

RESEARCH ARTICLE

Comparative Transcriptome Analysis between the Cytoplasmic Male Sterile Line NJCMS1A and Its Maintainer NJCMS1B in Soybean (*Glycine max* (L.) Merr.)

Jiajia Li, Shaohuai Han, Xianlong Ding, Tingting He, Jinying Dai, Shouping Yang*, Junyi Gai*

Soybean Research Institute, National Center for Soybean Improvement, MOA Key Laboratory of Biology and Genetic Improvement of Soybean (General), National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, Jiangsu, People's Republic of China

* spyung@126.com (SPY); sri@njau.edu.cn (JYG)



OPEN ACCESS

Citation: Li J, Han S, Ding X, He T, Dai J, Yang S, et al. (2015) Comparative Transcriptome Analysis between the Cytoplasmic Male Sterile Line NJCMS1A and Its Maintainer NJCMS1B in Soybean (*Glycine max* (L.) Merr.). PLoS ONE 10(5): e0126771. doi:10.1371/journal.pone.0126771

Academic Editor: Zhixi Tian, Chinese Academy of Sciences, CHINA

Received: November 2, 2014

Accepted: April 7, 2015

Published: May 18, 2015

Copyright: © 2015 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All of these RNA-Seq reads were deposited in Sequence Read Archive database (<http://www.ncbi.nlm.nih.gov/Traces/sra/>) under the accession number SRP052011.

Funding: Funding: This work was supported by the National Hightech R & D Program of China (2011AA10A105), the National Transgene Science and Technology Major Program of China (2011ZX08004-005, 2013ZX08004-005, 2014ZX08004-005), the National Key Basic Research Program of China (2011CB109301), and the Program for Changjiang Scholars and Innovative

Abstract

Background

The utilization of soybean heterosis is probably one of the potential approaches in future yield breakthrough as was the situation in rice breeding in China. Cytoplasmic male sterility (CMS) plays an important role in the production of hybrid seeds. However, the molecular mechanism of CMS in soybean remains unclear.

Results

The comparative transcriptome analysis between cytoplasmic male sterile line NJCMS1A and its near-isogenic maintainer NJCMS1B in soybean was conducted using Illumina sequencing technology. A total of 88,643 transcripts were produced in Illumina sequencing. Then 56,044 genes were obtained matching soybean reference genome. Three hundred and sixty five differentially expressed genes (DEGs) between NJCMS1A and NJCMS1B were screened by threshold, among which, 339 down-regulated and 26 up-regulated in NJCMS1A compared to in NJCMS1B. Gene Ontology (GO) annotation showed that 242 DEGs were annotated to 19 functional categories. Clusters of Orthologous Groups of proteins (COG) annotation showed that 265 DEGs were classified into 19 categories. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that 46 DEGs were assigned to 33 metabolic pathways. According to functional and metabolic pathway analysis combined with reported literatures, the relations between some key DEGs and the male sterility of NJCMS1A were discussed. qRT-PCR analysis validated that the gene expression pattern in RNA-Seq was reliable. Finally, enzyme activity assay showed that energy supply was decreased in NJCMS1A compared to in NJCMS1B.

Conclusions

We concluded that the male sterility of NJCMS1A might be related to the disturbed functions and metabolism pathways of some key DEGs, such as DEGs involved in carbohydrate and

Research Team in University (PCSIRT13073). The funders had no role in study design, data collection and analysis, decision or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

energy metabolism, transcription factors, regulation of pollen development, elimination of reactive oxygen species (ROS), cellular signal transduction, and programmed cell death (PCD) etc. Future research will focus on cloning and transgenic function validation of possible candidate genes associated with soybean CMS.

Introduction

Soybean (*Glycine max* (L.) Merr.) is an important source of plant protein and oil. However, low yield is a key factor restricting its development. The utilization of soybean heterosis is probably one of the potential approaches in the future yield breakthrough as was the situation in rice breeding in China. Cytoplasmic male sterility (CMS) plays an important role in the production of hybrid seeds [1]. However, the molecular mechanism of CMS in soybean remains unclear.

The transcriptome is the complete set of transcripts in a cell at a specific developmental stage or physiological condition, which can provide information on gene expression and gene regulation [2]. Transcriptome sequencing (RNA-seq) is a recently developed high-performance and comprehensive method of transcriptome analysis [3, 4]. Transcriptome analysis using RNA-seq technology has allowed for the comparison and analysis of thousands of genes within one experiment [5].

Liu et al. [6] analyzed differentially expressed genes between chili pepper cytoplasmic male sterile line 121A and its near-isogenic line-restorer line 121C at the transcriptional level using Solexa/Illumina technology, and found a group of key genes and significant pathways associated with male sterility. Wei et al. [7] conducted transcriptome analysis of differentially expressed genes in the process of development in wild type and nuclear male sterile cotton anthers using digital gene expression profiles, and illustrated that many key genes involved in anther development showed the opposite gene expression patterns in GMS mutant anthers compared with that of wild type anthers at the same development stage. Yan et al. [8] conducted analysis of genome-wide and high-throughput transcriptome sequencing on young floral buds of *B. napus* CMS line Nsa and its novel restorer line NR1 using Solexa/Illumina techniques, and found a group of candidate genes associated with male sterility. An et al. [9] compared the genomic expression profiles of fertile and sterile young flower buds of *pol*-CMS in *B. napus* by RNA-Seq, and found some unigenes controlling anther development were dramatically down-regulated in sterile buds. However, there is no related report on CMS in soybean so far.

The soybean cytoplasmic male sterile line NJCMS1A was developed through consecutive backcross procedures with the cultivar N8855 as donor parent and N2899 (designated as NJCMS1B afterwards) as recurrent parent [10–12]. So NJCMS1A and NJCMS1B were a pair of near-isogenic lines and fit for the study on the molecular mechanism of CMS in soybean. In the present paper, we tried to find important differentially expressed genes and metabolism pathways might related to the soybean CMS through the comparative transcriptome analysis between the flower buds of NJCMS1A and those of NJCMS1B using the Illumina sequencing technology.

Materials and Methods

Plant Materials

The soybean cytoplasmic male-sterile line NJCMS1A was developed through consecutive backcross procedures with the cultivar N8855 as donor parent and N2899 (designated as NJCMS1B

afterwards) as recurrent parent [10–12]. The genotypes of NJCMS1A and NJCMS1B were designated as S (rr) and N (rr), respectively. Both of NJCMS1A and NJCMS1B had similar nucleus genetic background, but with different cytoplasm, so they are a pair of near-isogenic lines of isonuclear alloplasmic type. NJCMS1A and NJCMS1B were planted in the summer of 2012 and 2014 at Jiangpu Experimental Station, National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, Jiangsu, China. The male-sterile plants were identified through three kinds of methods including the dehiscence of anthers, germination rate of pollens, and performance of plants at maturity. Cytological observation showed that the male abortion of NJCMS1A occurred mainly at the early binucleate pollen stage [13]. So during the flowering period, the flower buds in different sizes before abortion stage were collected and pooled from NJCMS1A and NJCMS1B plants respectively, then immediately frozen in liquid nitrogen and stored at -80 °C for further use. The samples collected in the summer of 2012 were used to RNA-seq and qRT-PCR experiments. The samples collected in the summer of 2014 were used to further qRT-PCR, enzyme activity assay and sugar content analysis.

Total RNA Extraction, cDNA Library Construction and Illumina Deep Sequencing

Total RNA (5 µg) from the flower bud tissue (0.5–0.8 g) of NJCMS1A and NJCMS1B respectively was extracted using the TRIzol kit (Invitrogen, Carlsbad, CA, USA). An Ultra-micro spectrophotometer NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) was used to detect total RNA concentration and purity. Biological analyzer Agilent 2100 (Agilent, Santa Clara, CA, USA) was employed to detect the integrity of RNA. A Truseq RNA Sample Prep Kit (Illumina, San Diego, CA, USA) was employed in mRNA purification and cDNA library construction according to the manufacturer's instructions. The cDNA library was amplified by PCR enrichment, and was examined by 2% electrophoresis agarose gel to recover PCR fragments. TBS380 micro fluorescence (QuantiFluor ST/P, Promega, Madison, WI, USA) was used for the quantification of the cDNA library. Illumina sequencing was conducted on a HiSeq 2000 sequencer (HiSeq 2000 Truseq SBS Kit v3-HS (200 cycles), Illumina). These experiments were completed by Shanghai Majorbio Bio-pharm Biotechnology Co. (<http://www.majorbio.com>, Shanghai, China).

Data Analysis of RNA-Seq

Software SeqPrep (<https://github.com/jstjohn/SeqPrep>) and Condetri_v2.0.pl (http://code.google.com/p/condetri/downloads/detail?Name=condetri_v2.0.pl) were used to filter noises for the original sequencing reads. Sequencing saturation and coverage in the two cDNA libraries were performed by the RSeQC-2.3.2 software [14]. The sequencing adapter sequence, low-quality reads, higher N rate sequences, and too short sequences were removed. The remaining high-quality reads were submitted for mapping analysis against soybean reference genome (ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/early_release/Gmax_275_Wm82.a2.v1/, version Glyma2.0) using Tophat (<http://tophat.cbcb.umd.edu/>) [15], allowing two base mismatches. The mapped reads was then assembled with Cufflinks (<http://cufflinks.cbcb.umd.edu/>) [16].

Differential Expression Analysis

The expression quantity of each gene (fragments per kilobase of exon model per million mapped fragments, FPKM) was estimated by Cuffdiff software [17]. “FDR (False Discovery

Rate) ≤ 0.05 [18, 19] and $|\text{Log}_2\text{FC (Fold Change)}| \geq 1$ ” were used as the threshold for judging the significant of gene expression difference.

Gene Ontology (GO) Annotation, COG Annotation, and KEGG Enrichment Pathway Analysis

Gene Ontology (GO, <http://www.geneontology.org/>) and functional enrichment analysis were conducted on all identified differentially expressed genes (DEGs) using the Goatools software [20] (<https://github.com/tang-haibao/goatools>) ($P \leq 0.05$). Functional classification of Clusters of Orthologous Groups of proteins (COG) was conducted on all identified DEGs using Blastx 2.2.24+ software in the STRING9.0 database (<http://string-db.org/>). Finally, metabolic pathway analysis was performed on all identified DEGs in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/genes.html>) using Blastx/Blastp 2.2.24+ and KOBAS (<http://kobas.cbi.pku.edu.cn/home.do>) [21].

Quantitative Real Time-PCR (qRT-PCR) Analysis

Quantitative real time-PCR (qRT-PCR) analysis was used to verify the RNA-Seq gene expression pattern. Total RNA was extracted using the TRIzol kit (Invitrogen, Carlsbad, CA, USA). Then, cDNA was synthesized by reverse transcription with DNA enzyme purified RNA samples using PrimeScript RT Reagent kits with gDNA Eraser (PrimeScript RT reagent Kit with gDNA Eraser, Takara, Dalian, China) following the manufacturer’s protocols. Gene-specific qRT-PCR primers were designed based on reference unigene sequences with Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, USA), gene-specific primers for qRT-PCR and genes annotation were listed in [S5 Table](#). The mixed solution of qRT-PCR reaction (25 μl) contained SybrGreen qRT-PCR Master Mix (2 \times concentration, Ruian Biotechnologies, Shanghai, China) 12.5 μl , reverse and forward primers (10 μM) 0.5 μl , cDNA 2 μl and ddH₂O 9.5 μl . qRT-PCR was performed in an ABI 7500 FAST Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). PCR conditions were 2 min at 95°C, followed by 40 cycles of heating at 95°C for 10 s and annealing at 60°C for 40 s. The *β -actin* gene was used as the internal control. $2^{-(\Delta\Delta\text{Ct})}$ algorithm was used to calculate the relative level of gene expression, NJCMS1B sample served as the control. The relative level of gene expression greater than 1 was regarded as up-regulated and less than 1 was regarded as down-regulated. All qRT-PCR reactions were performed with three biological replicates.

Enzyme Activity Assay and Sugar Content Analysis

Total ATPase activity was measured at 636 nm by the UV-spectrophotometer (Philes, Nanjing, China, <http://www.philes.cn/>) using the ultramicro total ATPase assay kit (Jiancheng, Nanjing, China, <http://mall.njcbio.com>). One unit of the total ATPase activity was defined as 1 μmol of inorganic phosphate (Pi) generated from ATP decomposed by ATPase in per hour per milligram tissue protein ($\mu\text{mol}\cdot\text{Pi}/\text{mg}\cdot\text{Protein}/\text{hour}$).

Sucrose phosphate synthase (SPS) activity was measured at 290 nm by the UV-spectrophotometer (Philes, Nanjing, China, <http://www.philes.cn/>) using the sucrose phosphate synthase assay kit (Jiancheng, Nanjing, China, <http://mall.njcbio.com>). One unit of the SPS activity was defined as 1 μmol of sucrose generated by converting the substrate required enzyme content in per minute per milligram tissue protein under 37°C condition (U/mg·Protein).

Soluble sugars (glucose, fructose, sucrose) and starch content were measured at 340 nm by the UV-spectrophotometer (Philes, Nanjing, China, <http://www.philes.cn/>) using the glucose-fructose-sucrose assay kit and starch assay kit (BioSenTec, France, <http://www.biosentec.fr/>), respectively. The content of various sugars (g/L) was calculated based on the formulas

according to the instructions in kits. All enzyme activity assay and sugar content analysis experiments were performed with three biological replicates.

Results

Transcriptome Sequencing and Sequence Alignment

In this study, the transcriptome sequencing analysis of flower buds of the cytoplasmic male sterile line NJCMS1A and its near-isogenic maintainer NJCMS1B in soybean was conducted using an Illumina Hiseq 2000 sequencer. The original image data obtained by sequencing base-calling were the original sequence reads. Each read in the Solexa paired-end (PE) sequencing was 101 bp in length. There were 112.10 million reads and a 11.32 Gb original data sets produced during sequencing. After the raw data were trimmed, 57,382,380 clean reads for NJCMS1A sample and 45,599,106 for NJCMS1B sample were obtained (Table 1). All clean reads were matched to the soybean reference genome by Tophat software, allowing two base mismatches [15, 22]. As a result, 53,376,483 mapped reads for NJCMS1A sample and 42,066,351 for NJCMS1B sample were obtained, with an average matching rate of 92.64% (Table 1).

Saturation and Coverage Analysis of Sequencing

To estimate whether the sequencing depth was sufficient for the transcriptome coverage, the sequencing saturation and coverage in the two cDNA libraries were analyzed. Saturation analysis showed that most genes with moderate contents of expression (genes with greater than 3.5 FPKM) became saturated when more than 40% of sequencing reads were aligned (vertical axis numerical approached 1), which indicated that the overall quality of sequencing saturation in the two cDNA libraries was high, and sequencing amount covered the vast majority of expressed genes (Fig 1). Coverage analysis showed that two ends of the sequencing coverage in the two cDNA libraries had no significant peaks, which indicated that sequencing data among the two cDNA libraries was normally distributed (Fig 2).

Analysis of Differentially Expressed Genes (DEGs)

A total of 88,643 transcripts were produced in the Illumina sequencing. Then 56,044 genes were obtained matching the soybean reference genome by Cufflinks software [16] (S1 Table). “FDR ≤ 0.05 and |Log₂FC| ≥ 1” were used as the threshold to screen the DEGs between NJCMS1A and NJCMS1B. It was found that there were 365 DEGs between NJCMS1A and

Table 1. Number of reads sequenced and mapped to soybean genome.

Reads	NJCMS1A	NJCMS1B	Sum
Raw bases (bp)	6,271,412,392	5,050,932,230	11,322,344,622
Raw reads	62,093,194	50,009,232	112,102,426
≥Q20 (%)	94.23	92.90	94.10
Clean bases (bp)	5,630,986,966	4,433,806,292	10,064,793,258
Clean reads	57,382,380	45,599,106	102,981,486
≥Q20 (%)	98.79	98.63	98.81
Mapped reads	53,376,483	42,066,351	95,442,834
Proportion (%)	93.02	92.25	92.64

Note: Sequence length was 2x101 bp, length of each read was 101 bp using double end sequencing.

doi:10.1371/journal.pone.0126771.t001

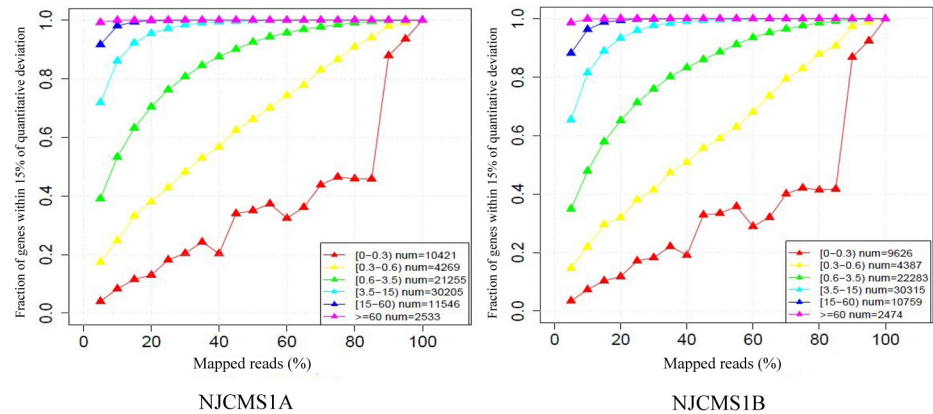


Fig 1. Saturation analysis of sequencing data of NJCMS1A and NJCMS1B. X-axis represented the percentage of mapped reads to soybean genome (%); Y-axis represented the fraction of genes within 15% of quantitative deviation. Each color line represented the saturation curve of different gene expression level, and the gene number within different FPKM interval was displayed in the lower right corner.

doi:10.1371/journal.pone.0126771.g001

NJCMS1B (S2 Table), among which, 339 down-regulated and 26 up-regulated in NJCMS1A compared to in NJCMS1B. Furthermore, 93 down-regulated DEGs were only expressed in NJCMS1B and 9 up-regulated DEGs were uniquely expressed in NJCMS1A. Results showed that the number of the down-regulated DEGs was obviously larger than that of the up-regulated DEGs in NJCMS1A compared to in NJCMS1B. All of these RNA-Seq reads were deposited in Sequence Read Archive database (<http://www.ncbi.nlm.nih.gov/Traces/sra/>) under the Accession number SRP052011.

Gene Ontology (GO) Annotation, COG Annotation and KEGG Enrichment Pathway Analysis

Gene ontology (GO) is an internationally standardized gene function classification system used to describe properties of genes and their products in any organism, containing three ontologies: biological process, cellular component and molecular function [8]. In this study, plant GO Slim annotation was conducted by Blast2GO software (<http://www.blast2go.com/b2ghome>)

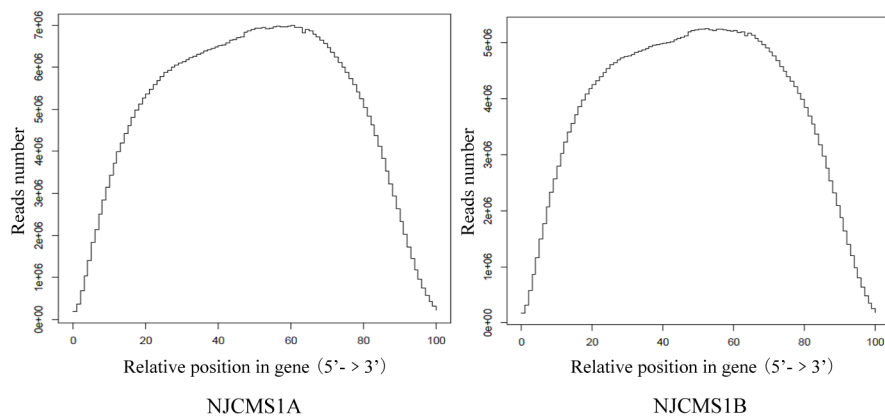


Fig 2. Coverage analysis of sequencing data of NJCMS1A and NJCMS1B. X-axis represented the relative position in gene, 0 and 100 indicated the 5' and 3' end of gene respectively; Y-axis represented the reads number mapped to soybean genome.

doi:10.1371/journal.pone.0126771.g002

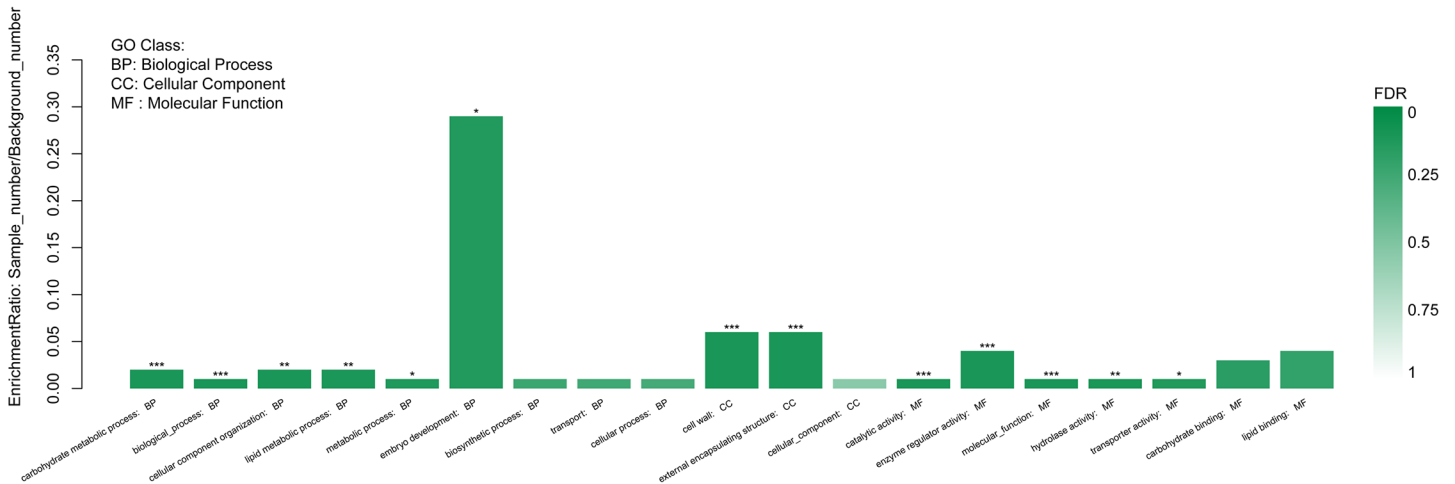


Fig 3. Gene Ontology functional analysis of differentially expressed genes between NJCMS1A and NJCMS1B. X-axis represented each GO term; Y-axis represented the enrichment ratio of genes in each main category.

doi:10.1371/journal.pone.0126771.g003

Version 2.3.5) [23]. Based on sequence homology, 242 DEGs were annotated to 19 functional categories, including 9 biological processes, 3 cellular components and 7 molecular functions (Fig 3, S3 Table). Among the biological process categories, “embryo development” was the main functional groups, followed by “cellular component organization” and “carbohydrate metabolic process”. Among the cellular component categories, “cellular component” was the main functional groups, followed by “external encapsulating structure”. Among the molecular function categories, “enzyme regulator activity” was the main functional groups, followed by “lipid binding” and “carbohydrate binding”.

All detected DEGs were blasted to STRING 9.0 for further annotation based on Cluster of Orthologous Groups (COG) protein categories [24]. A total of 265 DEGs were classified into 19 COG categories (Fig 4, S4 Table), among which, “general function prediction only” represented the largest group (58, 21.9%), followed by “carbohydrate transport and metabolism” (43, 16.2%), and “signal transduction mechanisms” (24, 9.1%). “RNA processing and modification” (2, 0.8%), “Translation, ribosomal structure and biogenesis” (2, 0.8%), and “Defense mechanisms” (2, 0.8%) were the smallest groups.

To identify the metabolic pathways in which the DEGs were involved and enriched, pathway-based analysis was performed using the KEGG pathway database [25]. In total, 46 DEGs were assigned to 33 KEGG pathways (Table 2), among which, “glycolysis/gluconeogenesis” was the most representative pathway (pathway: gmx00010, 8), followed by “carbon fixation in photosynthetic organisms” (pathway: gmx00710, 7), and “oxidative phosphorylation” (pathway: gmx00190, 6). Few DEGs were involved in “RNA transport” (pathway: gmx03013, 1) and “spliceosome” (pathway: gmx03040, 1) etc.

Analysis of DEGs Potentially Related to Male-Sterility in Soybean

Carbohydrate and energy metabolism is one of the most basic metabolic pathways in biological metabolism. Its main physiological function is to provide required energy and carbon sources [26]. In this study, many DEGs were found to involve in carbohydrate and energy metabolism (Table 2), for example, there were 8 DEGs participated in glycolysis/gluconeogenesis pathway, among which, 7 DEGs were down-regulated and 1 DEG was up-regulated in NJCMS1A compared to in NJCMS1B; there were 7 DEGs participated in Carbon fixation in photosynthetic

COG Function Classification

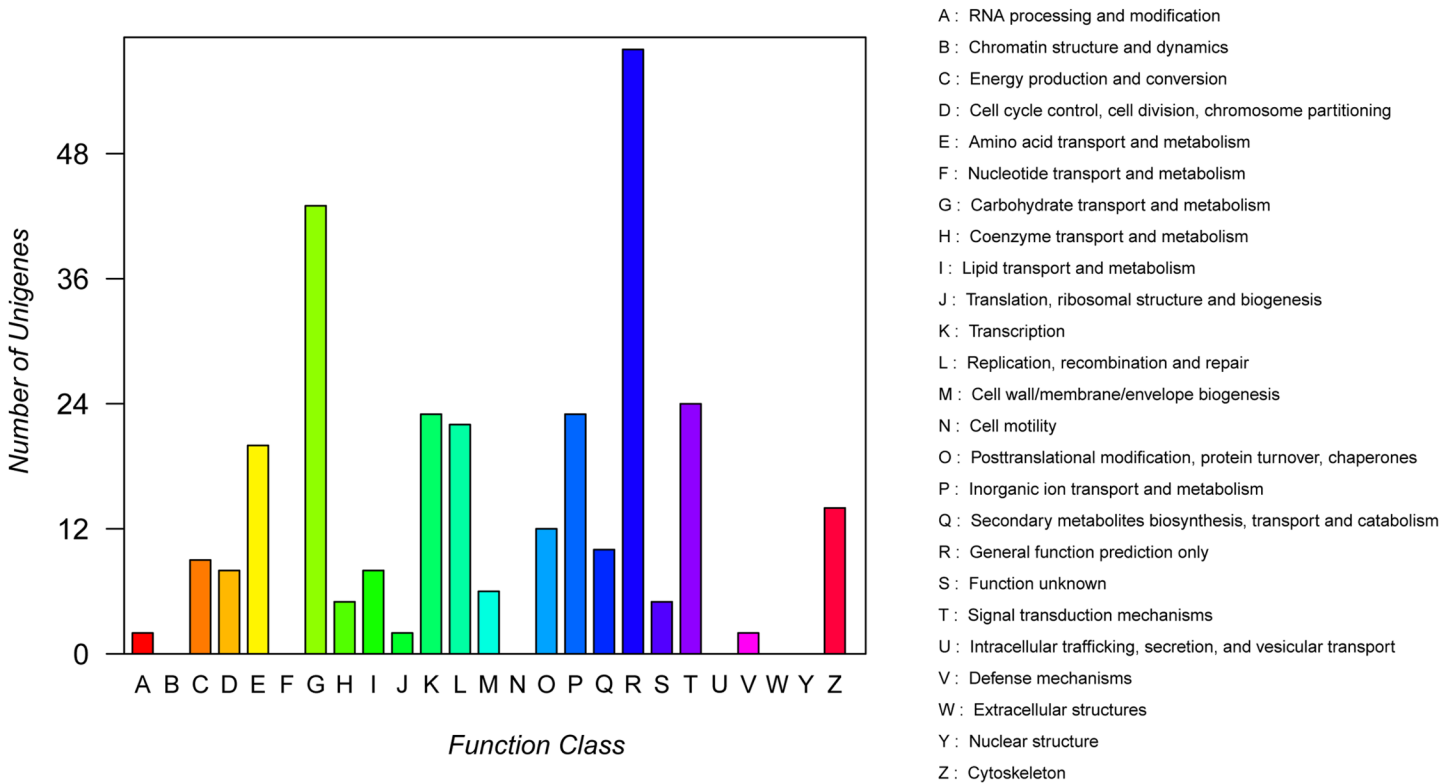


Fig 4. Function classification in Clusters of Orthologous Groups of proteins (COG) of differentially expressed genes between NJCMS1A and NJCMS1B. Capital letters on X-axis indicated the COG categories as listed on the right of the histogram; Y-axis indicated the number of differentially expressed genes.

doi:10.1371/journal.pone.0126771.g004

organisms and all down-regulated in NJCMS1A compared to in NJCMS1B; there were 4 DEGs participated in starch and sucrose metabolism and all down-regulated in NJCMS1A compared to in NJCMS1B. Three DEGs encoding H⁽⁺⁾-ATPase 9, one NADH-ubiquinone oxidoreductase 20 kDa subunit, one vacuolar H⁺-ATPase subunit and one pyrophosphorylase involved in the oxidative phosphorylation and all down-regulated in NJCMS1A compared to in NJCMS1B, and so on (See detail to [Table 2](#)).

Transcription factors are essential for the regulation of gene expression. Changes in gene transcription are associated with changes in expression of transcription factors [27]. Our results showed that there were 15 DEGs encoding transcription factors ([S2 Table](#)), among which, 14 DEGs were down-regulated and 1 DEG was up-regulated in NJCMS1A compared to in NJCMS1B. The 14 down-regulated DEGs were 7 zinc-finger family proteins, 2 WRKY family transcription factors, 1 phytochrome interacting factor 3-like 5, 1 sequence-specific DNA binding transcription factor, 1 MYB domain protein 101, 1 Ras-related small GTP-binding family protein, 1 F-box/RNI-like superfamily protein. One up-regulated DEG was F-box family protein with a domain of unknown function.

In the present study, 38 DEGs were found involved in regulation of pollen development ([S2 Table](#)). There were 34 DEGs participated in the pollen wall development and all down-regulated in NJCMS1A compared to in NJCMS1B, among which, 22 DEGs were might related to cell wall remodeling, including 13 genes encoding plant invertase/pectin methylesterase inhibitor

Table 2. KEGG pathways enriched of differentially expressed gene (DEGs) between NJCMS1A and NJCMS1B (core DEGs).

Gene Id	Gene Annotation	^a Log ₂ FC	^b Regulation	^c P Value
<i>Glycolysis / Gluconeogenesis</i>				
Glyma.10G066700	Aldolase superfamily protein	-2.806566398	down	5.00E-05
Glyma.01G097400	Phosphoenolpyruvate carboxykinase 1	-2.759838312	down	5.00E-05
Glyma.11G247600	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1	1.747097282	up	5.00E-05
Glyma.13G151800	Aldolase superfamily protein	-3.448031374	down	5.00E-05
Glyma.05G073100	Glyceraldehyde-3-phosphate dehydrogenase C2	-1.773605799	down	2.00E-04
Glyma.08G273100	Phosphoenolpyruvate carboxykinase 2	-3.29383008	down	5.00E-05
Glyma.01G118000	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	-4.408790632	down	5.00E-05
Glyma.19G078300	Glyceraldehyde-3-phosphate dehydrogenase C2	-2.910063151	down	5.00E-05
<i>Carbon fixation in photosynthetic organisms</i>				
Glyma.01G091000	Phosphoenolpyruvate carboxylase 4	-1.96777983	down	5.00E-05
Glyma.08G201200	NADP-malic enzyme 4	-3.040729085	down	5.00E-05
Glyma.02G130700	Phosphoenolpyruvate carboxylase 4	-1.854776055	down	5.00E-05
<i>Oxidative phosphorylation</i>				
Glyma.07G025900	H ⁽⁺⁾ -ATPase 9	-1.485862871	down	2.00E-04
Glyma.14G093300	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial	-2.291938094	down	5.00E-05
Glyma.12G040300	Vacuolar H ⁺ -ATPase subunit E isoform 2	-2.813773122	down	5.00E-05
Glyma.07G048300	Pyrophosphorylase 1	-2.009730385	down	5.00E-05
Glyma.15G004300	H ⁽⁺⁾ -ATPase 9	-5.168319998	down	5.00E-05
Glyma.13G369300	H ⁽⁺⁾ -ATPase 9	-4.890112683	down	5.00E-05
<i>Cysteine and methionine metabolism</i>				
Glyma.17G184900	Cobalamin-independent synthase family protein	-4.042843442	down	5.00E-05
Glyma.05G090100	Cobalamin-independent synthase family protein	-2.797929869	down	5.00E-05
Glyma.11G254700	S-adenosyl-L-homocysteine hydrolase	-1.551923295	down	1.00E-04
Glyma.13G141600	Methionine adenosyltransferase 3	-2.981274586	down	5.00E-05
Glyma.10G054500	Methionine adenosyltransferase 3	-2.914574198	down	5.00E-05
<i>Starch and sucrose metabolism</i>				
Glyma.19G195400	Cell wall invertase 2	-2.346102727	down	5.00E-05
Glyma.08G308600	Sucrose phosphate synthase 3F	-1.784517228	down	1.00E-04
Glyma.03G223400	Plant invertase/pectin methylesterase inhibitor superfamily	-6.481838357	down	5.00E-05
Glyma.02G008300	VANGUARD 1 homolog 2	-3.847662291	down	5.00E-05
<i>Pentose and glucuronate interconversions</i>				
Glyma.19G020200	Pectate lyase family protein	-5.637503879	down	5.00E-05
Glyma.08G296900	Pectate lyase family protein	-5.903267116	down	5.00E-05
<i>Nitrogen metabolism</i>				
Glyma.02G035000	Carbonic anhydrase 2	-1.358114642	down	2.00E-04
Glyma.17G148000	Glutamate dehydrogenase 2	-1.532668021	down	2.00E-04
<i>Alanine, aspartate and glutamate metabolism</i>				
Glyma.02G241400	Glutamate decarboxylase	-2.026581623	down	5.00E-05
Glyma.14G211100	Glutamate decarboxylase	-3.278453751	down	5.00E-05
<i>Phagosome</i>				
Glyma.05G126100	Tubulin beta-7 chain	-1.482128236	down	2.00E-04
Glyma.01G197500	Tubulin alpha-2 chain	-1.512186909	down	5.00E-05
<i>Sphingolipid metabolism</i>				
Glyma.15G072800	Fatty acid desaturase family protein	-3.078035515	down	5.00E-05
<i>Stilbenoid, diarylheptanoid and gingerol biosynthesis</i>				
Glyma.08G089500	Cytochrome P450, family 81, subfamily D, polypeptide 3	0.532775756	up	5.00E-05

(Continued)

Table 2. (Continued)

Gene Id	Gene Annotation	^a Log ₂ FC	^b Regulation	^c P Value
Biosynthesis of unsaturated fatty acids				
Glyma.14G121400	Plant stearoyl-acyl-carrier-protein desaturase family protein	-2.613215404	down	5.00E-05
Ascorbate and aldarate metabolism				
Glyma.07G225400	SKU5 similar 12	-3.417857252	down	5.00E-05
Glutathione metabolism				
Glyma.02G154400	Glutathione S-transferase 7	-3.023522601	down	5.00E-05
Endocytosis				
Glyma.18G289600	Heat shock cognate protein 70–1	-2.185713321	down	5.00E-05
Phenylalanine metabolism				
Glyma.08G292100	Peroxidase superfamily protein	-1.744501273	down	5.00E-05
RNA transport				
Glyma.10G212900	GTP binding Elongation factor Tu family protein	1.750839666	up	5.00E-05

^aLog₂FC ≥ 1 represented up-regulated and Log₂FC ≤ -1 represented down-regulated.

^bRegulation direction of DEGs (NJCMS1B was the control).

^cP value of ≤ 0.05 was considered statistically significant.

doi:10.1371/journal.pone.0126771.t002

superfamily proteins, 3 genes encoding pectin lyase-like superfamily proteins, 2 genes encoding pectate lyase family proteins, 1 gene encoding cell wall invertase 2, 1 gene encoding cellulose synthase like D4, 1 gene encoding callose synthase 5, 1 gene encoding hexokinase-like 3; the other 12 DEGs were might related to cytoskeletal structures, including 3 genes encoding myosin family proteins, 3 genes encoding profiling, 2 genes encoding tubulin, 2 genes encoding actin-11, 1 gene encoding myosin 2, and 1 gene encoding actin depolymerizing factor. In addition, there were 4 DEGs encoding pollen Ole e 1 allergen and extensin family proteins and all down-regulated in NJCMS1A compared to in NJCMS1B.

In this study, we also found other DEGs potentially related to male-sterility in soybean. There were 17 DEGs related to elimination of reactive oxygen species (ROS) (S2 Table), among which, 15 DEGs were down-regulated and 2 DEGs were up-regulated in NJCMS1A compared to in NJCMS1B. Fifteen down-regulated DEGs were 13 genes encoding late embryogenesis abundant protein (LEA) family proteins and 2 genes encoding peroxidase superfamily proteins. Two up-regulated DEGs were 1 gene encoding LEA family protein and 1 gene encoding alternative oxidase. In addition, 12 DEGs associated with calmodulin-like were found and all down-regulated in NJCMS1A compared to in NJCMS1B (S2 Table). They were 3 genes encoding calcium-binding EF-hand family proteins, 4 genes encoding calcium-dependent lipid-binding (CaLB domain) family proteins and 5 genes encoding calcium-dependent protein kinases. Notably, several DEGs with known or unknown function were also found in our results (S2 Table).

Analysis of DEGs by qRT-PCR

According to the functional and metabolic pathway analysis combined with previously reported literatures, 21 DEGs might related to CMS were chosen to be conducted qRT-PCR analysis using the same sample as that in RNA-seq. It was found that the expression patterns of qRT-PCR of 18 DEGs were consistent with those of RNA-Seq, while the other 3 DEGs were not (Fig 5, S5-1 Table). The coincidence rate between qRT-PCR results and RNA-Seq results was 85.71%. In addition, 5 DEGs might related to CMS were chosen to be conducted qRT-PCR analysis using different sample from that in RNA-seq (Fig 6, S5-2 Table). The results showed

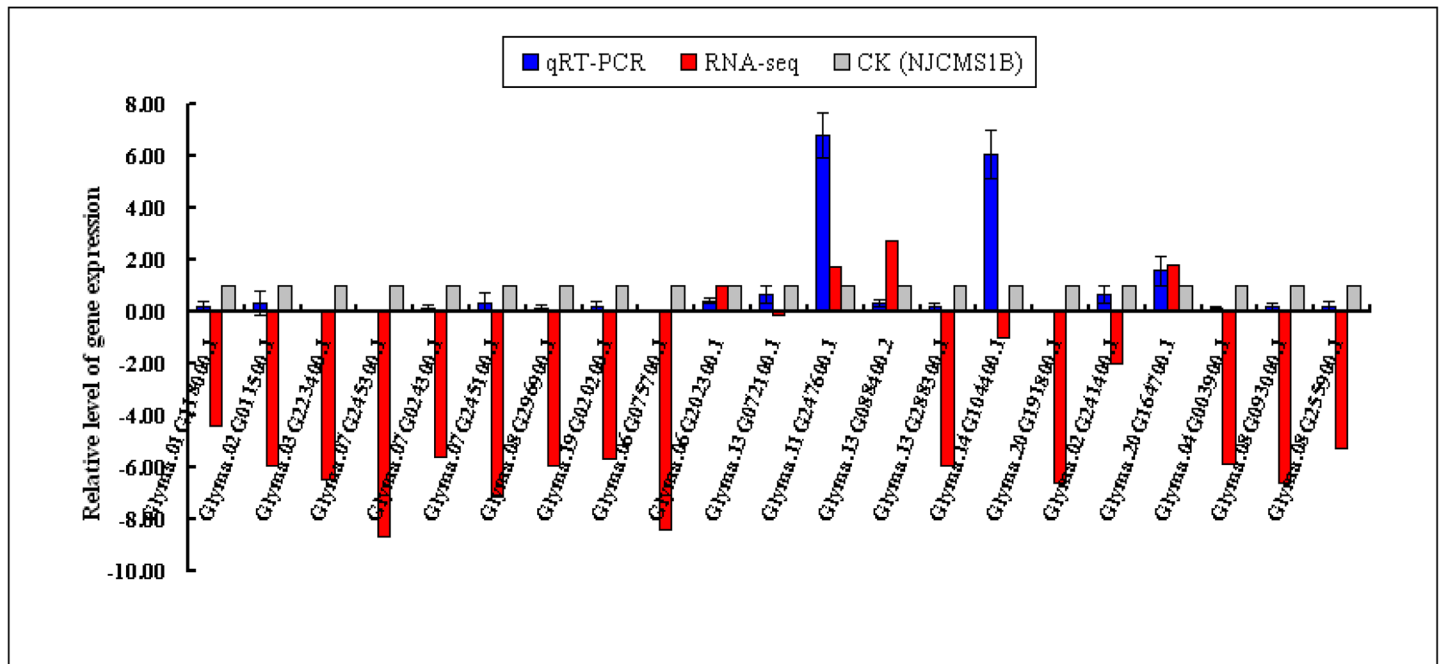


Fig 5. DEGs confirmed by qRT-PCR using the same sample as that in RNA-seq. X-axis represented gene name, the blue column represented qRT-PCR results, the red column represented RNA-seq results, and gray column represented CK (NJCMS1B); Y-axis represented the relative level of gene expression. Gene-specific qRT-PCR primers and gene name were listed in [S5-1 Table](#). All qRT-PCR reactions were performed with three biological replicates.

doi:10.1371/journal.pone.0126771.g005

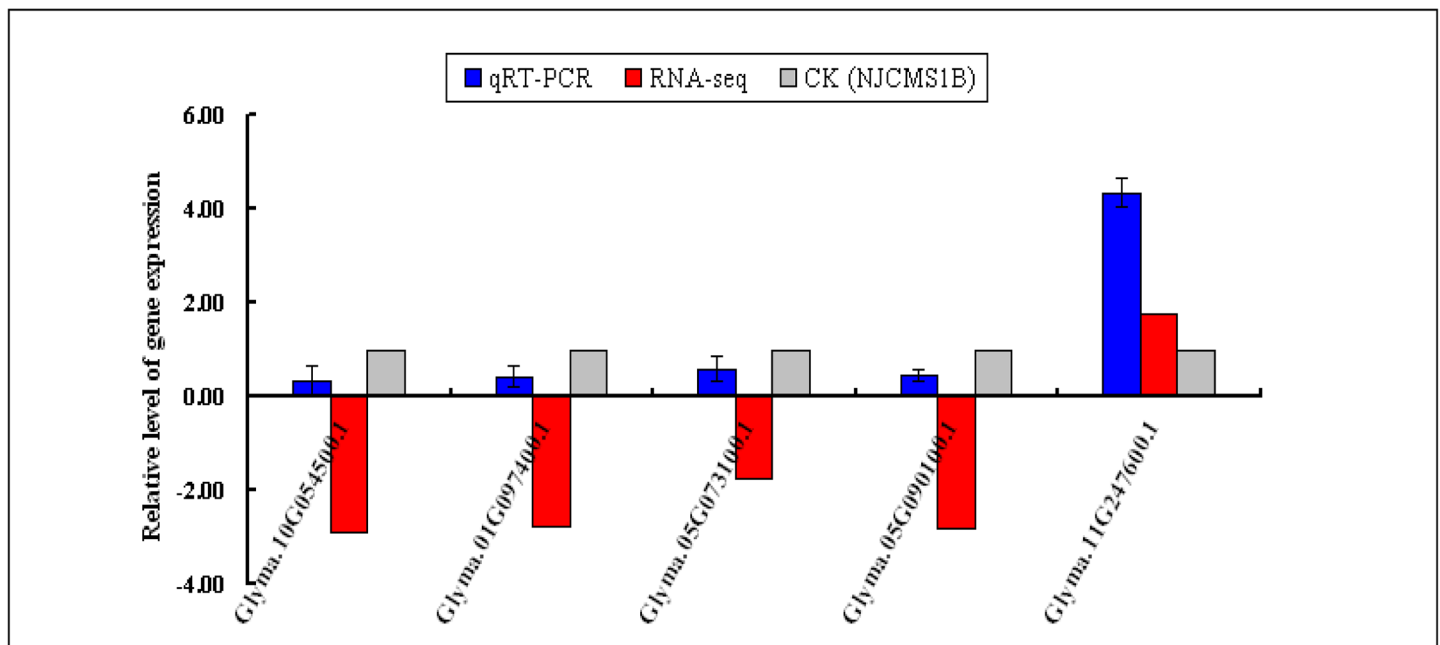


Fig 6. DEGs confirmed by qRT-PCR using different sample from that in RNA-seq. X-axis represented gene name, the blue column represented qRT-PCR results, the red column represented RNA-seq results, and gray column represented CK (NJCMS1B); Y-axis represented the relative level of gene expression. Gene-specific qRT-PCR primers and gene name were listed in [S5-2 Table](#). All qRT-PCR reactions were performed with three biological replicates.

doi:10.1371/journal.pone.0126771.g006

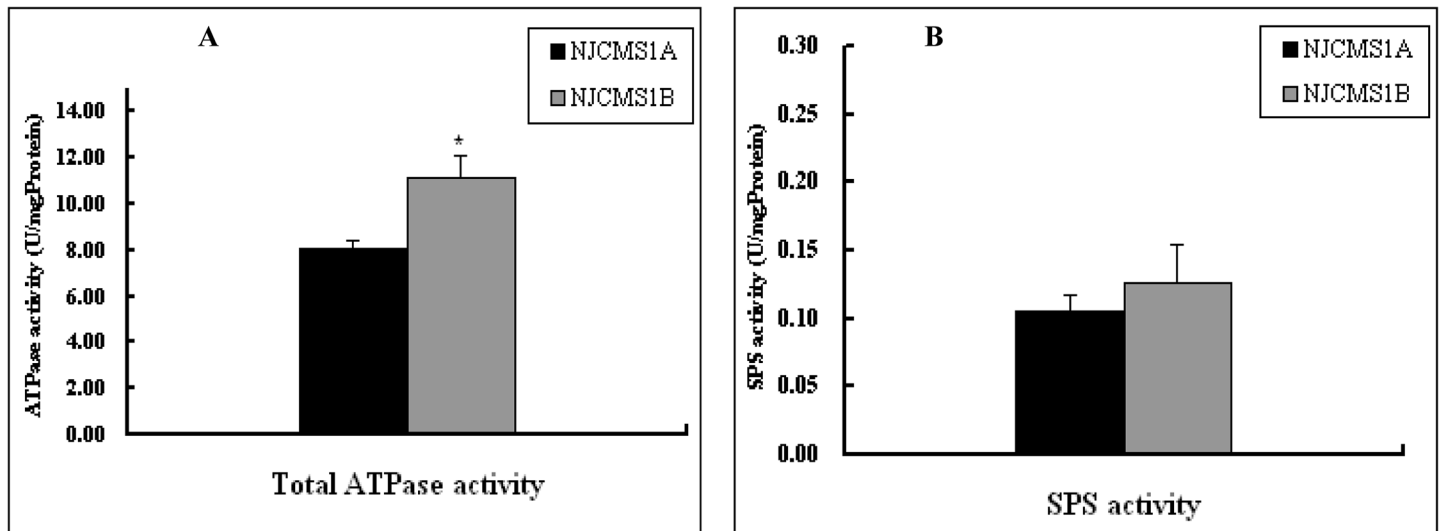


Fig 7. Enzyme activity assay in NJCMS1A and NJCMS1B. The black column represented NJCMS1A and gray column represented NJCMS1B on X-axis; Y-axis represented the enzyme activity. (A) Total ATPase activity and (B) Sucrose phosphate synthase (SPS) activity. The data were given as Mean \pm SD from three biological replicates.

doi:10.1371/journal.pone.0126771.g007

that the expression patterns of qRT-PCR of 5 DEGs were consistent with those of RNA-Seq. These results indicated that the RNA-Seq results in the present study were reliable.

Enzyme Activity Assay and Sugar Content Analysis

The ATPase, a key enzyme for the synthesis of ATP in cellular biosynthesis, plays an important role in material transport, energy transformation and information transmission [28]. Sucrose phosphate synthase (SPS) plays a key role for sucrose synthesis in photosynthetic organs of green plants. To investigate the enzymes activity associated with energy metabolism, we tested the total ATPase activity and SPS activity in flower buds of NJCMS1A and NJCMS1B. The results showed that the total ATPase activity was significantly decreased in male-sterile line NJCMS1A, relative to maintainer NJCMS1B (Fig 7), but, there was not significant difference of SPS activity between NJCMS1A and NJCMS1B (Fig 7), this might be caused by sample tissues and needed to be further studied.

To improve our understanding of the basis of energy deficiency in the male-sterile line, we tested the content of sugars, including glucose, fructose, sucrose and starch. The results showed that there were not significant difference of soluble sugars (glucose, fructose, sucrose) and starch content between NJCMS1A and NJCMS1B (Fig 8), this might be caused by sample tissues and needed to be further studied.

Discussion

In the present study, the comparative transcriptome analysis between the cytoplasmic male sterile line NJCMS1A and its near-isogenic maintainer NJCMS1B in soybean was conducted using the Illumina sequencing technology. Three hundred and sixty five DEGs were screened between NJCMS1A and NJCMS1B by threshold, among which, 339 down-regulated and 26 up-regulated in NJCMS1A compared to in NJCMS1B. According to GO, COG and KEGG functional and metabolic pathway analysis combined with previously reported literatures, the relations between some key DEGs and the male sterility of NJCMS1A would be discussed as follows.

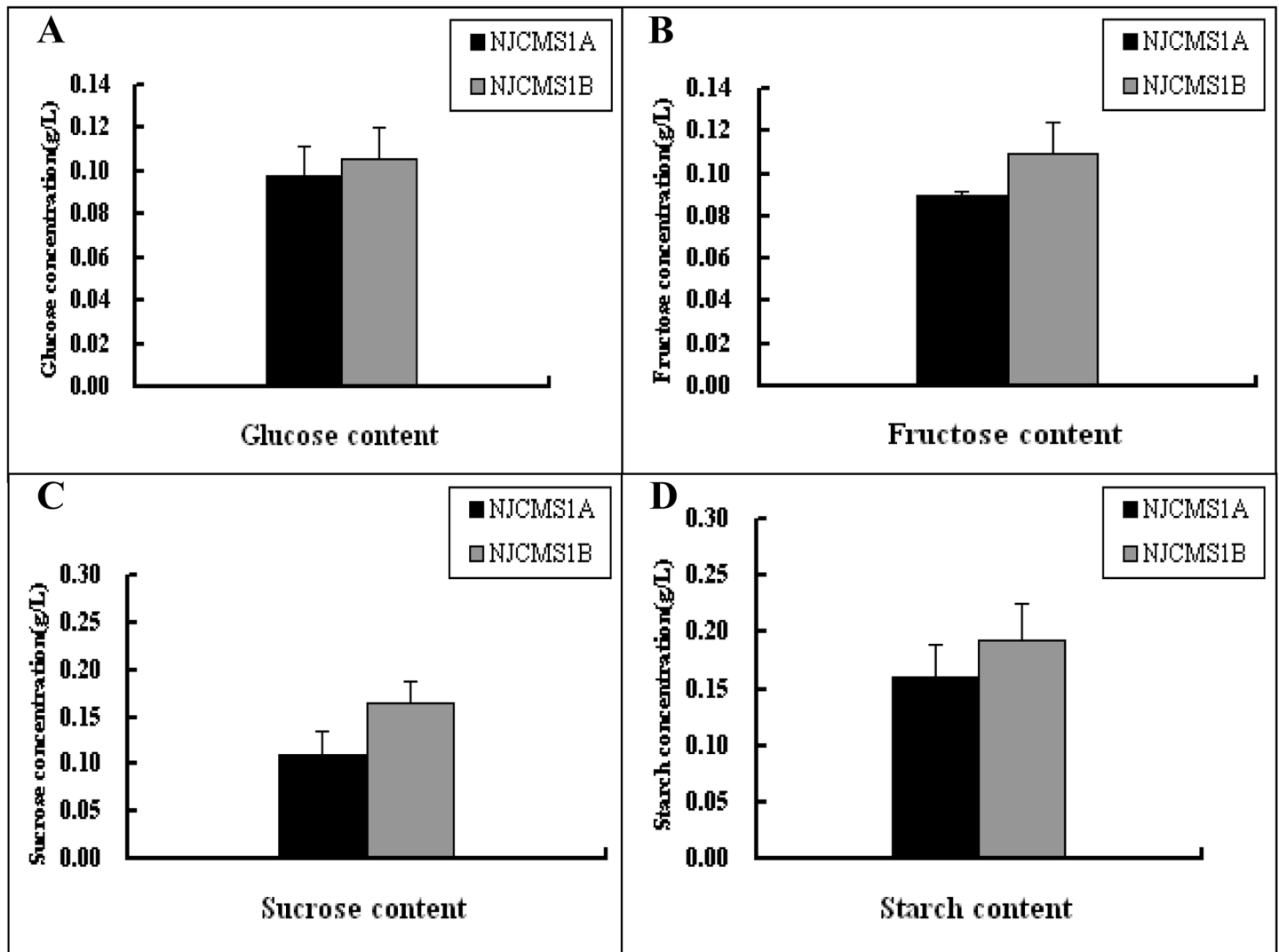


Fig 8. Sugar content analysis in NJCMS1A and NJCMS1B. The black column represented NJCMS1A and gray column represented NJCMS1B on X-axis; Y-axis represented sugar concentration. (A) Glucose; (B) Fructose; (C) Sucrose and (D) Starch. The data were given as Mean \pm SD from three biological replicates.

doi:10.1371/journal.pone.0126771.g008

Analysis of DEGs Involved in Carbohydrate and Energy Metabolism Potentially Related to CMS in Soybeans

The development of stamen and pollen in flowering plants is a complicated process that involves a series of well coordinated cytoplasmic and nuclear gene interactions leading to multifarious metabolic processes and structural changes [29]. Dorion et al. [30] stated that a dysfunction in a major metabolic pathway, such as sugar metabolism, could adversely affect the development of pollen grain. It has been suggested that a high respiration rate and great energy demand are usually observed during pollen development [31]. Carbohydrate not only provides nutrition for anther development, but also affects anther and pollen development as a signal substance [26]. Moreover, Bergman et al. [32] and Teixeira et al. [33] had proved that the ATP content related to supply of energy in male sterile lines was significantly decreased. In this study, many DEGs were found involved in carbohydrate and energy metabolism pathway

(Table 2), and were down-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B. Meanwhile, enzyme activity assay showed total ATPase activity was significantly decreased in NJCMS1A, relative to NJCMS1B. These results showed that the expression of genes related to the supply of energy in the male sterile line NJCMS1A was suppressed, which might result in a shortage in energy required for pollen maturation and the abnormal development of pollen, and ultimately led to the male sterility of NJCMS1A.

Analysis of DEGs Encoding Transcription Factors Potentially Related to CMS in Soybeans

Hao et al. [27] indicated that a single transcription factor could regulate the expressions of multiple genes in a metabolic pathway. Yan et al. [8] determined that transcription factors were essential for the regulation of plant gene expression, and changes in gene transcription were related to changes in the expression of transcription factors. Alteration in the expression of transcription factor genes normally results in dramatic changes during plant growth [34, 35]. In this study, 14 and 1 DEG encoding transcription factors were found down-regulated and up-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B respectively. This might lead to the expression of genes related to the development of flower organ in NJCMS1A was interfered, possibly resulted in male sterility of NJCMS1A.

Analysis of DEGs Involved in Regulation of Pollen Development Potentially Related to CMS in Soybeans

Pollen development is an essential process of sexual reproduction in flowering plants. Zhu et al. [36] stated that pollen cell wall development in pollen grains ensured plant sexual reproduction, and the majority of male sterile traits were associated with abnormal wall development. Zhang et al. [37] determined that *BoPME11* expression was suppressed and resulted in the retardation of pollen development and partial male sterility in the antisense expression studies of *BoPME11* in *Arabidopsis thaliana*. Hideaki et al. [38] showed that genes involved in the cytoskeleton category played key roles in cell wall expansion. Li et al. [39] determined that the over-expression of actin depolymerizing factor (*GhADF7*) in *Gossypium hirsutum* L. could alter the balance of actin depolymerization and polymerization, resulting in incomplete cytokinesis and partial male sterility. In addition, recent studies have suggested that the pollen Ole e 1 allergen and extensin family protein functioned as developmental regulators in many plant tissues [40–42]. In this study, 38 DEGs were found related to the pollen development and all down-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B, among which, 34 DEGs participated in the pollen wall development and 4 DEGs functioned as developmental regulators in different plant tissues. The above results showed that the abnormal expression of genes involved in the pollen development in NJCMS1A might directly influence the pollen development of NJCMS1A including cell wall remodeling, aberrant cytoskeletal structures, and lose of functions as developmental regulators in pollen etc., and ultimately resulted in the male sterility of NJCMS1A.

Analysis of DEGs Involved in Elimination of ROS and Cellular Signal Transduction Potentially Related to CMS in Soybeans

It has been suggested that an abnormality of activated oxygen metabolism in the development of the anther or young panicle might be related to male sterility [43–45]. Li et al. [43] showed that a higher concentration of ROS and mitochondrial damage existed in the microspores of male sterile rice lines. Liu et al. [45] demonstrated that an important characteristic of LEA,

which differed from that of other molecular chaperone protein functions, was that they could eliminate active oxygen and protect cell membrane stability. In the present study, 15 DEGs related to elimination of ROS were found down-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B. The results showed that the expression of a variety of active oxygen scavenging enzyme genes was inhibited in NJCMS1A, resulting in higher concentrations of ROS in NJCMS1A than in NJCMS1B, which might be a possible reason for the occurrence of male sterility of NJCMS1A.

Several studies had demonstrated that, Ca^{2+} , a messenger in cellular signal transduction, functioned as a pivotal regulator of the cell life cycle including cell division, differentiation, and apoptosis [46–49]. Rato et al. [50] showed that pollen development depended on multiple signaling pathways, in which calmodulin was a key element. In this study, 12 DEGs associated with calmodulin-like were identified and all down-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B. The results indicated that their differential expression might cause destruction of calcium signaling pathways and abnormal pollen development, resulting in male sterility of NJCMS1A.

Analysis of Other DEGs Potentially Related to CMS in Soybeans

It was suggested that aspartic protease acted as an anti-cell-death factor participating in programmed cell death (PCD) and the over-expression of this gene resulted in male sterility in *Arabidopsis* [51]. In this study, we found 1 gene encoding aspartic proteinase A1 up-regulated in NJCMS1A, the higher expression level of this gene in the male sterile line NJCMS1A could lead to PCD of the pollen cell, ultimately causing male sterility. In addition, other DEGs were also found in our study, for example, there were 2 genes encoding cytochrome P450 family protein, 1 gene encoding glutathione S-transferase tau 9, 1 gene encoding leucine-rich repeat transmembrane protein kinase and 1 gene encoding glyceraldehyde-3-phosphate dehydrogenase C subunit 1, etc., which were up-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B; seven genes encoding major facilitator superfamily proteins, 6 genes encoding NAD(P)-binding Rossmann-fold superfamily proteins and 3 genes encoding tetratricopeptide repeat (TPR)-containing proteins, etc., which were down-regulated in the male sterile line NJCMS1A compared to in the near-isogenic maintainer NJCMS1B; and 20 DEGs with unknown functions. The above DEGs might be associated with the male sterility of NJCMS1A, but their specific functions needed to be further studied.

Conclusions

In the present study, the comparative transcriptome analysis between the cytoplasmic male sterile line NJCMS1A and its near-isogenic maintainer NJCMS1B in soybean was conducted. The results showed that there were 365 DEGs between NJCMS1A and NJCMS1B, among which, 339 down-regulated and 26 up-regulated in NJCMS1A compared to in NJCMS1B. According to GO, COG and KEGG functional and metabolic pathway analysis combined with the previously reported literatures, we concluded that the male sterility of NJCMS1A might be related to the disturbed functions and metabolism pathways of some key DEGs, such as DEGs involved in carbohydrate and energy metabolism, encoding transcription factors, regulation of pollen development, elimination of ROS, cellular signal transduction, and PCD etc. These results will help to elucidate the molecular mechanism of CMS in soybean, and provides a theoretical basis for better utilization of soybean heterosis. Future research will focus on the cloning and transgenic function validation of possible candidate genes associated with soybean CMS.

Supporting Information

S1 Table. Total number of identified genes between NJCMS1A and NJCMS1B.

(XLS)

S2 Table. Number of differentially expressed genes between NJCMS1A and NJCMS1B.

(XLS)

S3 Table. Gene Ontology functional annotation of differentially expressed genes between NJCMS1A and NJCMS1B.

(XLS)

S4 Table. Clusters of Orthologous Groups of proteins classification of differentially expressed genes between NJCMS1A and NJCMS1B.

(XLS)

S5 Table. Comparison of expression patterns between RNA-seq and qRT-PCR. S5-1 Table.

DEGs confirmed by qRT-PCR using the same sample as that in RNA-seq. S5-2 Table. DEGs

confirmed by qRT-PCR using different sample from that in RNA-seq.

(XLS)

Acknowledgments

We thank the Shanghai Majorbio Bio-pharm Biotechnology Company (Shanghai, China) for analysis of transcriptome sequencing data.

Author Contributions

Conceived and designed the experiments: SPY JYG JJJ. Performed the experiments: JJJ SHH XLD TTH JYD. Analyzed the data: JJJ. Contributed reagents/materials/analysis tools: JJJ SHH. Wrote the paper: JJJ SPY.

References

1. Havey MJ. The use of cytoplasmic male sterility for hybrid seed production. In: Daniell H, Chase CD, editors. *Molecular Biology and Biotechnology of Plant Organelles*. Springer Netherlands; 2004. pp. 623–634.
2. Wei WL, Qi XQ, Wang LH, Zhang YX, Hua W, Li D, et al. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics*. 2011; 12: 451. doi: [10.1186/1471-2164-12-451](https://doi.org/10.1186/1471-2164-12-451) PMID: [21929789](https://pubmed.ncbi.nlm.nih.gov/21929789/)
3. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods*. 2008; 5: 16–18. doi: [10.1038/nmeth1156](https://doi.org/10.1038/nmeth1156) PMID: [18165802](https://pubmed.ncbi.nlm.nih.gov/18165802/)
4. Ansorge WJ. Next-generation DNA sequencing techniques. *N Biotechnol*. 2009; 25: 195–203. doi: [10.1016/j.nbt.2008.12.009](https://doi.org/10.1016/j.nbt.2008.12.009) PMID: [19429539](https://pubmed.ncbi.nlm.nih.gov/19429539/)
5. Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-seq. *Plant Physiol*. 2010; 152: 1787–1795. doi: [10.1104/pp.109.149716](https://doi.org/10.1104/pp.109.149716) PMID: [20118272](https://pubmed.ncbi.nlm.nih.gov/20118272/)
6. Liu C, Ma N, Wang PY, Fu N, Shen HL. Transcriptome sequencing and de novo analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). *PLoS One*. 2013; 8: e65209. doi: [10.1371/journal.pone.0065209](https://doi.org/10.1371/journal.pone.0065209) PMID: [23750245](https://pubmed.ncbi.nlm.nih.gov/23750245/)
7. Wei MM, Song MZ, Fan SL, Yu SX. Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. *BMC Genomics*. 2013; 14: 97. doi: [10.1186/1471-2164-14-97](https://doi.org/10.1186/1471-2164-14-97) PMID: [23402279](https://pubmed.ncbi.nlm.nih.gov/23402279/)
8. Yan XH, Dong CH, Yu JY, Liu WH, Jiang CH, Liu J, et al. Transcriptome profile analysis of young floral buds of fertile and sterile plants from the self-pollinated offspring of the hybrid between novel restorer line NR1 and Nsa CMS line in *Brassica napus*. *BMC Genomics*. 2012; 14: 26.

9. An H, Yang ZH, Yi B, Wen J, Shen JX, Tu JX, et al. Comparative transcript profiling of the fertile and sterile flower buds of pol CMS in *B.napus*. BMC Genomics. 2014; 15: 258. doi: [10.1186/1471-2164-15-258](https://doi.org/10.1186/1471-2164-15-258) PMID: [24707970](https://pubmed.ncbi.nlm.nih.gov/24707970/)
10. Gai JY, Cui ZL, Ji DF, Ren ZJ, Ding DR. A report on the nuclear cytoplasmic male sterility from a cross between two soybean cultivars. Soy Genet Newsl. 1995; 22: 55–58.
11. Ding DR, Gai JY, Cui ZL, Yang SP, Qiu JX. Development and verification of the cytoplasmic-nuclear male sterile soybean line NJCMS1A and its maintainer NJCMS1B. Chinese Sci Bull. 1999; 44: 191–192.
12. Ding DR, Gai JY, Cui ZL, Qiu JX. Development of a cytoplasmic-nuclear male-sterile line of soybean. Euphytica. 2002; 124: 85–91.
13. Fan JM. Studies on cyto-morphological and cyto-chemical features of cytoplasmic-nuclear male-sterile lines of soybeans (*Glycine max* (L.) Merr.). M. Sc. Thesis, Nanjing Agricultural University. 2003. <http://book.hzu.edu.cn/629383.html>.
14. Wang LG, Wang SQ, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012; 28: 2184–2185. doi: [10.1093/bioinformatics/bts356](https://doi.org/10.1093/bioinformatics/bts356) PMID: [22743226](https://pubmed.ncbi.nlm.nih.gov/22743226/)
15. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25: 1105–1111. doi: [10.1093/bioinformatics/btp120](https://doi.org/10.1093/bioinformatics/btp120) PMID: [19289445](https://pubmed.ncbi.nlm.nih.gov/19289445/)
16. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28: 511–515. doi: [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621) PMID: [20436464](https://pubmed.ncbi.nlm.nih.gov/20436464/)
17. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol. 2013; 31: 46–53. doi: [10.1038/nbt.2450](https://doi.org/10.1038/nbt.2450) PMID: [23222703](https://pubmed.ncbi.nlm.nih.gov/23222703/)
18. Benjamini BY, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001; 29: 1165–1188.
19. Benjamini BY, Hochberg Y. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. J R Statist Soc. 1995; 57: 289–300.
20. Tang HB, Wang XY, Bowers JE, Ming R, Alam M, Paterson AH. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res. 2008; 18: 1944–1954. doi: [10.1101/gr.080978.108](https://doi.org/10.1101/gr.080978.108) PMID: [18832442](https://pubmed.ncbi.nlm.nih.gov/18832442/)
21. Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011; 39: W316–W322. doi: [10.1093/nar/gkr483](https://doi.org/10.1093/nar/gkr483) PMID: [21715386](https://pubmed.ncbi.nlm.nih.gov/21715386/)
22. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7: 562–578. doi: [10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016) PMID: [22383036](https://pubmed.ncbi.nlm.nih.gov/22383036/)
23. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21: 3674–3676. PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/)
24. Wang Y, Tao X, Tang XM, Xiao L, Sun JL, Yan XF, et al. Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. BMC Genomics. 2013; 14: 841. doi: [10.1186/1471-2164-14-841](https://doi.org/10.1186/1471-2164-14-841) PMID: [24289302](https://pubmed.ncbi.nlm.nih.gov/24289302/)
25. Zhai RR, Feng Y, Wang HM, Zhan XD, Shen XH, Wu WM, et al. Transcriptome analysis of rice root heterosis by RNA-Seq. BMC Genomics. 2013; 14: 19. doi: [10.1186/1471-2164-14-19](https://doi.org/10.1186/1471-2164-14-19) PMID: [23324257](https://pubmed.ncbi.nlm.nih.gov/23324257/)
26. Wu ZM, Cheng JW, Qin C, Hu ZQ, Yin CX, Hu KL. Differential proteomic analysis of anthers between cytoplasmic male sterile and maintainer lines in *Capsicum annuum* L. Int J Mol Sci. 2013; 14: 22982–22996. doi: [10.3390/ijms141122982](https://doi.org/10.3390/ijms141122982) PMID: [24264042](https://pubmed.ncbi.nlm.nih.gov/24264042/)
27. Hao QN, Zhou XA, Sha AH, Wang C, Zhou R, Chen SL. Identification of genes associated with nitrogen use efficiency by genome-wide transcriptional analysis of two soybean genotypes. BMC Genomics. 2011; 12: 525. doi: [10.1186/1471-2164-12-525](https://doi.org/10.1186/1471-2164-12-525) PMID: [22029603](https://pubmed.ncbi.nlm.nih.gov/22029603/)
28. Siedow JN, Umbach AL. Plant mitochondrial electron transfer and molecular biology. Plant Cell. 1995; 7: 821–831. PMID: [12242388](https://pubmed.ncbi.nlm.nih.gov/12242388/)
29. Sheoran IS, Sawhney VK. Proteome analysis of the normal and Ogura (ogu) CMS anthers of *Brassica napus* to identify proteins associated with male sterility. Botany. 2010; 88: 217–230.
30. Dorion S, Lalonde S, Saini HS. Induction of male sterility in wheat by meiotic-stage water deficit is preceded by a decline in invertase activity and changes in carbohydrate metabolism in anthers. Plant Physiol. 1996; 111: 137–145. PMID: [12226280](https://pubmed.ncbi.nlm.nih.gov/12226280/)

31. Tadege M, Kuhlemeier C. Aerobic fermentation during tobacco pollen development. *Plant Mol Biol*. 1997; 35: 343–354. PMID: [9349258](#)
32. Bergman P, Edqvist J, Farbos I, Glimelius K. Male-sterile tobacco displays abnormal mitochondrial *atp1* transcript accumulation and reduced floral ATP/ADP ratio. *Plant Mol Biol*. 2000; 42: 531–544. PMID: [10798621](#)
33. Teixeira RT, Knorpp C, Glimelius K. Modified sucrose, starch, and ATP levels in two alloplasmic male-sterile lines of *B.napus*. *J Exp Bot*. 2005; 56: 1245–1253. PMID: [15753110](#)
34. Kater MM, Colombo L, Franken J, Busscher M, Masiero S, Van Lookeren Campagne MM, et al. Multiple *AGAMOUS* homologs from cucumber and petunia differ in their ability to induce reproductive organ fate. *Plant Cell*. 1998; 10: 171–182. PMID: [9490741](#)
35. Grotewold E, Chamberlin M, Snook M, Siame B, Butler L, Swenson J, et al. Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. *Plant Cell*. 1998; 10: 721–740. PMID: [9596632](#)
36. Zhu J, Yang ZN. The research progress of pollen wall development. *Chin J Nat*. 2013; 35 (2). doi: [10.3969/j.issn.0253-9608.2013.02.005](#)
37. Zhang GY, Feng J, Wu J, Wang XW. *BoPME11*, a pollen specific pectin methylesterase inhibitor, has an essential role in pollen tube growth. *Planta*. 2010; 231: 1323–1334. doi: [10.1007/s00425-010-1136-7](#) PMID: [20229192](#)
38. Hideaki S, Laura RU, Xu JN, Zhang JF. Transcriptome analysis of cytoplasmic male sterility and restoration in CMS-D8 cotton. *Plant Cell Rep*. 2013; 32: 1531–1542. doi: [10.1007/s00299-013-1465-7](#) PMID: [23743655](#)
39. Li XB, Xu D, Wang XL, Huang GQ, Luo J, Li DD, et al. Three cotton genes preferentially expressed in flower tissues encode actin depolymerizing factors which are involved in F-actin dynamics in cells. *J Exp Bot*. 2010; 61: 41–53. doi: [10.1093/jxb/erp280](#) PMID: [19861654](#)
40. de Dios AJ, M'Rani-Alaoui M, Castro AJ, Rodriguez-Garcia MI. Ole e 1, the major allergen from olive (*Olea europaea* L.) pollen, increases its expression and is released to the culture medium during in vitro germination. *Plant Cell Physiol*. 2004; 45: 1149–1157. PMID: [15509837](#)
41. Jiang SY, Jasmin PX, Ting YY, Ramachandran S. Genome-wide identification and molecular characterization of Ole_e_1, Allerg_1 and Allerg_2 domain-containing pollen-allergen-like genes in *Oryza sativa*. *DNA Res*. 2005; 12: 167–179. PMID: [16303748](#)
42. Hu B, Liu BY, Liu L, Liu CL, Xu L, Ruan Y. Epigenetic control of Pollen Ole e 1 allergen and extensin family gene expression in *Arabidopsis thaliana*. *Acta Physiol Plant*. 2014; pp. 2203–2209.
43. Li SQ, Wan CX, Kong J, Zhang ZJ, Li YS, Zhu YG. Programmed cell death during microgenesis in a Hong-lian CMS line of rice is correlated with oxidative stress in mitochondria. *Funct Plant Biol*. 2004; 31: 369–376.
44. Jiang PD, Zhang XQ, Zhu YG, Zhu W, Xie HY, Wang XD. Metabolism of reactive oxygen species in cotton cytoplasmic male sterility and its restoration. *Plant Cell Rep*. 2007; 26: 1627–1634. PMID: [17426978](#)
45. Liu GB, Xu H, Zhang L, Zheng YZ. Fe binding properties of two soybean (*Glycine max* (L.) Merr.) LEA4 proteins associated with antioxidant activity. *Plant Cell Physiol*. 2011; 52: 994–1002. doi: [10.1093/pcp/pcr052](#) PMID: [21531760](#)
46. Carafoli E. Calcium signaling: a tale for all seasons. *Proc Natl Acad Sci USA*. 2002; 99: 1115–1122. PMID: [11830654](#)
47. Ermak G, Davies KJ. Calcium and oxidative stress: from cell signaling to cell death. *Mol Immunol*. 2002; 38: 713–721. PMID: [11841831](#)
48. Orrenius S, Zhivotovsky B, Nicotera P. Regulation of cell death: the calcium-apoptosis link. *Nat Rev Mol Cell Biol*. 2003; 4: 552–565. PMID: [12838338](#)
49. Means AR, Rasmussen CD. Calcium, calmodulin and cell proliferation. *Cell Calcium*. 1988; 9: 313–319. PMID: [3224371](#)
50. Rato C, Monteiro D, Hepler PT, Malho R. Calmodulin activity and cAMP signalling modulate growth and apical secretion in pollen tubes. *Plant J*. 2004; 38: 887–897. PMID: [15165182](#)
51. Ge XC, Dietrich C, Matsuno M, Li GJ, Berg H, Xia YJ. An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. *EMBO Rep*. 2005; 6: 282–288. PMID: [15723040](#)