGTTGGTGGAAACCTTAGG
psbH	GGCGACTAAAGTTGCTGTTTCC	CAGTCTGGTTGCGAGGTCTTAA
rpl14	GGGTCGGCTGATACGCTTTA	ATGATGACAATGCTGCGGTTAT
ndhH	GCCAGCATATTCAGGACCAAT	AGAGCGAGTTGAAGGAGTAGG
ndhF	AGAGGAAGAAGTCGAGCTAGAA	TGGTCTTGGACCGTCAATAATG
Rpl32	GCAGCAATAGATGTCTTTCACA	TGGAGAAGATAGGTGGAAGTTG
ccsA	CGAGTGGCGGCATTCTTGA	TGGAAATGAAGGGACAGAGGTT
ndhA	GCTTCCGAATTGATCTCATCCT	AGTTAGTGAAGGGTTAGGAACA

Table 11 Primers for the qRT-PCR.

Gene	Forward primer	Reverse primer
atpB	TTGATAACCCACTGCGGAGG	CTGCCCTAACTATGGCGGAA
ccsA	CGGAGCGTTTGGATTCTTGG	GCCTCATTAGCCCATACTGC
matK	AGCGCATGAAAGTCGAAGTA	TCTTGATCGATTGTTGGTCGGA
ndhH	GTAGATCGGCGGCTACTCCT	TCGGCGCACAGACTCCTTT
rbcL	TGAATGCGACTGCGGGTACA	AGGCCATTGTCGCGGCAATA
rpoC2	GGCTGGTGCTTCCCTTGTTG	ATCGGTCCTGCACTTGTCCT

Figures

Figure 1

The structure observation of ear and anther of K519A and 519B. A. K519A. B. 519B.

Figure 2

Annotated map of the chloroplast genome structure in K519A. The transcripts of inside genes are the clockwise direction, while the transcripts of outside genes are in the opposite direction. Different functional genes are identified with different colors. The gray histogram shows genomic GC content, and the middle gray line is the 50% threshold line.

Figure 3

Annotated map of the chloroplast genome structure in 519B. The transcripts of inside genes are the clockwise direction, while the transcripts of outside genes are in the opposite direction. Different functional genes are identified with different colors. The gray histogram shows genomic GC content, and the middle gray line is the 50% threshold line.

Figure 4

Polymorphism SSR markers detection based on polyacrylamide gel electrophoresis. A. K519A. B. 519B.

Figure 5

Verification of gene sequences of non-synonymous mutate sites. SGS. second-generation sequencing. FGS. first-generation sequencing.

Figure 6

Relative expression level of selected non-synonymous mutant genes in K519A compared with 519B. The abscissa represents the development period. The ordinate represents the expression level of related genes.

Figure 7

Phylogeny of chloroplast genomes of proximal species.