

Identification of Nucleus Sterility Candidate Genes Using a Resequencing Technique in Sweet Pepper Sterile Line AB91

Yaning Meng*, Libin Yan*, Hongxiao Zhang, Yanqin Fan#

The Institute of Cash Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China

Email: *nkyfanyq@126.com

How to cite this paper: Meng, Y.N., Yan, L.B., Zhang, H.X. and Fan, Y.Q. (2021) Identification of Nucleus Sterility Candidate Genes Using a Resequencing Technique in Sweet Pepper Sterile Line AB91. *Agricultural Sciences*, **12**, 138-157.

<https://doi.org/10.4236/as.2021.122010>

Received: December 29, 2020

Accepted: February 23, 2021

Published: February 26, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Breeding hybrids with nuclear malesterile lines is an important method for the cross-breeding of sweet peppers. To date, few reports have been published on the nuclear malesterility gene of sweet pepper. Yet, there are approximately 20 pepper nuclear malesterility lines in the world. Using the self-developed testing material, sweet pepper nuclear malesterile dual-purpose line AB91, the genome-wide resequencing technique was applied to find that the mutation site causing the abortion of sweet pepper nuclear malesterility AB91 is on chromosome #5. The mutation gene Capana05g000747 was filtered out and validated by the flight mass spectrometry genotyping and quantitative real-time PCR method and determined to be the gene causing the abortion of sweet pepper nuclear male sterility AB91. The gene Capana05g000747 mutation site is a non-synonymous mutation site located at the 6th exon, the base C mutated into A, and the amino acid changed from alanine to serine. The three-dimensional protein structure of fertile and sterile plant Capana05g000747 was predicted. The results showed that the three-dimensional structure of the two proteins differed significantly. Sequence alignment analysis showed that the gene Capana05g000747 has a similar function to gene At2g02148. The gene At2g02148 contains a pentatricopeptide repeat protein which has important physiological functions in the gene expression process of organelles and is closely related to the performance of malesterility genes. Therefore, Capana05g000747 was selected as an important candidate gene for sweet pepper nuclear male sterile testing material AB91.

Keywords

Sweetpepper, Nuclear Malesterile Line, Genome-Wide Resequencing

*Contributed equally to this work.

1. Introduction

Sweet pepper (*Capsicum annuum*var. *grossum*) is a one-year or perennial crop in the Solanaceae and a variant of pepper. The heterosis of sweet pepper has high utilization value and using male infertility to breed hybrids is an important approach to solving the problem of artificial emasculation. Therefore, the research topic of male infertility has drawn a great deal of attention of domestic and international research scholars. Due to the lack of recovery genes for male sterility in the sweet pepper nucleus and the difficulty of finding excellent recovery lines, the breeding and application of cytoplasmic male sterility in sweet pepper hybrids are limited. However, the genetics of sweet pepper nucleus male sterility is relatively simpler and the number of recovery sources is relatively large, which are a great advantage inbreeding sweet pepper hybrids. Martin and Crawford [1] first reported capsicum sterility materials controlled by a recessive single-gene. Since then, new genes have been discovered in the pepper male sterile material. To date, nearly 20 nuclear sterility genes have been discovered in peppers [2] [3] [4]. However, there are very few reports on the sweet pepper nuclear male sterility genes.

Fan *et al.* [5] discovered a natural sterile sweet pepper plant. After years of research, improvement, and breeding, the sweet pepper nuclear male sterile dual-purpose line AB91 was bred, and its sterility feature was controlled by recessive nuclear genes named msc2 [2]. The sterile line AB91 is completely aborted and stable in sterility with no adverse cytoplasmic effect and has excellent agronomic traits and wide recovery sources. Usually, inbred lines can easily become their own recovery lines, the degree of freedom for association is large, and they are more likely to obtain advantageous combinations. So far, 14 sweet pepper hybrids have been bred using AB91 and have been widely planted [6]. We have found that the abortion of sterile plants occurs after the development of quarter microspores at the cytological level; the reasons for microspores abortion are disintegration of the tetraploid enamel wall, tapetum cell dysplasia, and delayed disintegration [7]; however, there are no reports of co-segregation markers or candidate genes for this gene site.

In recent years, with rapid development of modern molecular biology technology in the field of plants, male infertility research has risen to the level of genetic engineering. At present, a lot of genes are connected to the sterility traits for the pepper sterile lines. Ma *et al.* [8] and Li *et al.* [9] discovered the gene PAP3 in peppers, which was found to cause atrophy of pollen grains leading to male sterility. Two functional genes of the *MADS-box* gene family are APETALA3 (AP3)/DEFICIENS (DEF) and GLOBOSA (GLO)/PISTILLATA (PI), which are transcription indexes controlling petals and stamens [10] [11]. Fan *et al.* [12]

found a new gene CaCTS from the pepper male sterile line by cloning, which is highly expressed in flowers and seeds, moderately expressed in the placenta and pericarp, and weakly expressed in stems and leaves. Deng *et al.* [13] expanded out the malate dehydrogenase (MDH) gene by reverse transcription PCR. For the MDH gene, the performance level is low in abortion and the performance trend is not stable, which may interfere with the energy metabolism balance of the sterile line. Guo *et al.* [14] discovered the protein CaAMS, the down-regulation of which causes partial filament shortening, withering, stamen non-cracking, and pollen abortion. CaAMS plays an important role in the development of pepper tapetum and pollen through a complex regulatory network. Deng *et al.* [15] expanded the complete coding sequence for the triose phosphate isomerase (TPI) gene by reverse transcriptase PCR. In the abortion process, the activity and performance level of TPI in the anthers of male sterile lines were significantly decreased, and the levels of TPI in F1 hybrids and maintainer lines remained normal, which indicated that the stable TPI transcripts maintained the energy metabolism at a normal level. Qing *et al.* [16] performed genomic resequencing and comparison analysis for the male sterile line and fertile line of capsicum and selected Capana02g002096 as a candidate gene for the msc1 site according to genetic variation and annotation. Capana02g002096 encodes a homologous gene of AtDYT1, which is a bHLH transcription index involved in the early development of the tapetum. In addition, a deletion of 7 bp was found in the exon of Capana02g002096, which leads the codon to terminate prematurely, resulting in infertility due to a function loss. These prior genetic studies related to infertility have provided important insights for the further study of the mechanism of male infertility, and they also provided a theoretical foundation for related research.

Based on the studies mentioned above, our study used sweet pepper nuclear male sterile line AB91 as the testing material. According to the pepper genome information published [17], the difference sites and differential candidate genes of fertile and sterile plants for sweet pepper nuclear infertility AB91 were obtained through whole genome resequencing technology, and the genome resequencing results were analyzed by bioinformatics methods. The candidate gene sites were filtered out and validated by a mass spectrometry genotyping, quantitative real-time PCR and three-dimensional protein structure prediction method, and the candidate gene of male sterility for AB91 was finally identified, which provides a theoretical and technical foundation for further cloning as well as transformation and utilization of the gene.

2. Methods

2.1. Plant Materials

The sweet pepper recessive male sterile dual-purpose line AB91 was provided by the Sweet Pepper Research Group from the Institute of Cash Crops of Hebei Academy of Agriculture and Forestry Sciences. Fifty percent of the sweet pepper

recessive male sterile dual-purpose line AB91 groups are fertile plants and the remainder are sterile plants. The sterility trait, which is controlled by recessive nuclear genes, is genetically stable.

Basic group construction: F2 generation isolated groups containing recessive sterile plants (msms) and fertile plants (Msms, MsMs) were obtained from the self-crossing of fertile plants (Msms) in the sweet pepper male sterile line AB91. In order to determine the genotype of the fertile plants, the fertile plants in the F2 plants were self-crossed, and the homozygous dominant fertile plants (MsMs) and hybrid plants (Msms) were separated and identified according to offspring fertility to complete the basic group construction.

2.2. Gene Bank Construction and Illumina Sequencing

The homozygous fertile plants (MsMs) and recessive sterile plants (msms) in the F2 generation isolated group of sweet pepper AB91 were used as testing materials, and the DNA of the young leaves of plants was extracted by the CTAB method [18] to construct the MsMs and Msms gene banks. The DNA sample was fragmented into a size of 350 bp by sonication, then the DNA fragments were finally polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR expansion. At last, the PCR products were purified (AMPure XP system). Then, the size distribution of gene banks was analyzed by an Agilent2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Quantified analysis was performed using real-time PCR. The genebanks constructed above were sequenced by an Illumina HiSeq4000 platform (Illumina, San Diego, CA, USA) and 150 bp paired-end reads were generated with an insert size of approximately 350 bp.

2.3. Data Analysis

The raw reads from the msms and MsMs pools were filtered and aligned to the pepper sequence genomics (<http://peppersequence.genomics.cn/page/species/download.jsp>) using the Burrows Wheeler BWA alignment tool [19]. GATK software was used to detect single-nucleotide polymorphisms (SNPs) and insertion-deletion (InDels) [20]. The reads depth information for the above homozygous SNPs/InDels in the offspring pools was obtained to calculate the SNP/InDel index [21]. We filtered out those points that had an SNP/InDel index in both pools of less than 0.3. The sliding window method was used to present the SNP/InDel index of the whole genome. The average of all SNP/InDel indexes in each window was the SNP/InDel index for this window. Usually, we use a window size of 1 Mb and step size of 10 Kb as default settings. The difference of the SNP/InDel index of the two pools was calculated as the delta SNP/InDel index. The differential candidate genes for male sterile and fertile plants of sweet pepper male sterile line AB91 are identified by calculating the indexes of the SNP and SNP/InDel index. The differential candidate genes of male sterile and female fertile AB91 nuclear male sterile lines were

determined by calculating the values of SNP/InDel index and Δ (SNP/InDel-index).

2.4. Mass Spectrometry Method Identification

The self-crossing F2 generation of the sweet pepper male sterile line AB91 fertile line (Msms) was tested by the flying mass spectrometry Sequenom platform typing technique [22], and the testing results were read in real-time and completely analyzed by the software Mass ARRAY® (Agena Bioscience Inc., San Diego, CA, USA). The detection for all samples was repeated to verify accuracy. The test results were compared with the test results with the results of phenotypic identification in the field to calculate the accuracy rate (the number of test results consistent with the phenotype identification in the field divided by the total number of F2 generation groups), and filter out the candidate genes for sweet pepper nuclear male sterility msc2 according to the accuracy rate.

2.5. Confirmation of Candidate Gene by Quantitative Real-Time PCR

During the full bloom period, flower buds of different sizes were collected, and one anther in each flower bud was taken out for microscopic examination. According to the results of microscopic examination, the flower buds were divided into three periods of microspore tetrad prophase, microspore tetrad, and microspore tetrad anaphase. The anthers were stripped from the flower buds, and the total RNA of the anthers at three different periods was extracted and the RNA was reverse transcribed into cDNA. Primers used for qRT-PCR were designed using Primer 5.0 (<http://www.premierbiosoft.com/primerdesigner/>). The primers of candidate genes used were: 5'-CTTCTTTGAATGGAATGCGTGA-3' and 5'-CACTAATTAGGAAACCTTGCC-3'. The reaction procedure was 95°C for 10 min; 94°C for 15 s, 60°C for 1 min and 72°C for 30 s, for 40 cycles. All qRT-PCR analyses were performed with three replicates of a biological sample.

2.6. Three-Dimensional Protein Structure Prediction

The protein sequences of fertile and sterile plants differential gene were analyzed by the online software Zhang Lab QUARK (<https://zhanglab.ccmb.med.umich.edu/QUARK/>) based on the protein structure., Compare the difference in protein structure of the candidate gene between fertile and sterile plants by predicting the three-dimensional structures of the protein.

3. Results

3.1. Fertility Results and Statistical Analysis

As **Figure 1** shows for the male sterile dual-purpose line of the sweet pepper AB91, the anthers of the fertile plant are full and bright yellow, the stigma is lower than the anther, and the whole anther is covered in pollen; however, the

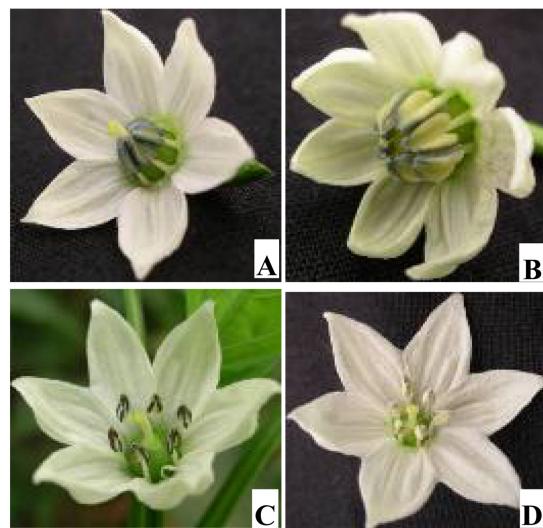


Figure 1. Morphological identification of sterile and fertile plants of the sweet pepper AB91 male sterile dual-purpose line: (A) The sterile plant anther is small and lavender, and the stigma is higher than the anther; (B) fertile plant is full of anthers and bright yellow, and the stigma is lower than the anther; (C) sterile plant anther is shriveled and has no pollen; (D) pollen coversentire anther.

sterile plant anther is small and lavender, the stigma is higher than the anther, the anther is shriveled, and there is no pollen. A total of 473 plants of self-crossing F2 generation of the sweet pepper AB91 dual-purpose fertility plant (Msms) were planted in a field, and fertility identification was performed at the flowering stage. As the investigation results for the F2 groups imply, there were 359 fertile plants and 114 sterile plants; the ratio between fertile and sterile plants was in line with the expected theoretical value 3:1 ($\chi^2 = 0.101 < \chi^2_{0.05} = 0.751$). The fertile plants (Msms, MsMs) of the male sterile F2 generation of sweet peppers were singly seeded and planted separately; the group size for each single plant ranges from 30 to 40. The fertile plant (Msms) and pure dominant fertile plant (MsMs) of sweet pepper nuclear male sterile hybrid were filtered out according to the fertility separation of eachlineof F3 generation. There were 110 pure dominant fertile plants and 249 heterozygous fertile plants in F2.

3.2. Bulked-Segregant Analysis and Data Analysis

MsMs and msms pools were sequenced by the bulked-segregant analysis sequencing (BSA-Seq) technique; 417.04 million and 467.1 million filtered reads were obtained from the MsMs and msms pools, respectively. The MsMs-pool comparison ratio was 99.31%, the average coverage depth for the reference genome (excluding the N region) was 19.26X, and the 1X coverage degree (at least one base got covered) was 92.78%. The msms-pool comparison ratio was 99.03%, the average coverage depth for the reference genome (excluding the N

region) was 21.55X, and the 1X coverage degree (at least one base got covered) was 92.95%. According to the results in **Table 1**, the comparison results are normal and can be used for subsequent mutation detection and correlation analysis.

In order to reduce the impact of sequencing errors and comparison errors, the SNP index and InDel index polymorphism sites were filtered out, and the number of polymorphic marker sites was 11,348,482 after filtration. By making a difference between the SNP index and InDel index for two descendants by choosing a 95% confidence level. The window larger than the threshold was selected as the candidate interval, and the numbers of polymorphism marker sites selected out for SNP and InDel were 27,541 and 1865 respectively. For the candidate site extraction ANNOVAR annotation results, stop loss, and stop gain were selected with the priority, non-synonymous mutation, or alternative splicing site selected as the candidate gene. The number of SNP candidate genes filtered out from the fertile and sterile gene pools was 35 (InDel has no annotated candidate genes above), and 33 candidate genes of them located at chromosome #5 (**Figure 2**, **Table S1**).

3.3. Functional Annotations of Genes

The 34 genes in the correlated region are compared to the databases of NR, SwissProt [23]. GO [24] and KEGG [25] used BLAST software to identify 10 candidate genes and 11 candidate sites connecting to sweet pepper nuclear infertility, respectively, and they are speculated to be related to putative pentatrico-peptide repeat-containing protein., galactinol synthase, glucosyltransferase, zeatin O-xylosyltransferase, threonine-protein kinase, histone acetyltransferase, U-box domain-containing protein, cell division cycle, and apoptosis regulator

Table 1. Sequencing depth and coverage ratio statistics.

Sample	Mapped reads	Total reads	Mapping rate (%)	Average depth (X)	Coverage at least 1X (%)
MsMs	414,182,342	417,041,448	99.31	19.26	92.78
msms	462,562,325	467,111,904	99.03	21.55	92.95

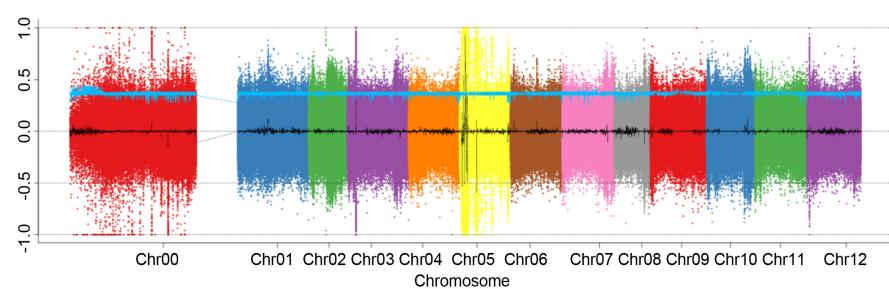


Figure 2. Distribution of msms and MsMs-bulk Δ All-index at chromosome for sweet pepper male sterile. Horizontal axis: chromosome length (Mb); vertical axis: Δ (All-index); and the blue horizontal line is the threshold at the 95% confidence level.

protein. All these candidate gene sites are located at chromosome #5. Hence, the candidate gene controlling the sweet pepper nuclear male sterile AB91 is determined at chromosome #5 (**Table 2**).

3.4. Mass Spectrometry Analysis

A mass spectrometry genotyping technique was used to detect the 11 mutation sites obtained by BSA-sequencing a population of 222 self-crossed F2 generation of sweet pepper nuclear male sterile dual-purpose AB91 fertile plant (Msms). Both sites 19,194,137 and 34,599,677 did not exhibit complete cosegregation with msc2. Therefore, the two corresponding genes Capana05g000617 and Capana05g000896 were excluded. The remaining 9 locations showed co-segregation, but it was slightly different from the msc2 phenotype detection in the field; the mutation site 28,594,037 had the highest similarity with msc2, and the accuracy rate (the number consistent with the msc2 phenotype detection in the field/F2 generation total population size) was 99.5%, and the corresponding gene for this site is Capana05g000747. Hence, the Capana05g000747 gene was determined to be the most powerful candidate gene for msc2 (**Figure 3**).

3.5. Expression Pattern of Fertile and Sterile Plants in Different Periods

The qRT-PCR was used to detect the changes of msc2 expression in fertile and

Table 2. Candidate genes comments.

TransID	Pos	Subject Description
Capana05g000747	28594037	PREDICTED: uncharacterized protein At2g02148 isoform X1 [Solanum lycopersicum]
Capana05g000767	30026988	PREDICTED: pentatricopeptide repeat-containing protein At1g03100, mitochondrial-like isoform X1 [Solanum tuberosum]
Capana05g000608	18775367 18776004	PREDICTED: galactinol synthase 2-like [Nicotiana tomentosiformis]
Capana05g000611	18835766	PREDICTED: putative glucosyltransferase [Solanum lycopersicum]
Capana05g000612	18852022	PREDICTED: zeatin O-xylosyltransferase-like [Solanum lycopersicum]
Capana05g000617	19194137	PREDICTED: zeatin O-xylosyltransferase-like [Solanum lycopersicum]
Capana05g000657	22059680	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At1g34110-like [Solanum tuberosum]
Capana05g000711	25862908	PREDICTED: cell division cycle and apoptosis regulator protein 1-like [Solanum tuberosum]
Capana05g000715	26374526	PREDICTED: histone acetyltransferase HAC1-like isoform X1 [Solanum tuberosum]
Capana05g000896	34599677	PREDICTED: U-boxdomain-containing protein 21-like [Solanum tuberosum]

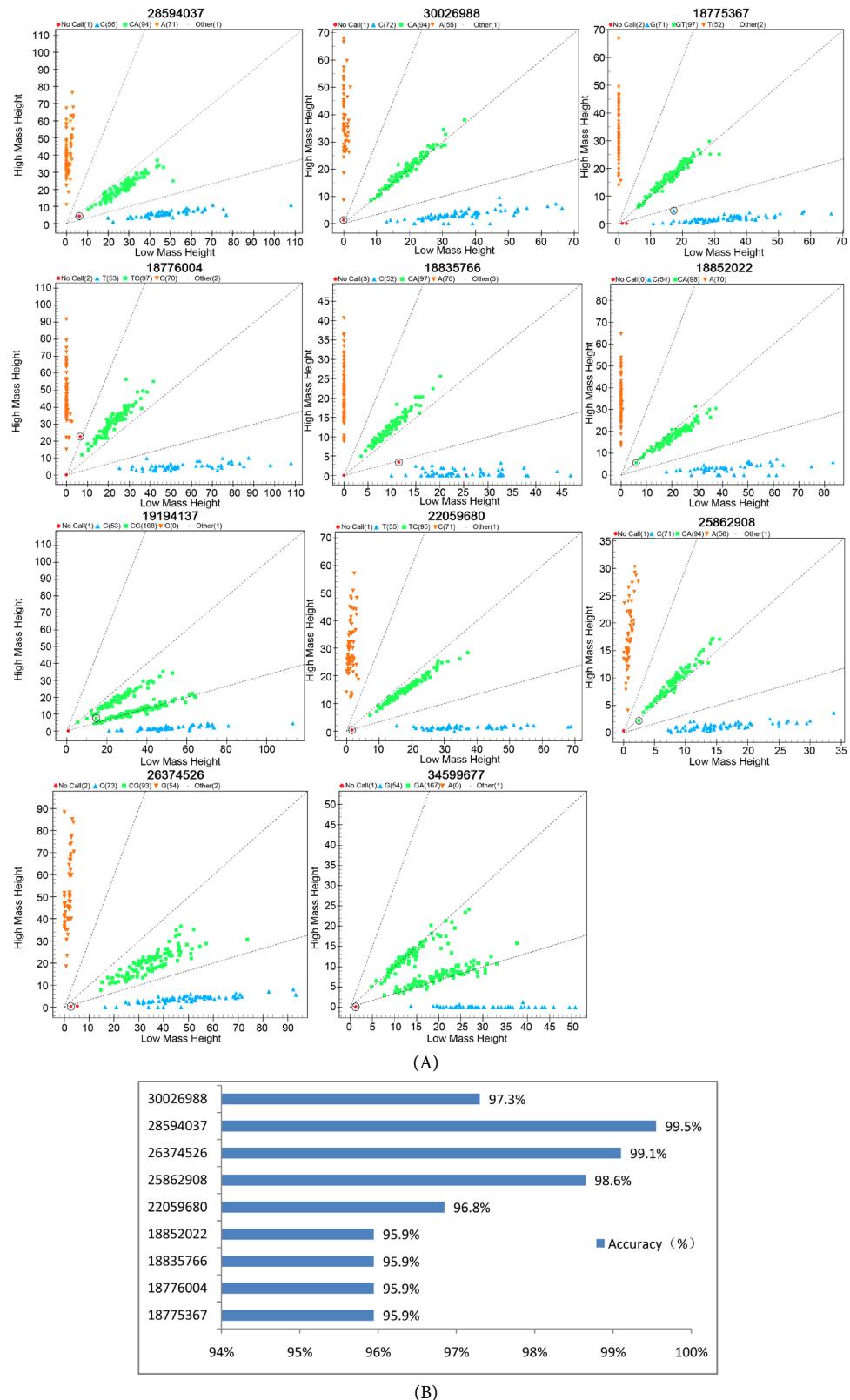


Figure 3. Mass spectrometry genotyping test results for sweet pepper nuclear male sterile AB91: (A) mass spectrometry detection result of a high accuracy candidate site; (B) accuracy histograms of sweet pepper nuclear male sterile AB91F2 generation for 9 candidate sites.

sterile strains of sweet pepper male sterile AB91 at three stages of microspore tetrad prophase, microspore tetrad and microspore tetrad anaphase. Gene of *msc2* were expressed in the three stages, and the expression of sterile plants were higher than fertile plants. However, the expression of sterile plants was significantly higher than fertile plants in the tetraspore prophase stage ($p < 0.05$) (Figure 4). Therefore, it can be presumed that *msc2* is a specific gene. The mutation of bases was significantly affected, and the mutation occurred in the prophase of tetraspore, which was earlier than the cytological observation of tetraspore [7].

3.6. Sequence Analysis of Male Sterility Gene *msc2* for Sweet Pepper Line AB91

The gene *msc2* of sweet pepper male sterile AB91 is composed of eight exons and seven introns. The cDNA sequence is 1350 bp in length and encodes 450 amino acids. The mutation site of the sterility gene is located at the 6th exon of Capana05g000747, the base C is mutated into A, and the amino acid is changed from alanine to serine, which causes a change in the fertility expression of the sterile plant of sweet pepper nuclear male sterile dual-purpose line AB91, leading to male sterility (Figure 5).

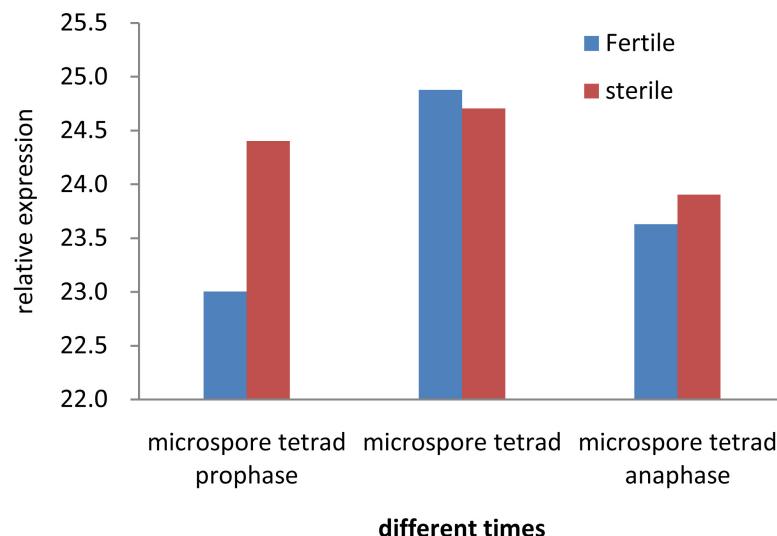


Figure 4. The changes of *msc2* expression in fertile and sterile plants of sweet pepper male sterile AB91 at three stages.

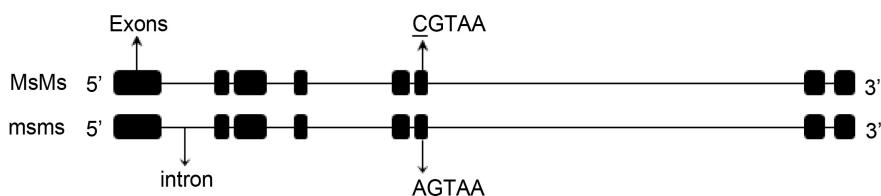


Figure 5. Sequence information of the MSMS gene and the msms gene in sweet pepper nuclear male sterile line AB91.

3.7. Three-Dimensional Protein Structure Prediction

The protein sequences of fertile and sterile plants candidate gene Capana05g000747 were analyzed by the online software Zhang Lab QUARK based on the protein structure. To construct the three-dimensional protein structure of the fertile and sterile plants candidate gene Capana05g000747 in the sweet pepper male sterile line AB91. Three-dimensional structures revealed the site “c” base mutation of Capana05g000747 gene lead to significant differences (**Figure 6**). Therefore, it can be concluded that the base “c” mutation of gene Capana05g000747 may lead to pollen abortion.

3.8. Homology of Nuclear Male Sterility Gene *msc2* in Sweet Pepper AB91 with Other Species

For a further functional study of *msc2*, we performed a sequence alignment and homology relationship analysis for the *genemsc2* and other species, which showed that *msc2* gene was highly conserved and at the same branch as tomato, potato, and tobacco, indicating that their homology relationship is very close and *msc2* is more likely to have the same effect as a close homology. However, the function of gene *msc2* in tomato, potato, and tobacco sequences is predicted, and its function was found to be similar to the At2g02148 gene, which encodes the pentatricopeptide repeats (PPR) protein. This finding preliminarily implies that the *msc2* gene has a connection with the PPR protein (**Figure 7**).

4. Discussion

Employing male sterility to breed advantageous hybrids is the most economical and efficient approach. To date, two types of recessive nuclear male sterility resources for sweet peppers have been discovered in China; one is the pepper male sterile material discovered by Shizhou Yang with the nuclear sterility gene named *msc1*, and the other is the sweet pepper male sterile line AB91 used in this study with its nuclear sterility gene named *msc2*. To date, sweet pepper nuclear male sterile line AB91 has been used to breed five national recognized

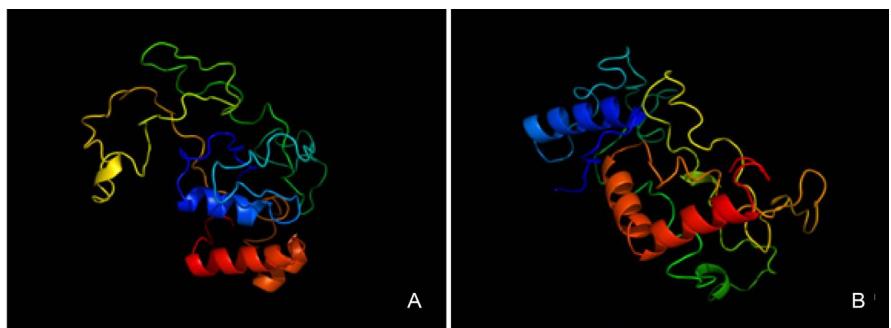
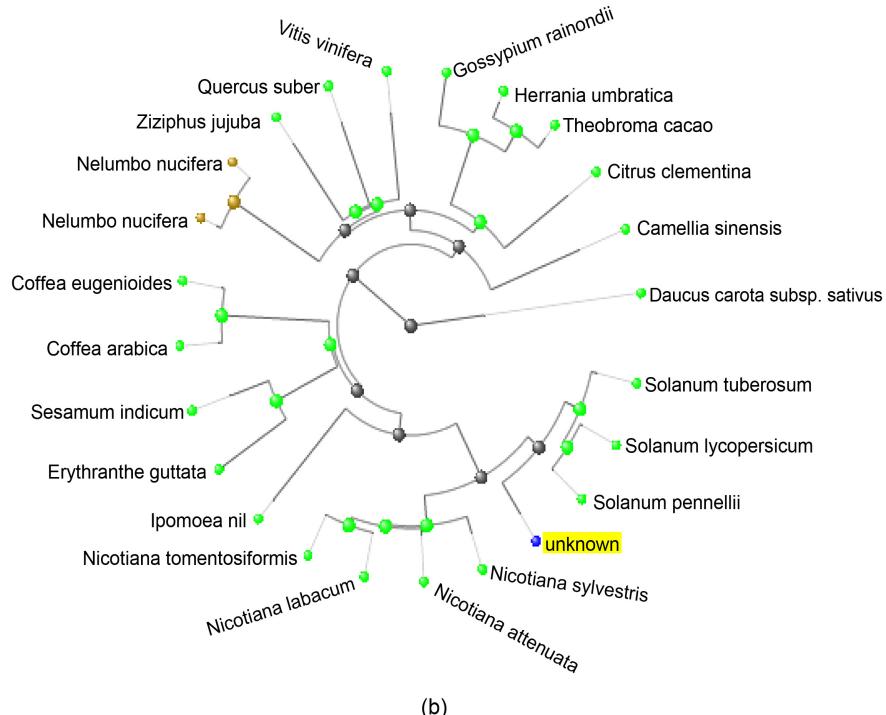


Figure 6. Three-dimensional protein structure prediction of fertile and sterile plants candidate genes Capana05g000747. (A) Three-dimensional protein structure prediction of Capana05g000747 in fertile plants; (B) Three-dimensional protein structure prediction of Capana05g000747 in sterile plants.

capana05g00747	ALEAAGGLYRKLAKKFPRKGFC	NFKRRELATSFETRARFP	238
lycopersicum			239
tabacum			240
tuberosum			239
Consensus	aleaagglyrklaakkfprkgfc	fkrelatsfetrrarfp	
capana05g00747	EIVIQEEKRVRFVVVNGLAIVEKP	TSLCICLLAEWFPRMTG	278
lycopersicum			279
tabacum			280
tuberosum			279
Consensus	elviqeeekrvrfvvvnglaivekp	sl iddaewf rmtg	
capana05g00747	RNEVAISPRDYKEYAPRHKYRRASN	SISNITGISAATSTD	318
lycopersicum			319
tabacum			320
tuberosum			319
Consensus	rnev sprdykfyaprhkyrrasn	isnitg saatstd	
capana05g00747	NASSISACQSYRSVSEESQQTTSKQHMQPIAHQAQFHPLQ	358	
lycopersicum			359
tabacum			360
tuberosum			359
Consensus	nssl agqsyrsveseqqttskqhmqplvhqaqfhplq		

(a)



(b)

Figure 7. Homologous relationship for gene *msc2* between sweet pepper nuclear male sterility AB91 and other species. (a) Blastp analysis of *msc2* in Capsicum, lycopersicum, tabacum and tuberosum. Red box: Represents the mutant site; (b) Phylogenetic tree of *msc2* in different species. Unknown: represents annotated with gene *msc2*.

vegetable varieties (JiYan5, JiYan6, JiYan15, JiYan16, JiYan108) and nine provincial recognized vegetable varieties (JiYan8, JiYan12, JiYan13, JiYan19, JiYan4, JiYan105, JiYan102, JiYan20, JiYan28), which are widely planted throughout China. Therefore, it is of great meaning to study the sterility gene *msc2* of the sweet pepper male sterile line AB91. However, the research on the infertility mechanism of the *msc2* gene has not been reported yet. Whole genome sequencing technology (WGS) is currently the most effective method for group function genetic mining and has many advantages such as its ability to contain

comprehensive, efficient, and accurate information. WGS is a bioinformatics technique that performs differential sequence analysis of individuals or groups at the genomic level to explore species and filter out functional genes [26]. Currently, WGS has been widely used in rice [27], cucumber [28], potato [29], watermelon [30], sorghum [31], and other crops. Therefore, this study carried out in-depth research to analyze the reason for abortion caused by gene *msc2* using genetic sequencing and genotyping technology.

4.1. Material with Prominent and Genetically Stable Objective Traits

The sweet pepper male sterile dual-purpose line AB91 is a sister cross-hybrid whose agronomic traits within groups have remained stable after years of breeding and whose sterility is controlled by recessive nuclear genes. The pure dominant fertile plant has a genetic background that is consistent with that of the pure recessive sterile plant, which avoids the wrong interference analysis caused by the genetic background difference between fertility materials.

4.2. Capana05g000747 Is an Important Candidate Gene of *msc2*

The phenomenon of nuclear male sterile caused by point mutation has been reported in cucumber [32], maize [33] and other crops, but has not been reported in sweet pepper. In this study, the mutation site that caused the abortion of sweet pepper male sterile AB91, which is on chromosome #5, was first discovered through WGS technology. Flight mass spectrometry genotyping and quantitative real-time PCR were used to verify the differential gene and filter out the gene Capana05g000747, which is most likely to be the important gene leading to the abortion of sweet pepper male sterile AB91. For the gene Capana05g000747 including eight exons and seven introns, the mutation site 28,549,037 is a non-synonymous mutation site with its position at the 6th exon of Capana05g000747; the features with base C mutated into A and the amino acid changed from alanine to serine. The gene was further verified by quantitative real-time PCR and three-dimensional protein structure prediction. It was found that the period of base mutation was the microspore tetrad prophase, which was earlier than the cytological observation of tetraspore. It may be because the cytological identification is relatively late to the genetic phenotype change. Although the genes changed in the early stage of the tetraspores, the cells did not change. In addition, by predicting the three-dimensional protein structure, it shows that the point mutations in the sterile plants have significant changes at the protein level. Therefore, it was concluded that Capana05g000747 was an important candidate gene of sterility expression in sweet pepper male sterile AB91.

4.3. Sequence Analysis of the Capana05g000747 Gene

For a further function of the gene Capana05g000747, we performed a sequence

alignment analysis that showed that the gene homology is close to tomato, potato, and tobacco, but their functional annotations were all predicted and are similar to that of the gene At2g02148. At2g02148 encodes a PPR gene protein, which is encoded by a nuclear gene and consists of 35 degenerate amino acids in a series of repeating units; most of the protein is transported into organelles to fulfill their functions [8]. The PPR gene protein has important physiological functions in the gene performance process of organelles and is involved in almost all stages of gene performance, including transcription [34], RNA splicing [35], RNA editing [36], translation [37], and RNA stability maintenance [38]. Previous studies have shown that the PPR gene protein is dispersed all over the entire pepper genome [39] [40], while most of the Rf candidate genes are on chromosome #6 [41] [42] and the sterility performance is realized by controlling the male sterility related genes. However, our research shows that the candidate gene protein PPR of the nuclear male sterility line AB91 on chromosome #5 has a base mutation, which causes a protein function change, resulting in abortion microspores, but the molecular mechanism of the PPR gene protein in the sweet pepper nuclear male sterility AB91 is still unclear and its verification will require further research.

Acknowledgements

We thank LetPub (<http://www.letpub.com/>) for its linguistic assistance during the preparation of this manuscript; and thank Novogene for performing the BSA- Seq.

Availability of Data and Material

All data generated or analysed during this study are included in this published article and its supplementary information files or are available from the corresponding author on request.

Funding

This work was supported by National Key Research and Development Projects (No. 2016YFD0101704), Hebei Science and Technology Support Program Project (No. 16226308D-9).

Authors' Contributions

Yanqin Fan is the project leader and designed the project. Yaning Meng designed the project and performed the experiment. Libin Yan designed the project and conducted field trials for identification. Hongxiao Zhang participated in the design of experiments.

Conflicts of Interest

The authors declare that they have no competing interests.

References

- [1] Martin, J.A. and Crawford, J.H. (1951) Several Types of Sterility in *Capsicum frutescens*. *Horticultural Science*, **57**, 335-338.
- [2] Wang, D.Y., Yang, F.M., Li, Y. and Wang, H.M. (2008) Advances in Research on Male Sterility Genes in Pepper. *China Vegetables*, **9**, 40-43.
- [3] Bartoszewski, G., Waszczak, C., Gawronski, P., Stepien, I., Bolibok, B.H., Palloix, A., Lefebvre, V., Korzeniewska, A. and Niemirowicz, S.K. (2012) Mapping of the ms8 Male Sterility Gene in Sweet Pepper (*Capsicum annuum* L.) on the Chromosome P4 Using PCR Based Markers Useful for Breeding Programmes. *Euphytica*, **186**, 453-461. <https://doi.org/10.1007/s10681-012-0637-9>
- [4] Lee, J., Han, J.H., An, C.G., Lee, W.P. and Yoon, J.B. (2010) A CAPS Marker Linked to a Genic Male-Sterile Gene in the Colored Sweet Pepper "Paprika" (*Capsicum annuum* L.). *Breeding Science*, **60**, 93-98. <https://doi.org/10.1270/jbsbbs.60.93>
- [5] Fan, Y.Q. and Guo, J.Y. (1993) A Preliminary Report on Breeding of Male-Sterile Lines AB91 in Sweet Pepper. *Acta Agriculturae Universitatis Pekinensis*, **19**, 118-122.
- [6] Fan, Y.Q., Liu, Y. and Yan, L.B. (2003) Breeding and Application of Male-Sterile Lines in Sweet Pepper. *Acta Agriculturae Boreali Sinica*, **18**, 81-83.
- [7] Luo, S.X., Fan, Y.Q., Shen, S.X., Chen, X.P., Liu, Y. and Yan, L.B. (2006) Observation on Microspore Development of Genetic Male-Sterile Pepper. *Acta Agriculturae Boreali Sinica*, **21**, 99-103.
- [8] Ma, N., Liu, C., Yang, W.C. and Shen, H.L. (2016) *PAP3* Regulates Stamen but Not Petal Development in *Capsicum annuum* L. *Horticultural Plant Journal*, **2**, 91-96. <https://doi.org/10.1016/j.hpj.2016.02.009>
- [9] Li, X., Liu, C., Da, F.J., Ma, N. and Shen, H.L. (2013) Expression Pattern of Class B Gene *PAP3* in Flower Development of Pepper. *International Journal of Molecular Sciences*, **14**, 24643-24655. <https://doi.org/10.3390/ijms141224643>
- [10] Becker, A. and Theissen, G. (2003) The Major Clades of MADS-Box Genes and Their Role in the Development and Evolution of Flowering Plants. *Molecular Phylogenetics and Evolution*, **29**, 464-489. [https://doi.org/10.1016/S1055-7903\(03\)00207-0](https://doi.org/10.1016/S1055-7903(03)00207-0)
- [11] Whipple, C.J., Ciceri, P., Padilla, C.M., Ambrose, B.A., Bandong, S.L. and Schmidt, R.J. (2004) Conservation of B-Class Floral Homeotic Gene Function between Maize and Arabidopsis. *Development*, **131**, 6083-6091. <https://doi.org/10.1242/dev.01523>
- [12] Fan, S., Deng, M.H., et al. (2016) Isolation and Characterization of Citrate Synthase Gene *CaCTS* in Pepper Cytoplasmic Male Sterility. *Research Journal of Biotechnology*, **11**, 9-15.
- [13] Deng, M.H., Lv, J.H. and Zhao, K. (2017) Molecular Cloning Sequence Characterization of a Novel Pepper Gene *MDH* and Its Effect on Cytoplasmic Male Sterility. *Research Journal of Biotechnology*, **12**, 62-68.
- [14] Guo, J.J., Liu, C., Wang, P., Cheng, Q., Sun, L., Yang, W.C. and Shen, H.L. (2018) The Aborted Microspores (AMS)-Like Gene Is Required for Anther and Microspore Development in Pepper (*Capsicum annuum* L.). *International Journal of Molecular Sciences*, **19**, 12-20. <https://doi.org/10.3390/ijms19051341>
- [15] Deng, M.H., Wen, J.F., Huo, J.L., et al. (2012) cDNA Cloning Sequence Characterization and Expression Analyses of a Novel Gene-*TPI* from Pepper Cytoplasmic Male Sterility(CMS). *Journal of Medicinal Plants Research*, **6**, 1997-2003.
- [16] Cheng, Q., Wang, P., Liu, J.Q., et al. (2018) Identification of Candidate Genes Un-

derlying Genic Male-Sterile msc1 Locus via Genome Resequencing in *Capsicum annuum* L. *Theoretical and Applied Genetics*, **131**, 1861-1872.
<https://doi.org/10.1007/s00122-018-3119-1>

- [17] Qin, C., Yu, C.S., Shen, Y.O., Fang, X.D., Chen, L., et al. (2014) Whole-Genome Sequencing of Cultivated and Wild Peppers Provides Insights into Capsicum Domestication and Specialization. *Proceedings of the National Academy of Sciences*, **111**, 5135-5140. <https://doi.org/10.1073/pnas.1400975111>
- [18] Healey, A., Furtado, A., Cooper, T. and Henry, R.J. (2014) Protocol: A Simple Method for Extracting Next-Generation Sequencing Quality Genomic DNA from Recalcitrant Plant Species. *Plant Methods*, **10**, 21.
<https://doi.org/10.1186/1746-4811-10-21>
- [19] Langmead, B. and Salzberg, S.L. (2012) Fast Gapped-Read Alignment with Bowtie 2. *Nat Methods*, **9**, 357-359. <https://doi.org/10.1038/nmeth.1923>
- [20] McKenna, A., Hanna, M., Banks, E., et al. (2010) The Genome Analysis Toolkit: A Map Reduce Framework for Analyzing Next Generation DNA Sequencing Data. *Genome Research*, **20**, 1297-1303. <https://doi.org/10.1101/gr.107524.110>
- [21] Takagi, H., Abe, A., Yoshida, K., et al. (2013) QTL-seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole Genome Resequencing of DNA from Two Bulked Populations. *The Plant Journal*, **74**, 174-183. <https://doi.org/10.1111/tpj.12105>
- [22] Schaeffeler, E., Zanger, U.M., Eichelbaum, M., et al. (2008) Highly Multiplexed Genotyping of Thiopurine s-Methyltransferase Variants Using MALDI-TOF Mass Spectrometry: Reliable Genotyping in Different Ethnic Groups. *Clinical Chemistry*, **54**, 1637-1647. <https://doi.org/10.1373/clinchem.2008.103457>
- [23] Yip, Y.L., Famiglietti, M., Gos, A., Duek, P.D., David, F.P.A., Gateau, A. and Bairoch, A. (2008) Annotating Single Amino Acid Polymorphisms in the UniProt Swiss-Prot Knowledge Base. *Human Mutation*, **29**, 361-366.
<https://doi.org/10.1002/humu.20671>
- [24] Sherlock, G. (2009) Gene Ontology: Tool for the Unification of Biology. *Canadian Institute of Food Science & Technology Journal*, **22**, 415.
- [25] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2016) KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Research*, **44**, 457-462. <https://doi.org/10.1093/nar/gkv1070>
- [26] Pan, Z.Y., He, X.Y., Liu, Q.Y., Guo, X.F., Cao, X.H., Hu, W.P., Di, R., Wang, X.Y. and Chu, M.X. (2016) Application of Whole Genome Sequencing (WGS) in Population Evolution and Gene Discovery of Domestic Animals. *Journal of Agricultural Biotechnology*, **24**, 1945-1954.
- [27] Li, D.Y., Huang, Y., Song, S.H., et al. (2016) Integrated Analysis of Phenome Genome and Transcriptome of Hybrid Rice Uncovered Multiple Heterosis-Related Loci for Yield Increase. *PNAS*, **113**, E6026-E6035.
<https://doi.org/10.1073/pnas.1610115113>
- [28] Lu, H., Lin, T., Klein, J., et al. (2014) QTL-seq Identifies an Early Flowering QTL Located near Flowering Locus T in Cucumber. *Theoretical and Applied Genetics*, **127**, 1491-1499. <https://doi.org/10.1007/s00122-014-2313-z>
- [29] Zhang, G.D., Liu, B.L., Li, X.B., Si, H.J. and Li, X.Q. (2016) Effects of Potato Cultivars and Tuber Sections on the Sequencing Depth Pattern in Genome. *Journal of Gansu Agricultural University*, **51**, 43-48.
- [30] Guo, S.G., Zhang, J.G., Sun, H.H., Salse, J., Lucas, W.J., Zhang, H.Y., Zheng, Y., Mao, L.Y., Ren, Y. and Wang, Z.W. (2013) The Draft Genome of Water Melon (*Ci-*

trullus lanatus) and Resequencing of 20 Diverse Accessions. *Nature Genetics*, **45**, 51-58. <https://doi.org/10.1038/ng.2470>

- [31] Zheng, L.Y., Guo, X.S., He, B., Sun, L.J., Peng, Y., Dong, S.S., Liu, T.F., Jiang, S., Ramachandran, S., Liu, C.M. and Jing, H.C. (2011) Genome Wide Patterns of Genetic Variation in Sweet and Grain Sorghum (*Sorghum bicolor*). *Genome Biology*, **12**, R114. <https://doi.org/10.1186/gb-2011-12-11-r114>
- [32] Han, Y.K., Zhao, F.Y., Gao, S., Wang, X.Y., Wei, A.M., et al. (2018) Fine Mapping of a Male Sterility Gene ms-3 in a Novel Cucumber (*Cucumis sativus* L.) Mutant. *Theoretical and Applied Genetics*, **131**, 449-460. <https://doi.org/10.1007/s00122-017-3013-2>
- [33] Liu, C.X., Wang, G.Q., Gao, J., Li, C.Y., Zhang, Z.R., et al. (2018) Characterization, Fine Mapping and Candidate Gene Analysis of Novel, Dominant, Nuclear Male-Sterile Gene Ms53 in Maize. *Euphytica*, **214**, 52. <https://doi.org/10.1007/s10681-018-2132-4>
- [34] Hashimoto, M., Endo, T., Peltier, G., et al. (2003) A Nucleus Encoded Factor CRR2 Is Essential for the Expression of Chloroplast ndhB in Arabidopsis. *The Plant Journal*, **36**, 541-549. <https://doi.org/10.1046/j.1365-313X.2003.01900.x>
- [35] Ichinose, M., Tasak, I.E., Sugita, C., et al. (2012) A PPR-DYW Protein Is Required for Splicing of a Group II Intron of cox1 Pre-mRNA in *Physcomitrella patens*. *The Plant Journal*, **70**, 271-278. <https://doi.org/10.1111/j.1365-313X.2011.04869.x>
- [36] Rudinger, M., Volkmar, U., Lenz, H., et al. (2012) Nuclear DYW-Type PPR Gene Families Diversify with Increasing RNA Editing Frequencies in Liverwort and Moss Mitochondria. *Journal of Molecular Evolution*, **74**, 37-51. <https://doi.org/10.1007/s00239-012-9486-3>
- [37] Tavares, C.F., Camacho, V.Y., Zamudio, O., et al. (2007) The Pentatricopeptide Repeats Present in PET 309 Are Necessary for Translation But Not for Stability of the Mitochondrial COX1 mRNA in Yeast. *Journal of Biological Chemistry*, **283**, 1472-1479. <https://doi.org/10.1074/jbc.M708437200>
- [38] Yamazaki, H., Tasaka, M. and Shinkanai, T. (2004) PPR Motifs of the Nucleus Encoded Factor PGR3 Function in the Selective and Distinct Steps of Chloroplast Gene Expression in Arabidopsis. *The Plant Journal*, **38**, 152-163. <https://doi.org/10.1111/j.1365-313X.2004.02035.x>
- [39] Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., et al. (2004) Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. *Plant Cell*, **16**, 2089-2103. <https://doi.org/10.1105/tpc.104.022236>
- [40] Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., Long, Y., Zhong, Y. and Liu, Y. (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**, 676-687. <https://doi.org/10.1105/tpc.105.038240>
- [41] Jo, Y.D., Kim, Y.M., Park, M.N., Yoo, J.H., Park, M., Kim, B.D. and Kang, B.C. (2010) Development and Evaluation of Broadly Applicable Markers for Restorer of Fertility in Pepper. *Molecular Breeding*, **25**, 187-201. <https://doi.org/10.1007/s11032-009-9318-3>
- [42] Ding, A.M., Li, F.X., Chen, Y.Q., et al. (2014) Homology Based Cloning and Expression Analysis of Rf Genes Encoding PPR-Containing Proteins in Tobacco. *Genetics and Molecular Research*, **13**, 2310-2322. <https://doi.org/10.4238/2014.March.31.11>

Appendix

Tables S1. The number of SNP candidate genes filtered out from the fertile and sterile gene pools.

	TransID	Variant	Chrom	Pos	Ref	Alt
1	Capana05g000605	nonsynonymous	Chr05	18709543	C	T
	Capana05g000605	nonsynonymous	Chr05	18709619	A	C
	Capana05g000605	nonsynonymous	Chr05	18709862	A	C
	Capana05g000605	nonsynonymous	Chr05	18709888	C	T
	Capana05g000605	nonsynonymous	Chr05	18709978	G	C
	Capana05g000605	nonsynonymous	Chr05	18709985	C	G
	Capana05g000605	nonsynonymous	Chr05	18710039	A	C
	Capana05g000605	nonsynonymous	Chr05	18710143	G	A
	Capana05g000605	nonsynonymous	Chr05	18710148	A	T
	Capana05g000605	nonsynonymous	Chr05	18710158	T	G
2	Capana05g000605	nonsynonymous	Chr05	18710174	A	T
	Capana05g000605	nonsynonymous	Chr05	18710176	T	C
	Capana05g000605	nonsynonymous	Chr05	18710504	T	C
	Capana05g000605	nonsynonymous	Chr05	18710859	C	G
	Capana05g000608	nonsynonymous	Chr05	18775367	G	T
	Capana05g000608	nonsynonymous	Chr05	18776004	C	T
	Capana05g000609	nonsynonymous	Chr05	18824359	A	G
	Capana05g000609	nonsynonymous	Chr05	18824551	T	A
	Capana05g000609	nonsynonymous	Chr05	18824575	A	G
	Capana05g000609	nonsynonymous	Chr05	18824725	G	A
3	Capana05g000609	nonsynonymous	Chr05	18825073	G	A
	Capana05g000609	nonsynonymous	Chr05	18825633	T	C
	Capana05g000609	nonsynonymous	Chr05	18825660	C	T
	Capana05g000609	nonsynonymous	Chr05	18825686	G	A
	Capana05g000609	nonsynonymous	Chr05	18827731	C	T
4	Capana05g000611	nonsynonymous	Chr05	18835766	A	C
	Capana05g000612	nonsynonymous	Chr05	18852022	A	C
5	Capana05g000612	nonsynonymous	Chr05	18852023	G	A
	Capana05g000617	nonsynonymous	Chr05	19194110	T	A
6	Capana05g000617	nonsynonymous	Chr05	19194137	C	G
	Capana05g000618	nonsynonymous	Chr05	19354650	G	A
7	Capana05g000618	nonsynonymous	Chr05	19354663	C	T
	Capana05g000657	nonsynonymous	Chr05	22059680	T	C
8	Capana05g000663	nonsynonymous	Chr05	22406793	T	G

Continued

	Capana05g000674	nonsynonymous	Chr05	22684482	T	G
10	Capana05g000674	nonsynonymous	Chr05	22684490	T	A
	Capana05g000674	nonsynonymous	Chr05	22684594	G	A
	Capana05g000674	nonsynonymous	Chr05	22684717	A	G
	Capana05g000678	nonsynonymous	Chr05	22721087	C	T
11	Capana05g000678	nonsynonymous	Chr05	22722060	C	T
	Capana05g000678	nonsynonymous	Chr05	22722965	C	T
12	Capana05g000694	nonsynonymous	Chr05	23907563	G	A
	Capana05g000704	nonsynonymous	Chr05	25471889	A	C
13	Capana05g000704	nonsynonymous	Chr05	25471955	C	A
	Capana05g000704	nonsynonymous	Chr05	25472935	T	A
14	Capana05g000707	nonsynonymous	Chr05	25683162	C	G
15	Capana05g000711	nonsynonymous	Chr05	25862908	C	A
	Capana05g000712	nonsynonymous	Chr05	25922395	A	G
16	Capana05g000712	nonsynonymous	Chr05	25922533	C	T
	Capana05g000712	nonsynonymous	Chr05	25922547	G	A
	Capana05g000712	nonsynonymous	Chr05	25922584	T	C
17	Capana05g000713	nonsynonymous	Chr05	25933940	A	T
18	Capana05g000715	nonsynonymous	Chr05	26374526	G	C
19	Capana05g000744	nonsynonymous	Chr05	28464968	A	G
20	Capana05g000747	nonsynonymous	Chr05	28594037	C	A
21	Capana05g000750	nonsynonymous	Chr05	28736970	C	T
22	Capana05g000755	nonsynonymous	Chr05	29101829	G	T
23	Capana05g000764	nonsynonymous	Chr05	29853146	A	G
24	Capana05g000765	nonsynonymous	Chr05	29860325	A	G
25	Capana05g000767	nonsynonymous	Chr05	30026988	A	C
26	Capana05g000768	nonsynonymous	Chr05	30034996	T	A
	Capana05g000769	nonsynonymous	Chr05	30038608	G	A
	Capana05g000769	nonsynonymous	Chr05	30038737	T	C
	Capana05g000769	nonsynonymous	Chr05	30038856	A	G
	Capana05g000769	nonsynonymous	Chr05	30039309	G	C
27	Capana05g000769	nonsynonymous	Chr05	30039555	A	G
	Capana05g000769	nonsynonymous	Chr05	30039607	A	C
	Capana05g000769	nonsynonymous	Chr05	30039711	A	C
	Capana05g000769	nonsynonymous	Chr05	30039771	A	G
	Capana05g000769	nonsynonymous	Chr05	30040332	C	T
28	Capana05g000792	nonsynonymous	Chr05	31103431	T	C

Continued

	Capana05g000792	nonsynonymous	Chr05	31105598	T	C
	Capana05g000877	nonsynonymous	Chr05	33245788	T	C
	Capana05g000877	nonsynonymous	Chr05	33245913	A	G
	Capana05g000877	nonsynonymous	Chr05	33246027	A	C
	Capana05g000877	nonsynonymous	Chr05	33246057	G	A
29	Capana05g000877	nonsynonymous	Chr05	33246067	A	G
	Capana05g000877	nonsynonymous	Chr05	33246135	G	A
	Capana05g000877	nonsynonymous	Chr05	33246156	C	T
	Capana05g000877	nonsynonymous	Chr05	33246201	A	G
	Capana05g000877	nonsynonymous	Chr05	33246279	A	G
	Capana05g000878	nonsynonymous	Chr05	33248086	A	G
	Capana05g000878	nonsynonymous	Chr05	33248111	T	C
	Capana05g000878	nonsynonymous	Chr05	33248768	A	G
30	Capana05g000878	nonsynonymous	Chr05	33248777	G	A
	Capana05g000878	nonsynonymous	Chr05	33249522	T	C
	Capana05g000878	nonsynonymous	Chr05	33249873	C	T
	Capana05g000878	nonsynonymous	Chr05	33249930	T	C
31	Capana05g000892	nonsynonymous	Chr05	34156958	G	A
	Capana05g000892	nonsynonymous	Chr05	34157577	G	A
32	Capana05g000894	nonsynonymous	Chr05	34277686	C	A
33	Capana05g000896	nonsynonymous	Chr05	34599677	G	A

Abbreviations

- AP: Apetala
 bHLH: Basic Helix-Loop-Helix
 BSA: Bulked Segregant Analysis
 CDS: Coding Sequences
 DEF: Deficiens
 GLO: Globosa
 MDH: Malate Dehydrogenase
 PAP: Pokeweed Antiviral Protein
 PI: Pistillata
 PPR: Pentatricopeptide Repeat
 qRT-PCR: Quantitative RT-PCR
 SNP: Single Nucleotide Polymorphism
 TPI: Triose Phosphate Isomerase
 WGS: Whole Genome Sequencing Technology