

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

YANING MENG

Hebei Academy of Agriculture and Forestry Sciences

libin Yan

the institute of cash crops, hebei academy of agriculture and forestry science
secretariat Tsukuba Office
Department of Research and Information: Norin Suisan Kenkyu Joho Sogo Center

hongxiao zhang

the institute of cash crops, hebei academy of agriculture and forestry science

Yanqin Fan (✉ nkyfanyq@126.com)

The institute of cash crop, hebei academy of agriculture and forestry science <https://orcid.org/0000-0001-5724-0097>

Research Article

Keywords: sweetpepper, nuclear malesterile line, genome-wide resequencing technique, candidate gene

Posted Date: July 20th, 2021

DOI: <https://doi.org/10.21203/rs.2.18377/v5>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Agricultural Sciences on January 1st, 2021. See the published version at <https://doi.org/10.4236/as.2021.122010>.

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 was 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.

Introduction

Sweet pepper (*Capsicum annum* 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 is 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 in breeding 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^[2-4]. 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*^[6]. 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^[7]. 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^[8]; 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.^[9] and Li et al.^[10] 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^[11-12]. Fan et al.^[13] found a new gene CaCTS from the pepper malesterile 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.^[14] 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.^[15] 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.^[16] 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 malesterile 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.^[17] performed genomic resequencing and comparison analysis for the malesterile 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^[18], 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.

Results

Fertility results and statistical analysis

As Fig. 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 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 implies, 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.

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.

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

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 1,865 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 34 candidate genes of them located at chromosome #5 (Fig. 2,Table S1).

Functional Annotations of Genes

The 34 genes in the correlated region are compared to the databases of NR, SwissProt^[19]. GO^[20] and KEGG^[21] 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 pentatricopeptide 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 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).

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]

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 19194137 and 34599677 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 28594037 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*(Fig. 3,TableS2,S3).

Expression pattern of Fertile and sterile plants in different periods

The qRT-PCR were used to detect the changes of *msc2* expression in fertile and 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 ($p < 0.05$) (Fig. 4, Table S4). 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 [8].

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 (Fig. 5).

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 (Fig. 6). Therefore, it can be concluded that the base "c" mutation of gene Capana05g000747 may lead to pollen abortion.

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 gene *msc2* 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 (Fig. 7).

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 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^[22]. Currently, WGS has been widely used in rice^[23], cucumber^[24], potato^[25], watermelon^[26], sorghum^[27], 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.

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.

Capana05g000747 is an important candidate gene of *msc2*

The phenomenon of nuclear male sterile caused by point mutation has been reported in cucumber^[28], maize^[29] 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 28549037 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.

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^[30]. 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^[31], RNA splicing^[32], RNA editing^[33], translation^[34], and RNA stability maintenance^[35]. Previous studies have shown that the PPR gene protein is dispersed all over the entire pepper genome^[36-37], while most of the Rf candidate genes are on chromosome #6^[38-39] 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.

Methods

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.

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^[40] to construct the MsMs and Msms gene banks. The DNA sample

was fragmented into a size of 350bp 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 (Aligent, 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 150bp paired-end reads were generated with an insert size of approximately 350bp.

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^[41]. GATK software was used to detect single-nucleotide polymorphisms (SNPs) and insertion-deletion (InDels)^[42]. The reads depth information for the above homozygous SNPs/InDels in the offspring pools was obtained to calculate the SNP/InDel index^[43]. 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 as 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 $\Delta(\text{SNP/InDel index})$.

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^[44], 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.

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.premierbio.com/primerdesign/>). The primers of candidate genes used were: 5'-CTTCTTTGAATGGAATGCGTGA-3' and 5'-CACTAATTCAGGAAACCTTGCC-3'..The reaction procedure was 95°C for 10 min;94°C for 15s,60°C for 1 min and 72°C for 30s, for 40 cycles. All qRT-PCR analyses were performed with three replicates of a biological sample.

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.

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

Declarations

Ethics approval and consent to participate Not applicable.

Consent to publish Not applicable.

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.

Competing interests The authors declare that they have no competing interests.

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

Authors' contributions Yanqin Fan and Libin Yan designed the project. Yaning Meng analyzed the data, carried out the experiments and wrote the manuscript. Hongxiao Zhang participated in the design of experiments. All authors contributed to the revision of this manuscript and approved the final manuscript.

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

References

1. Martin J.A., and Crawford J.H. (1951) Several types of sterility in *Capsicum frutescens*. Hort. Sci. 57:335-338.
2. Wang D.Y., Yang F.M., Li Y., 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., Stepień I., Bolibok B.H., Palloix A., Lefebvre V., Korzeniewska A., 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(2):453-461.
4. Lee J., Han J.H., An C.G., Lee W.P., Yoon J.B. (2010a) A CAPS marker linked to a genic male-sterile gene in the colored sweet pepper 'Paprika' (*Capsicum annuum* L.). Breed Sci. 60(1): 93-98.
5. Fan Yanqin, Guo Jingyin (1993) A Preliminary Report on Breeding of Male-sterile Lines AB91 in Sweet Pepper. Acta agriculture universitatis pekinensis 19:118-122.

6. Wang D.Y., Yang F.M., Li Y., Wang H.M. (2008) Research progress of genetic male sterility of pepper. *China Vegetables* 9:40-43.
7. Fan Yanqin, Liu Yun, Yan Libin (2003) Breeding and Application of Male-sterile Lines in Sweet Pepper. *Acta Agriculturae Boreali Sinica* 18:81-83.
8. Luo Shuangxia, Fan Yanqin, Shen Shuxing, Chen Xueping, Liu Yun, Yan Libin (2006) Observation on Microspore Development of Genetic Male-sterile Pepper. *Acta Agriculturae Boreali-Sinica* 21(2):99-103.
9. Ma Ning, Liu Chen, Yang Wencai, Shen Huolin (2016) *PAP3* regulates stamen but not petal development in *Capsicum annuum* L. *Horticultural Plant Journal* 2(2):91-96.
10. Li Xin, Liu Chen, Da Fengjiao, Ma Ning, Shen Huolin (2013) Expression pattern of class B gene *PAP3* in flower development of pepper. *International Journal Of Molecular Sciences* 14(12):24643-24655.
11. Becker A., Theissen G (2003) The major clades of *MADS-box* genes and their role in the development and evolution of flowering plants. *Phylogenet Evol* 29:464-489.
12. Whipple C.J., Ciceri P., Padilla C.M., Ambrose B.A., Bandong S.L., Schmidt R.J. (2004) Conservation of B-class floral homeotic gene function between maize and Arabidopsis. *Development* 131:6083-6091.
13. Fan Shi, Deng Minghua, et al (2016) Isolation and characterization of citrate synthase gene *CaCTS* in pepper cytoplasmic male sterility. *Research Journal Of Biotechnology* 11(10):9-15.
14. Deng Minghua, Lv Junheng, Zhao Kai (2017) Molecular cloning sequence characterization of a novel pepper gene *MDH* and its effect on cytoplasmic male sterility. *Research Journal Of Biotechnology* 12(4):62-68.
15. Guo Jinju, Liu Chen, Wang Peng, Cheng Qing, Sun Liang, Yang Wencai, Shen Huolin (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(5):12-20.
16. Deng Minghua, Wen Jinfen, Huo Jinlong, 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(10):1997-2003.
17. Qing Cheng, Peng Wang, Jinqiu Liu, et al (2018) Identification of candidate genes underlying genic male-sterile *msc1* locus via genome resequencing in *Capsicum annuum* L. *Theoretical and Applied Genetics* 131:1861-1872.
18. Cheng Qin, Changshui Yu, Yaou Shen, Xiaodong Fang, Lang Chen, 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(14):5135-5140.

19. Yip YL,Famiglietti M,Gos A,Duek PD,David FPA,Gateau A,Bairoch A(2008)Annotating single amino acid polymorphisms in the UniProt Swiss-Prot knowledgebase.Hum Mutat 29(3):361-6.
20. Sherlock G.(2009)Gene ontology: tool for the unification of biology.Canadian Ins J Food Scie Tech.22(4):415.
21. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M(2016)KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res.44:457-62.
- 22.Pan Z Y,He X Y,Liu Q Y,Guo X F,Cao X H,Hu W P,Di R,Wang X Y,Chu M X(2016)Application of Whole Genome Sequencing(WGS) in Population Evolution and Gene Discovery of Domestic Animals.Journal of Agricultural Biotechnology 24(12): 1945-1954.
23. Dayong lia,Zhiyuan Huangb,Shuhui Song,et al(2016)Integrate analisys of phenome genome and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase.PNAS 113(41):E6026-6035.
24. Lu H,Lin T,Klein J,et al(2014)QTL-seq indentifies an early flowering QTL located near flowering locus T in cucumber.Theoretical and applied genetics 127(7):1491-1499.
- 25.Zhang G D,Liu B I,Li X B,SI HuaiJ,Li X Q(2016)Effects of potato cultivars and tuber sections on the sequencing depth pattern in genome.Journal of Gansu Agricultural University. 51(3): 43-48.
26. Guo S G,Zhang J G,Sun H H,Salse J,Lucas W J,Zhang H Y,Zheng Y,Mao L Y,Ren Y,Wang Z W(2013)The draft genome of water melon(*Citrullus lanatus*)and resequencing of 20 diverse accessions.Nat Genet 45: 51-58.
27. Zheng L Y,Guo X S,He B,Sun L J,Peng Y,Dong S S,Liu T F,Jiang SRamachandran S,Liu C M,Jing H C(2011)Genome wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*).Genome Biol 12:R114.
- 28.Yike Han,Fengyue Zhao,ShangGao,Xianyun Wang,Aimin Wei,et al(2018)Fine mapping of a male sterility gene ms-3 in a novel cucumber(*Cucumis sativus* L.) mutant.Theoretical and Applied Genetics (2018) 131:449-460.
- 29.Chaoxian Liu,Guoqiang Wang,Jie Gao,Chunyan Li,Ziru Zhang,et al(2018)Characterization,fine mapping and candidate gene analysisof novel,dominant,nuclear male-sterilegene Ms53 in maize.Euphytica (2018) 214:52.
- 30.Ma Ning,Liu Chen,Yang Wencai,Shen Huolin(2016)*PAP3* regulates stamen but not petal development in *Capsicum annuum* L.Horticultural Plant Journal 2(2):91-96.
31. Hashimoto M,Endo T,Peltier G,et al(2003)A nucleus encoded factor CRR2 is essential for the expreision of chloroplast ndhB in Arabidopsis.The Plant Journal 36(4):541-549.

32. Ichinose M, Tasak I E, Sugita C, et al (2012) A PPR-DYW protein is required for splicing of a group intron of *cox1* pre-mRNA in *Physcomitrella patens*. *The Plant Journal* 70(2):271-278.
33. 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(1-2):37-51.
34. 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(3):1472-1479.
35. Yamazaki H, Tasaka M, 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(1):152-163.
36. Lurin, C., C. Andres, S. Aubourg, M. Bellaoui, et al (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16:2089–2103.
37. Wang, Z., Y. Zou, X. Li, Q. Zhang, L. Chen, H. Wu, D. Su, Y. Chen, J. Guo, D. Luo, Y. Long, Y. Zhong, and Y. Liu (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.
38. Jo Y.D., Kim Y.M., Park M.N., Yoo J.H., Park M., Kim B.D., Kang B.C. (2010) Development and evaluation of broadly applicable markers for Restorer of fertility in pepper. *Mol. Breeding* 25:187-201.
39. 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. *Genet Mol Res* 13: 2310-2322.
40. Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* 10:21.
41. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
42. McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next generation DNA sequencing data. *Genome research* 20(9): 1297-1303.
43. 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(1):174-183.
44. Schaeffeler E, Zanger U M, Eichelbaum M, et al (2008) Highly multiplexed genotyping of thiopurine smethyl transferase variants using malditof mass spectrometry: Reliable genotyping in different ethnic groups. *Clin Chem* 54(10):1637-1647

Figures



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 covers entire anther.

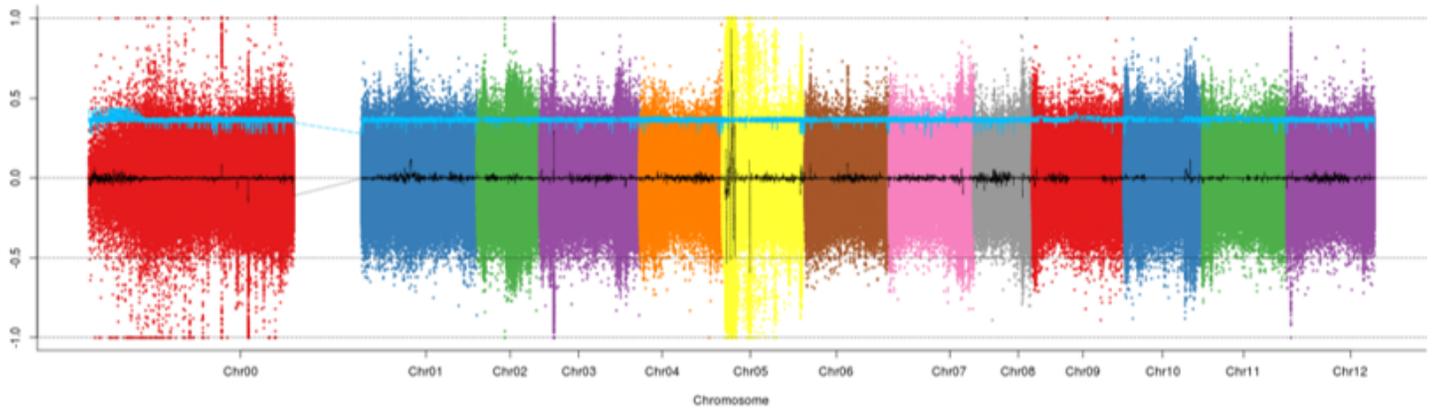
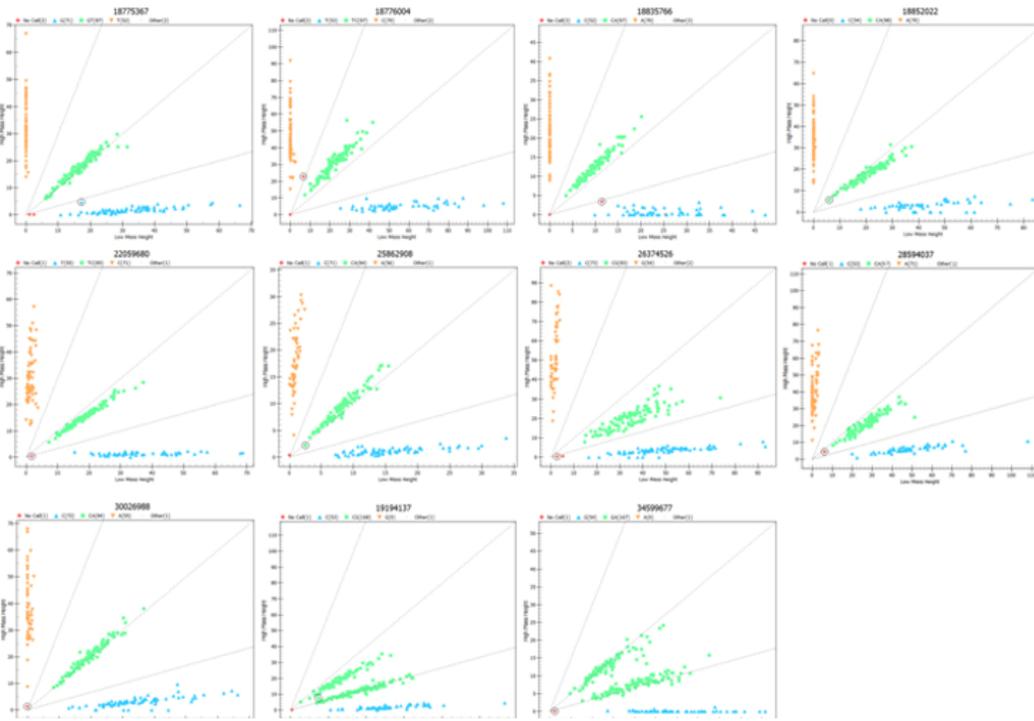
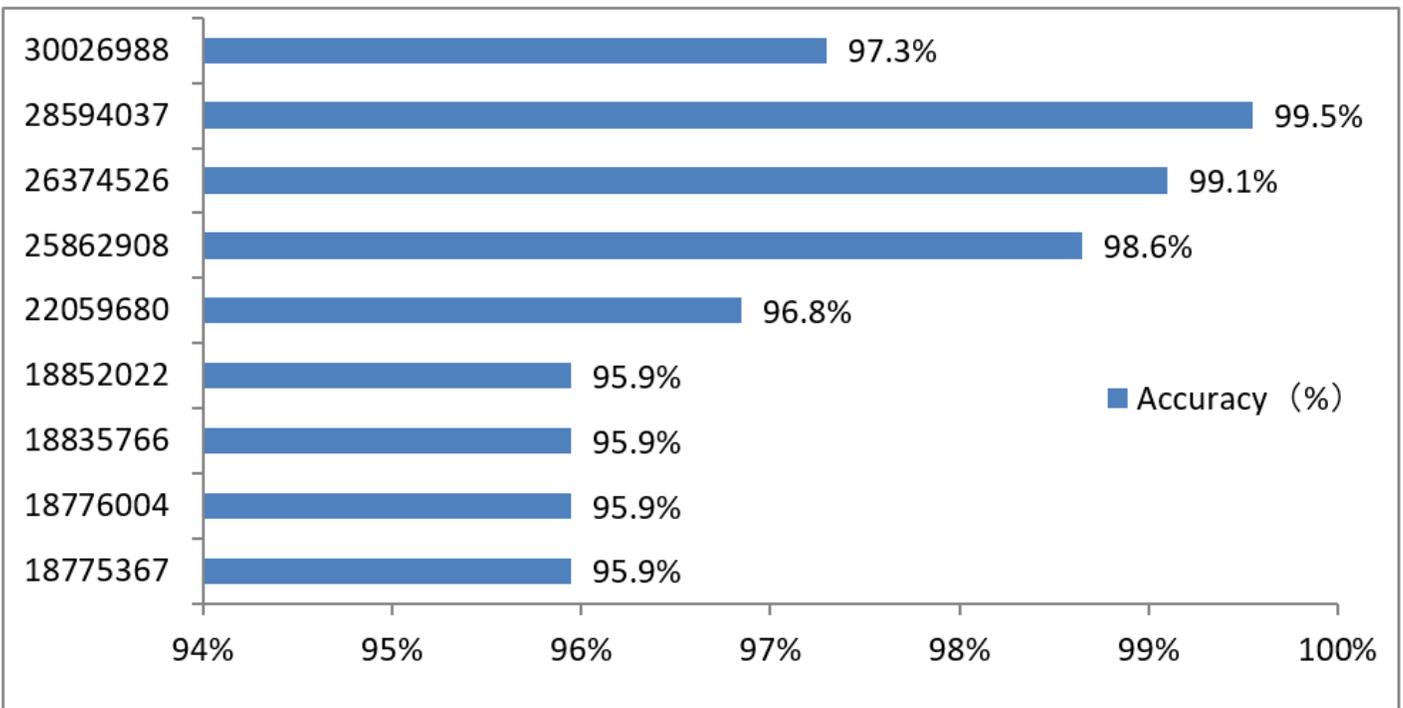


Figure 2

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.

A**B****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.

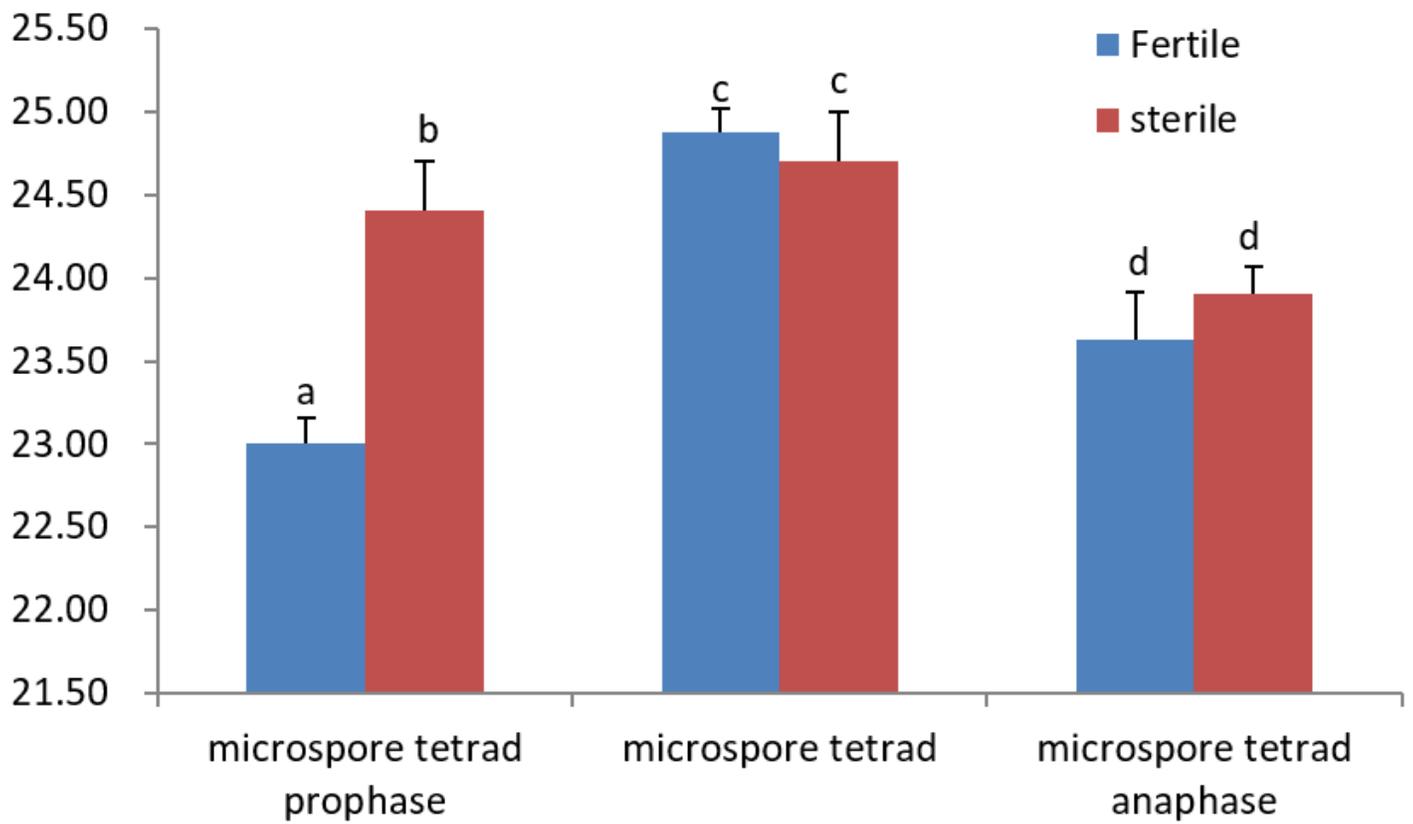


Figure 4

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

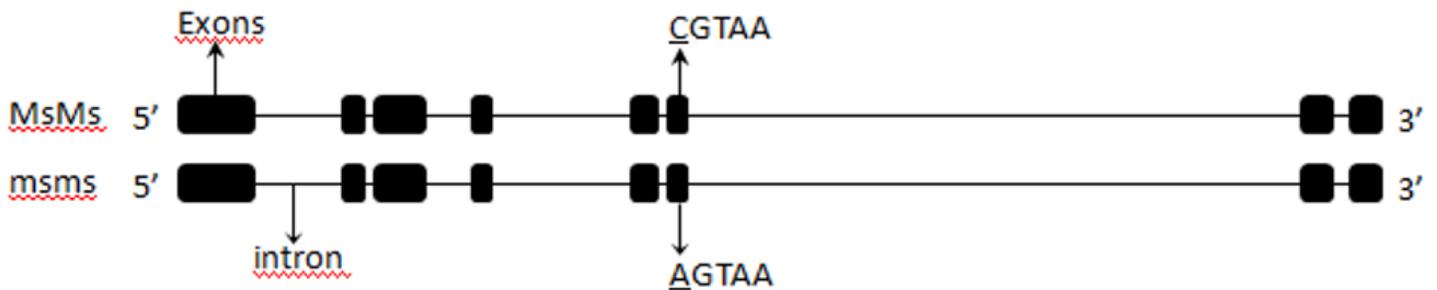


Figure 5

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

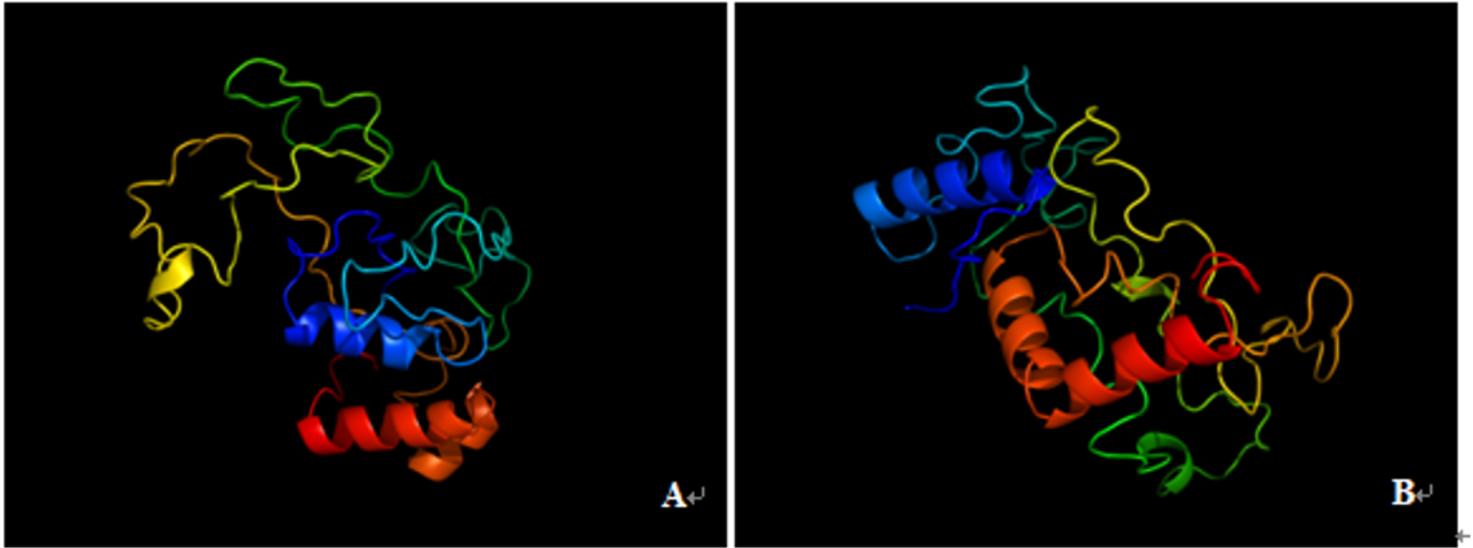


Figure 6

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

A

capana05g00747	ALEAAGGLYRKLAKKFPRKGFQVFKRRELATSFETRARFP	238
lycopersicum	ALEAAGGLYRKLAKKFPRKGFQVFKRRELATSFETRARFP	239
tabacum	ALEAAGGLYRKLAKKFPRKGFQVFKRRELATSFETRARFP	240
tuberosum	ALEAAGGLYRKLAKKFPRKGFQVFKRRELATSFETRARFP	239
Consensus	aleaagglyrklakkfprkqpc fkrrelatsfetrarfp	
capana05g00747	ELVIQEEKRVRFVVVNGLAIVEKPTSLCIDDAEWFRRMTG	278
lycopersicum	ELVIQEEKRVRFVVVNGLAIVEKPTSLCIDDAEWFRRMTG	279
tabacum	ELVIQEEKRVRFVVVNGLAIVEKPTSLCIDDAEWFRRMTG	280
tuberosum	ELVIQEEKRVRFVVVNGLAIVEKPTSLCIDDAEWFRRMTG	279
Consensus	elviqeekrvrfvvvnglaivekpsl iddaewf rmtg	
capana05g00747	RNEVAISPRDYKFYAPRHKYRRASNSISNITGLSAFTSTE	318
lycopersicum	RNEVTVSPRDYKFYAPRHKYRRASNSISNITGLSAFTSTD	319
tabacum	RNEVAISPRDYKFYAPRHKYRRASNSISNITGLSAFTSTD	320
tuberosum	RNEVTISPRDYKFYAPRHKYRRASNSISNITGLSAFTSTD	319
Consensus	rnev sprdykfyaprhkyrrasn isnitglsaftst	
capana05g00747	NASSISAGQSYRSVSEESQCTTSKQCMQPLAHQAQFHPLQ	358
lycopersicum	NASSISAGQSYRSVSEESQCTTSKQCMQPLVHQAQFHPLQ	359
tabacum	NASSISAGQSYRSVSEESQCTTSKQCMQPLSHQAQFHPLQ	360
tuberosum	NASSISAGQSYRSVSEESQCTTSKQCMQPLVHQAQFHPLQ	359
Consensus	n ssl agqsyrsvse sqqtskq mqpl hqaqfhplq	

B

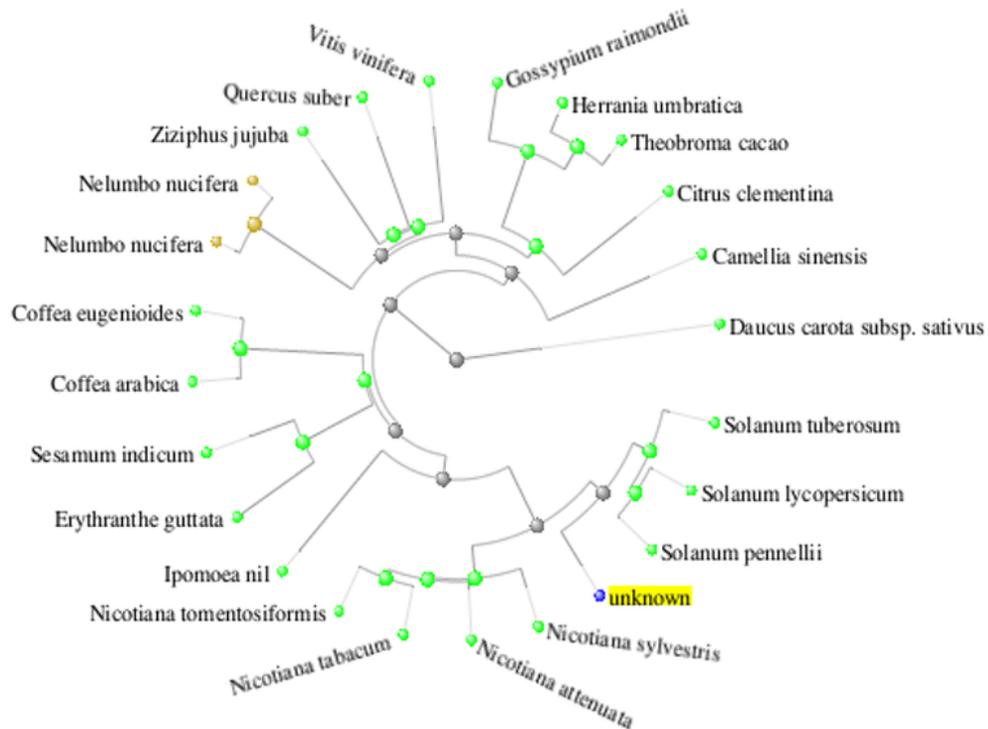


Figure 7

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*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryinformation.docx](#)