


RESEARCH ARTICLE

Open Access



Comparative transcriptional analysis of *Capsicum* flower buds between a sterile flower pool and a restorer flower pool provides insight into the regulation of fertility restoration

Bingqiang Wei^{1*} , Lanlan Wang², Paul W. Bosland³, Gaoyuan Zhang¹ and Ru Zhang²

Abstract

Background: Cytoplasmic male sterility (CMS) and its restoration of fertility (Rf) system is an important mechanism to produce F₁ hybrid seeds. Understanding the interaction that controls restoration at a molecular level will benefit plant breeders. The CMS is caused by the interaction between mitochondrial and nuclear genes, with the CMS phenotype failing to produce functional anthers, pollen, or male gametes. Thus, understanding the complex processes of anther and pollen development is a prerequisite for understanding the CMS system. Currently it is accepted that the *Rf* gene in the nucleus restores the fertility of CMS, however the *Rf* gene has not been cloned. In this study, CMS line 8A and the Rf line R1, as well as a sterile pool (SP) of accessions and a restorer pool (RP) of accessions analyzed the differentially expressed genes (DEGs) between CMS and its fertility restorer using the conjunction of RNA sequencing and bulk segregation analysis.

Results: A total of 2274 genes were up-regulated in R1 as compared to 8A, and 1490 genes were up-regulated in RP as compared to SP. There were 891 genes up-regulated in both restorer accessions, R1 and RP, as compared to both sterile accessions, 8A and SP. Through annotation and expression analysis of co-up-regulated expressed genes, eight genes related to fertility restoration were selected. These genes encode putative fructokinase, phosphatidylinositol 4-phosphate 5-kinase, pectate lyase, exopolysaccharidase, pectinesterase, cellulose synthase, fasciclin-like arabinogalactan protein and phosphoinositide phospholipase C. In addition, a phosphatidylinositol signaling system and an inositol phosphate metabolism related to the fertility restorer of CMS were ranked as the most likely pathway for affecting the restoration of fertility in pepper.

Conclusions: Our study revealed that eight genes were related to the restoration of fertility, which provides new insight into understanding the molecular mechanism of fertility restoration of CMS in *Capsicum*.

Keywords: CMS, Fertility restorer, Pepper, RNA-sequencing

* Correspondence: bqwei@gsau.edu.cn

¹College of Horticulture, Gansu Agricultural University, Lanzhou 730070, China

Full list of author information is available at the end of the article



Background

Capsicum species are one of the most popular spice and vegetable crops in the world and *Capsicum annuum* is the most widely grown among the five domesticated species (*C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, and *C. pubescens*) [1]. Hybrid vigor is a phenomenon that is advantageous for breeders because of increased fruit yield, enhance resistance, and improve quality. However, the production of F₁ hybrid seed needs manual emasculation that can lead to a high cost for F₁ seed production [2]. The unique mechanism cytoplasmic male sterility (CMS) is one of the most valuable methods to utilize plant heterosis or hybrid vigor [3] because lack of pollen production removes the need for manual emasculation. Thus, breeders use the CMS/Rf system to produce hybrid seeds more economically, and the system is widely exploited for hybrid seed production of a number of crops including *Capsicum* [4–6].

Male sterility is the failure of plants to produce functional anthers, pollen, or male gametes. The CMS phenotype is maternally inherited because it is controlled by the plant's mitochondrial genome [4]. It has been proven that CMS is caused by chimeric open reading frames (ORF) resulting from the rearrangements of the mitochondrial genome [7, 8]. These ORF may disturb the function of ATPase [9], destroy the mitochondrial membrane structure [10], and produce proteins that are cytotoxic [11], which, in turn, affect the normal development of pollen [12]. In pepper, CMS was first reported in 1958 from an Indian *Capsicum annuum* accession (PI164835) [13]. Since then, the CMS/Rf system has been used to produce F₁ hybrid seeds of pepper [14]. Two candidate CMS genes, *orf456* and *atp6-2* loci, have since been identified [15, 16]. Another ORF, designated *orf507*, is a modified version of *orf456*, elongated through the deterioration of a stop codon. This ORF was also proven to be related to CMS and inhibit the formation of microspore in pepper [17].

The CMS phenotype can be reversed by a nuclear *Rf* gene. These genes have been found to restore fertility through several different mechanisms. The CMS transcript can be processed by a *Rf* gene at the post-transcriptional level [18]. In rice, the transcript of B-*atp6-orf79* is silenced by cleavage from the restorer gene RF1A [19]. Restorer of fertility genes can also act at the translational level, as *Rfo* in radish is thought to repress the translational processing of *orf138* [20]. In maize, a putative aldehyde dehydrogenase (*rf2*) was found that may act to restore fertility through detoxification during pollen development [21].

Restorer-of-fertility genes have been found in a range of flowering plant species. They have been reported in petunia [22], radish [20, 23–26], sugar beet [27, 28], maize [21], sorghum [29], and rice [19, 30–32]. These *Rf*

genes have encoded a variety of different proteins. For example, *Rf1* (*bvORF20*) encodes an OMA1-like protein in sugar beet [27, 28], *Rf2* encodes an aldehyde dehydrogenase protein in maize [21], *Rf17* gene encodes an acyl-carrier protein synthase [33] and the *Rf2* gene encodes a glycine-rich protein in rice [34], indicating the existence of a diverse set of *Rf* genes. However, most *Rf* genes are known to encode Pentatricopeptide Repeat (PPR) proteins [8, 35, 36]. In pepper, so far all predicted *Rf* genes have encoded PPRs [36, 37]. A PPR gene, *CaPPR6*, was identified as a strong *Rf* candidate based on expression pattern and characteristics of coding sequence [37]. Recently, 12 candidate PPR genes with similarity to previously reported *Rf* genes were also identified in pepper [38]. These candidate *Rf* genes provide a basis for further study of fertility restoration in pepper.

It has been suggested that fertility restorer in pepper may be controlled by two complementary genes [39]. Previous studies show that the fertility restorer of CMS in pepper is controlled by two major additive-dominant epistatic genes and an additive-dominant polygene [40], and two major QTLs and several minor QTLs [41]. It could be that the two major QTLs correspond with the two major additive-dominant genes, in which case the studies support one another. In contrast, another study indicates that one major QTL and four minor QTLs relate to fertility restoration in pepper [42]. Another phenotype of partial restoration has also been reported, in which the flower simultaneously produces functional and aborted pollen, which is thought to be controlled by a gene (*pr*) in the nucleus separate from *Rf* genes [43]. In addition, the CMS phenotype can be restored temporarily under low temperature, suggesting that temperature affects the expression of some fertility modification genes [44]. Together, these various types of fertility restoration demonstrate that CMS is complex, and currently do not have a complete understanding of the molecular mechanisms that underlie the CMS/Rf system in pepper.

The RNA-Seq method directly sequences transcripts by using high-throughput sequencing technologies, and it has considerable advantages for providing genome-wide information, detection of novel transcripts, and allele-specific expression [45]. Bulked segregate analysis (BSA) is an efficient method for the rapid identification of molecular markers for specific traits or target gene loci [46]. Combining the advantage of BSA and RNA-seq, BSA RNA-seq (BSR) can be used to analyze the differentially expressed genes (DEGs), and single nucleotide polymorphisms (SNPs) between the two genetic pools [47, 48].

In this study, BSR-seq was applied to identify DEGs related to the fertility restorer of CMS in pepper. In addition, the transcriptomes of two parent lines were sequenced. A set of candidate genes were selected that are associated with the fertility restoration in CMS in pepper

based on both the BSR-seq and parental transcriptome sequencing. The results provide new insights into the study of molecular mechanisms of restorer fertility of CMS in pepper.

Results

Database estimation of transcript sequencing

Through the RNA-sequence of the fertility restorer line (R1), the CMS line (8A), a population of 30 fertile plants pool (RP), and a population of 30 sterile plants pool (SP), a total of 41.84 GB of aligned data were obtained. The aligned data of R1 and 8A were 8.24 GB and 6.98 GB, respectively, and that of RP and SP were 14.16 GB and 12.46 GB, respectively (Table 1). These raw data can be found in NCBI (<https://www.ncbi.nlm.nih.gov/>) with an accession number of SRA895207. The base ratios for quality scores of each aligned data was greater than Q30% at more than 91%, and GC content was more than 41%. This indicates that the aligned data were good for the subsequent searches.

The aligned data had a high success rate of being mapped to a reference genome. The percentage of each aligned data being mapped to a reference genome was more than 85.50%, and the unique mapped rate was greater than 82.47% (Table 2).

Differentially Expressed Genes (DEGs) analysis

DEGs were discovered between R1/8A and between RP/SP (Table 3). There were 3790 DEGs between R1 and 8A, in which 2274 were up-regulated and 1516 were down-regulated for expression in R1 as compared to 8A. There were also 1762 DEGs between RP and SP, in which 1490 were up-regulated and 272 were down-regulated for expression in RP as compared to SP. In an overall comparison, there were 944 co-DEGs among the above groups (R1/8A, RP/SP). Within this comparison, 891 out of 944 DEGs were commonly up-regulated in restorer accessions and the remaining 53 out of 944 DEGs were down-regulated in restorer accessions (R1 and RP) as compared to CMS accessions (8A and SP). This indicates that 891 DEGs were up-regulated not only in RP as compared to SP, but also in R1 as compared to 8A, and 53 DEGs were down-regulated not only in RP as compared to SP, but also in R1 as compared to 8A.

Table 1 The estimation of RNA sequence data

Samples	Read Number ^a	Base Number ^b	GC Content ^c	% ≥ Q30 ^d
R1	32,694,918	8,239,119,336	41.70%	91.55%
8A	27,708,115	6,982,444,980	41.82%	91.32%
RP	56,202,271	14,162,972,292	42.48%	92.06%
SP	49,443,495	12,459,760,740	42.23%	91.50%

Note: ^apair-end reads number in aligned data; ^bbase number in aligned data; ^cGC content in aligned data; ^dThe percentage of bases whose quality value of aligned data is more than or equal to Q30

The 891 commonly up-regulated DEGs were the subject of subsequent study.

Gene Ontology (GO) Annotation

The 891 common up-regulated DEGs in restorer accessions were assigned to three main categories, cellular component, molecular function, and biological process. These three categories were composed of 53 functional groups using GO assignment (Fig. 1).

In the cellular component category, the majority functional groups were cell part, cell, organelle, and membrane, including 589, 573, 475, and 397 genes, respectively. The most significant GO node that is enriched to the DEGs is external side of plasma membrane (GO:0009897), composed of Capana01g001287, Capana01g000392, Capana06g000150, Capana01g004065 and Capana09g002397 totaling 5 up-regulated DEGs (Table 4). The next were GO:0009535, GO:0005615, GO:0030173, GO:0005618, GO:0048046 and GO:0016021, including 4, 5, 6, 91, 60 and 97 up-regulated DEGs, respectively (Table 4).

In the molecular function category, the dominant function types were catalytic activity and binding with 369 and 360 genes, respectively. The most significant GO node that is enriched to the DEGs is serine-type endopeptidase activity (GO:0004252), composed of Capana04g000159, Capana06g000150, Capana09g002397, Capana05g000136, Capana11g002184 and Capana01g001287 totaling 6 up-regulated DEGs (Table 5). The next GO nodes were protein kinase activity (GO:0004672), transmembrane receptor protein kinase activity (GO:0019199), 8-hydroxyquercetin 8-O-methyltransferase activity (GO:0030761) and isoflavone 4'-O-methyltransferase activity (GO:0030746), including 56, 4, 1 and 1 up-regulated DEGs, respectively (Table 5).

In the biological process category, the dominant terms were metabolic process, response to stimulus, biological regulation, developmental process and cellular component organization or biogenesis including 574, 564, 511, 460, 368, 351, and 332 genes, respectively. The most significant GO node that is enriched to the DEGs is mucilage extrusion from seed coat (GO:0080001), composed of Capana04g000159, Capana05g000136, Capana09g002397, Capana01g001287 and Capana06g000150 totaling five up-regulated DEGs (Table 6). The next nodes were GO:0016045, GO:0006898, GO:0010204, GO:0048359, GO:0010359, GO:0010102 and GO:0052544, including 9, 6, 1, 10, 7, 9 and 8 up-regulated DEGs, respectively (Table 6).

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Metabolic Pathway of DEGs

The up-regulated shared DEGs were annotated to 49 KEGG metabolic pathways in five categories including

Table 2 The comparison between sequence data and reference genomics

Samples	Total Reads	Mapped Reads ^a	Mapped Ratio ^b	Unique Mapped Reads ^c	unique Mapped Ratio ^d
R1	65,389,836	56,002,822	85.64%	54,068,310	82.69%
8A	55,416,230	47,886,185	86.41%	46,275,567	83.51%
RP	112,404,542	97,929,810	87.12%	94,135,100	83.75%
SP	98,886,990	84,549,668	85.50%	81,549,039	82.47%

Note: ^aaligned data that were mapped to the reference genome; ^b percentage of aligned data being mapped to the reference genome; ^cGC content of aligned data; ^d percentage of bases whose quality value of aligned data is more than or equal to Q30

genetic information processing, metabolism, organismal systems, cellular processes, and environmental information processing (Fig. 2). The metabolic pathways composed of the most up-regulated DEGs were starch and sucrose metabolism, oxidative phosphorylation, and plant-pathogen interaction. The next metabolic pathways composed of nine up-regulated DEGs were inositol phosphate metabolism, and the phosphatidylinositol signaling system (Fig. 2). KEGG pathways that may be involved in fertility recovery or pollen development include energy metabolism, carbohydrate metabolism, protein and amino acid metabolism, lipid metabolism, substance absorption and transport, and signal transduction (Additional File 1: Table S1). Interestingly, two metabolic pathways, phosphatidylinositol signaling system and inositol phosphate metabolism, were enriched as the most reliable pathway for enrichment significance (Fig. 3). Within these two metabolic pathways most genes were the same, with eight out of nine genes being shared between the two pathways.

In the phosphatidylinositol signaling system (ko04070), scored as the most enriched system, nine genes were up-regulated (Fig. 4). Among these nine genes, five (Capana00g002844; Capana00g004424; Capana10g001436; Capana10g002170; Capana10g002470) encode phosphatidylinositol-4-phosphate 5-kinase (PI(4)P5K) and catalyzes the phosphatidylinositol-4-phosphate (PI(4)P) to phosphatidylinositol-4,5-bisphosphate (PI(4,5)P₂). Meanwhile, two genes (Capana03g002795 and Capana06g002131) encode phospholipase C (PLC), which could hydrolyze phosphatidylinositol (PI), PI(4)P or PI(4,5)P₂ to generate double messenger molecules inositol triphosphate (IP₃) and diacylglycerol (DG). Two genes (Capana05g000173 and Capana07g002321) encode inositol phosphate phosphatase and phosphatidate cytidyltransferase, respectively, which could participate in the reduction of IP₃ and DG to PI.

Table 3 Three groups of DEGs

DEG Set	DEG number	Up-regulated DEG number	Down-regulated DEG number
R1/8A	3790	2274	1516
RP/SP	1762	1490	272
R1/RP n 8A/ SP	944	891	53

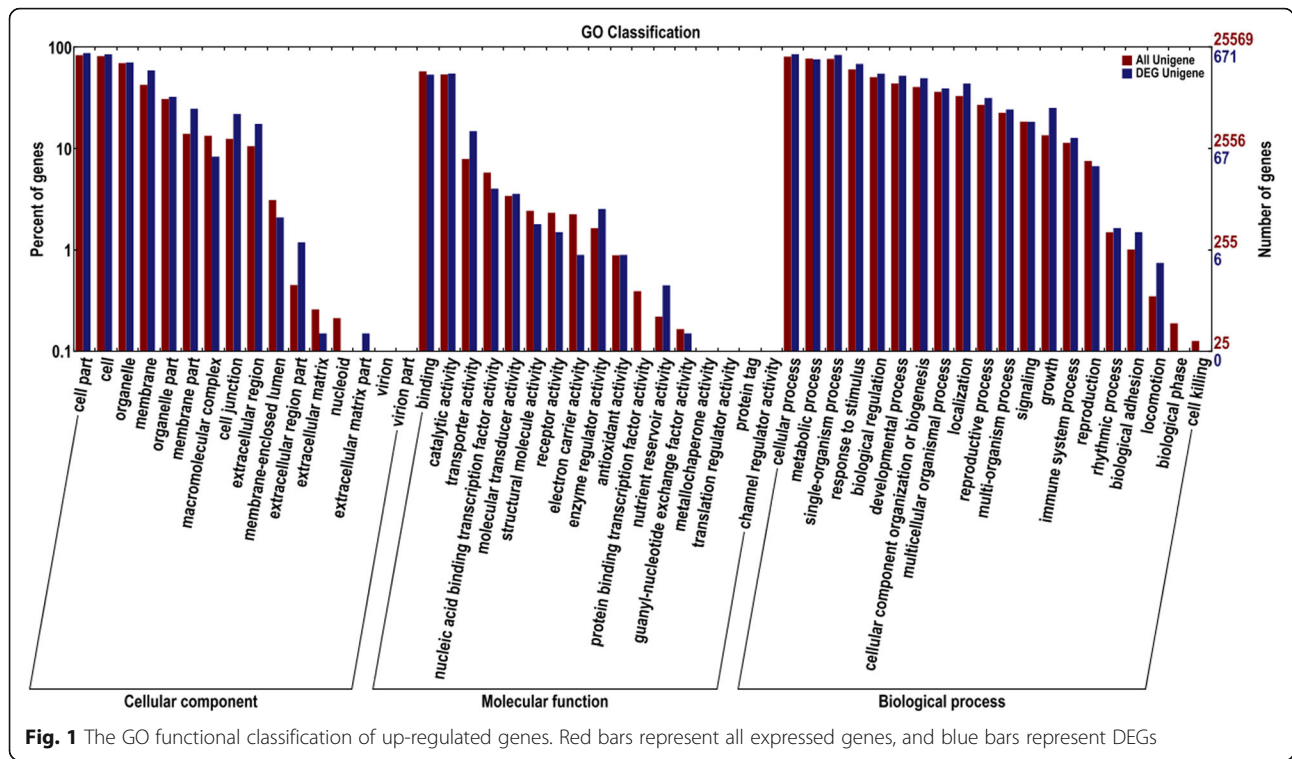
Related Genes Selection and expression

According to conjoined analysis of the gene FPKM (fragments per kilobase of exon per million fragments mapped) value, difference multiple, qRT-PCR value, functional annotation and metabolic pathway classification, eight genes, related to fertility restoration of CMS in *Capsicum* were selected. The eight genes were Capana00g002348, Capana00g004424, Capana02g000930, Capana00g003267, Capana01g002849, Capana05g002270, Capana01g004065 and Capana06g002131, and they encode fructose kinase, phosphatidylinositol phosphokinase, pectin lyase, extragalacturonase, pectin esterase, cellulose synthase, and bundle arabinogalactan protein, respectively (Table 7). All listed genes are key enzymes and proteins in anther and pollen development.

According to RNA-sequencing, these genes were dramatically up-regulated in the restorer parent (R1) and restorer pool (RP). The majority of these selected genes related to fertility restoration showed very little expression in the CMS 8A and SP plants, and one selected gene, Capana00g004424, had no detected expression in CMS line 8A and SP (Table 7). The lack of expression of these genes was validated by qRT-PCR between 8A and R1, as well as SP and RP. The qRT-PCR results indicated that these genes were up-regulated in both R1 and RP, as compared to 8A and SP, which was completely consistent with the sequencing results (Fig. 5).

The relative expression of two obviously down-regulated genes (NewGene11661 and NewGene949) both in fertile accessions R1 and RP as compared to sterile accessions 8A and SP were validated by qRT-PCR. The results indicated that two genes were also down-regulated in R1 compared to 8A (Fig. 6).

Tissue from four different developmental stage buds tested the relative expression, and the different stages are shown in Fig. 7. The qRT-PCR results show that the expression of these genes had a lower and relatively stable level among four developmental stages of flower buds in 8A. However, in the F₁ generation, the relative expression varied among the different developmental stages. In the F₁ generation, as the flower buds developed, the relative expression had a dramatic increase. At stage III, the expression increased rapidly until it peaked at stage IV (Fig. 8).



In addition, the relative expression of two down-regulated genes were analyzed in four developmental stages of the buds between 8A and F₁. The relative expression of NewGene11661 improved with the development of flower buds in two accessions, and the expression in every stage of 8A were higher than in R1 (Fig. 9). Unfortunately, although the relative expression of NewGene949 was higher in every stage in 8A than in R1, there was not agreement with the expected tendency (Fig. 9).

Discussion

Male sterility and the fertility restorer system are an extremely complex biological process involving substance and energy metabolism, signal transduction pathway, substance transportation, pollen wall morphogenesis, tapetum formation and programmed cell death (PCD), and a series of related gene expression regulation processes.

Anthers are the strongest energy reservoir in the flower organs, and have a very active metabolism during

Table 4 The enrichment results for the cellular component DEGs by topGO

GO ID ^a	Term ^b	Annotated ^c	Significant ^d	Expected ^e	KS ^f
GO:0009897	external side of plasma membrane	131	5	3.63	3.20E-07
GO:0009535	chloroplast thylakoid membrane	715	4	19.83	5.90E-05
GO:0009522	photosystem I	89	0	2.47	0.00023
GO:0005615	extracellular space	95	5	2.63	0.00035
GO:0030173	integral to Golgi membrane	24	6	0.67	0.00047
GO:0005618	cell wall	2279	91	63.21	0.00064
GO:0048046	apoplast	1336	60	37.06	0.00085
GO:0009523	photosystem II	104	0	2.88	0.00133
GO:0030076	light-harvesting complex	43	0	1.19	0.00134
GO:0016021	integral to membrane	2151	97	59.66	0.00153

Note: ^aGO term ID; ^bGO function; ^call genes annotated the function; ^dDEGs annotated the function; ^eExpected value of the DEGs annotated the function; ^fStatistical significance of enrichment nodes, the smaller the KS value, the more significant enrichment

Table 5 The enrichment results of molecular function DEGs by topGO

GO ID ^a	Term ^b	Annotated ^c	Significant ^d	Expected ^e	KS ^f
GO:0004252	serine-type endopeptidase activity	175	6	4.71	1.20E-13
GO:0004672	protein kinase activity	2053	56	55.27	1.50E-08
GO:0019199	transmembrane receptor protein kinase activity	203	4	5.46	5.30E-08
GO:0036202	ent-cassa-12,15-diene 11-hydroxylase activity	32	0	0.86	4.90E-06
GO:0010327	acetyl CoA:(Z)-3-hexen-1-ol acetyltransferase activity	14	0	0.38	8.70E-06
GO:0048038	quinone binding	53	0	1.43	2.20E-05
GO:0030761	8-hydroxyquercetin 8-O-methyltransferase activity	44	1	1.18	2.40E-05
GO:0030746	isoflavone 4'-O-methyltransferase activity	44	1	1.18	2.40E-05
GO:0047203	13-hydroxylupinine O-tigloyltransferase activity	15	0	0.4	3.10E-05
GO:0033793	aureusidin synthase activity	10	0	0.27	4.30E-05

Note: ^aGO term ID; ^bGO function; ^call genes annotated the function; ^dDEGs annotated the function; ^eExpected value of the DEGs annotated the function; ^fStatistical significance of enrichment nodes, the smaller the KS value, the more significant enrichment

development. A large number of sugars are transported to the anther [49], which can regulate the expression of genes as both a substrate of carbohydrate and a signal molecule in anthers [50]. In this study, one gene encoding fructokinase had an up-regulated expressed value of log₂FC more than 8.5 times in R1/8A as well as RP/SP in a sucrose and starch metabolism pathway, which indicated that fructokinase is one of the key genes that regulates the fertility restorer of CMS in pepper.

Generally, pectin and callose in pollen mother cell (PMC) are degraded absolutely, otherwise the microspore cannot be separated from the tetrad [51, 52]. The pectin methylesterification, is degraded by de-methylated esterification first by methylesterases (PMEs), then the combined action of PME, pectin lyases (PLs), and polygalacturonases (PGs) [53, 54]. In Arabidopsis, QRT1 gene encodes a PME, QRR2, and QRT3 encoding an external PG and an endonuclear PG, respectively. The loss of function of QRT1 didn't reduce the level of pectin methylesterification. If the level of the pectin methylesterification is higher, it cannot be degraded by PG [55]. In this study,

three genes encoding PME, PL and exo-PG were selected. What is more, these genes had up-regulated expression and the log₂FC were as high as 11.4473, 10.2657, and 8.97438, respectively. The results presented are in agreement with Hamid et al [56] that showed that eight pectin lyase-like superfamily protein coding genes and five pectin methylesterase genes were up-regulated in fertile plants as compared to Cytoplasmic Genic Male Sterility (CGMS) in cotton. In addition, a fasciclin-like protein gene was also up-regulated in restorer materials in this study. This result was similar to the reports of Hamid et al. [56], in which three cytoskeleton organization genes were up-regulated in fertile lines in cotton. These up-regulated genes may have a positive role in the degradation of pectin and the normal release of microspores.

It is well known that celluloses and hemicelluloses are the important components of pollen intine. Cellulose composed of a class of β -1,4-glucan molecular long chain plays an important role in cell wall toughness and strength. The deposition of cellulose on intine is mainly carried out by cellulose synthase complexes (CSCs) that

Table 6 The enrichment results of the biological process DEGs by topGO

GO ID ^a	Term ^b	Annotated ^c	Significant ^d	Expected ^e	KS ^f
GO:0080001	mucilage extrusion from seed coat	109	5	2.96	1.30E-15
GO:0002764	immune response-regulating signaling pathway	135	1	3.66	3.80E-13
GO:0016045	detection of bacterium	271	9	7.36	2.50E-12
GO:0006898	receptor-mediated endocytosis	139	6	3.77	3.30E-10
GO:0010204	defense response signaling pathway, resistance gene-independent	157	1	4.26	3.40E-09
GO:0048359	mucilage metabolic process involved seed coat development	205	10	5.56	7.90E-09
GO:0010359	regulation of anion channel activity	339	7	9.2	8.60E-09
GO:0010102	lateral root morphogenesis	322	9	8.74	5.00E-08
GO:0009864	induced systemic resistance, jasmonic acid mediated signaling pathway	107	0	2.9	1.20E-06

Note: ^aGO term ID; ^bGO function; ^call genes annotated the function; ^dDEGs annotated the function; ^eExpected value of the DEGs annotated the function; ^fStatistical significance of enrichment nodes, The smaller the KS value, the more significant enrichment

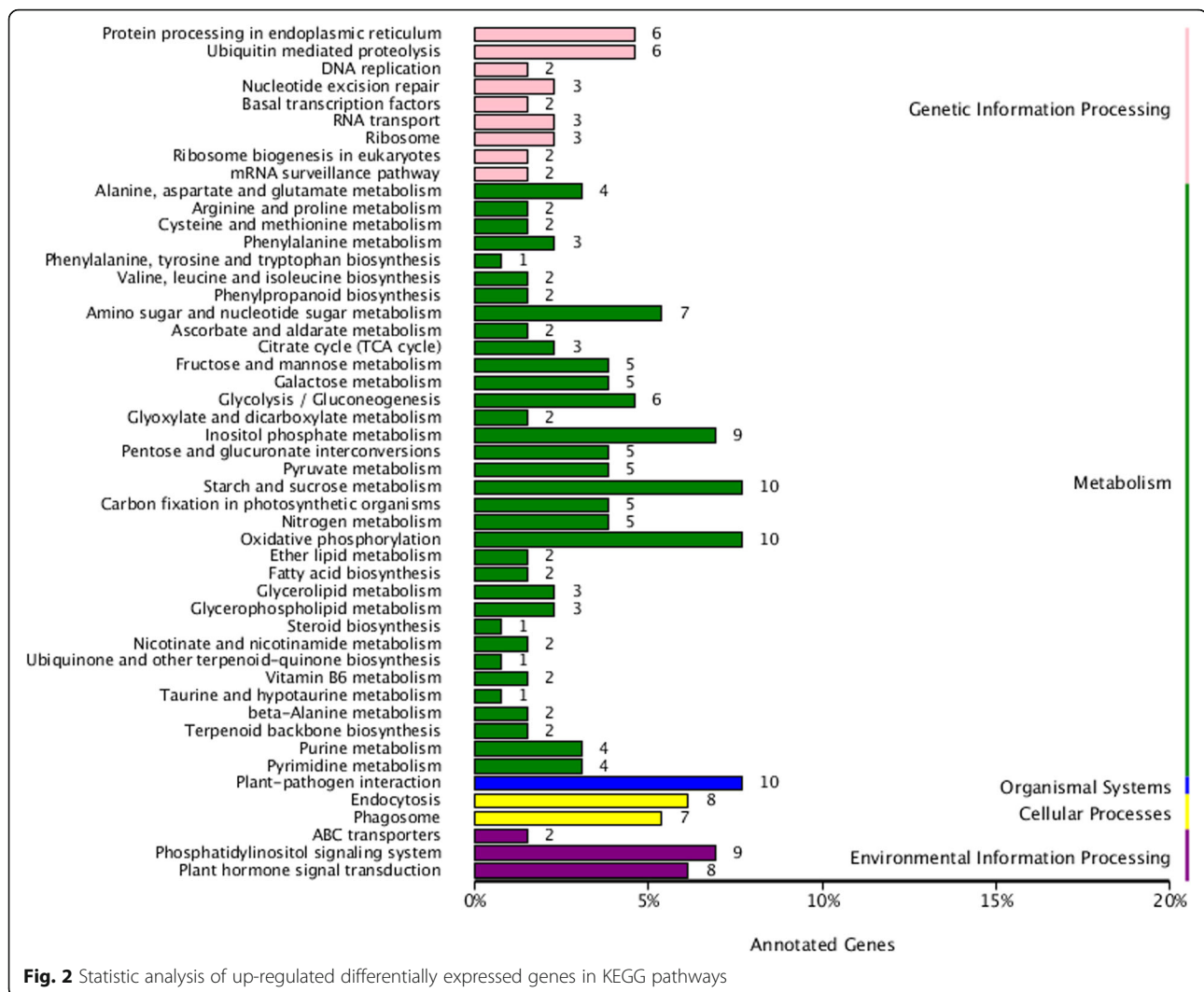
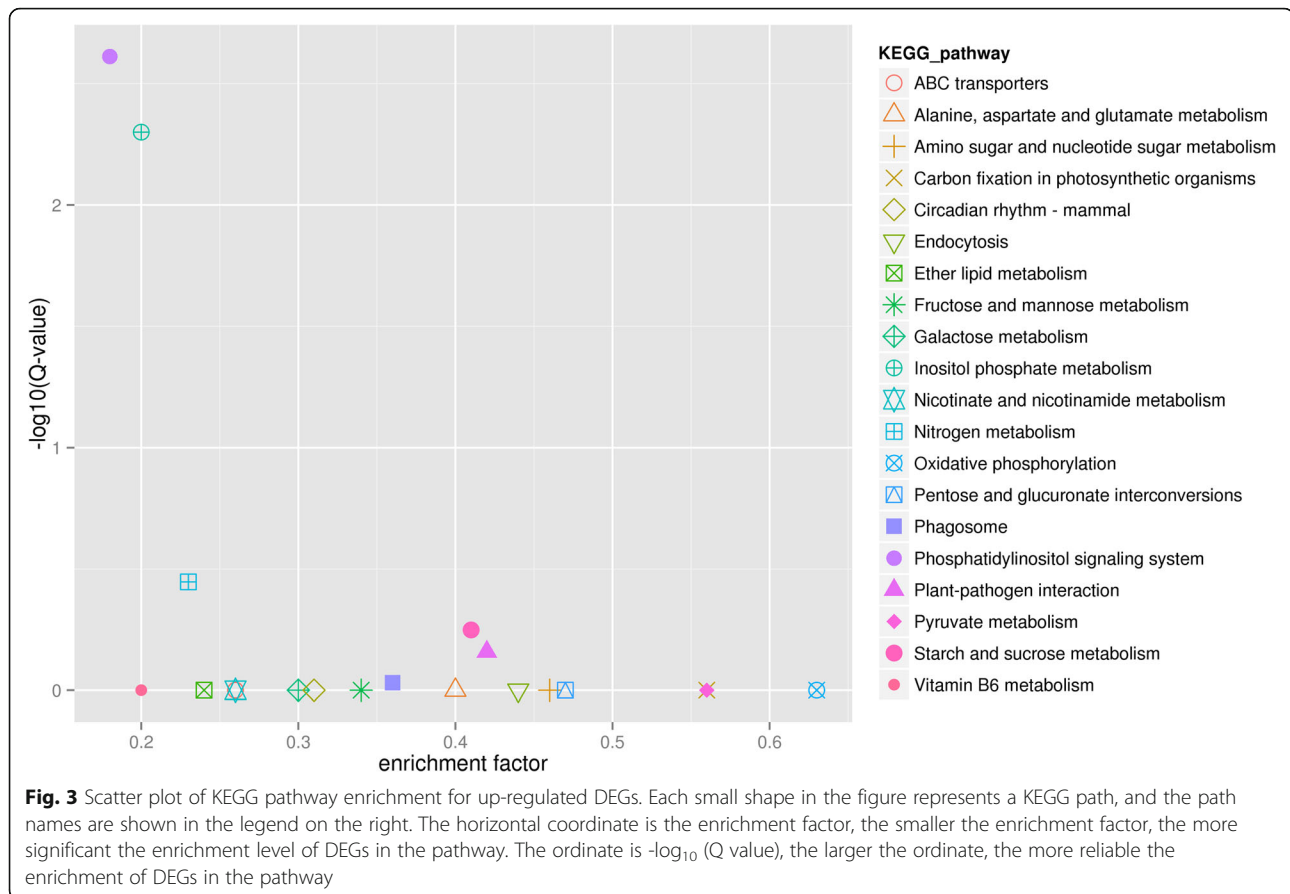


Fig. 2 Statistic analysis of up-regulated differentially expressed genes in KEGG pathways

is located on the cell membrane [57, 58]. In *Arabidopsis*, the deposition of cellulose in intine and extine was not well-distributed in the mutant of CESA1 gene and CESA3 gene, which resulted in abnormal pollen development. Fortunately, we also choose one gene encoding cellulose synthase that is up-regulated dramatically in restorer accessions.

It is commonly thought that the phosphatidylinositol signal system is one of the important signal systems for plant seed germination, growth and reproduction, senescence and response to environmental factors [59]. Pollen development undergoes a series of complex cell division and differentiation processes, which involve the dynamic changes of many cell components and internal subcellular components, including vacuole and cytoskeleton [60]. Many components of phosphatidylinositol signaling system participate in the vacuolar diversification during pollen development and vesicle transport in pollen tube growth. In this study, eight genes were up-regulated in

phosphatidylinositol signal system. There are five genes that encode PI(4)P5K, which can catalyze PI(4)P to PI(4,5)P₂. Previous studies have shown that PI(4)P is very important for pollen and stigma affinity, and PI(4,5)P₂ plays an important role in vesicle transport and cell skeleton rearrangement [61]. PI(4)P5K is a very important enzyme in the development of the anther, it would lead to an abnormal morphology of the pollen tube if PI(4)P5K was lost, thus inhibiting the germination of pollen and the growth of the pollen tube [62–65]. In *Arabidopsis*, PIP5K1 and PIP5K2 are important for vacuole biogenesis and early pollen development, pollen grains from flowers of the pip5k1+/-pip5k2+/-mutants show defects in vacuoles and exine wall formation [65]. In addition, two genes encode PLC, and PLC is the most important in phosphatidylinositol signal system, which could hydrolyze PI, PI(4)P and PI(4,5)P₂, to double as the messenger molecules inositol trisphosphate (IP₃) and diacylglycerol (DG). It is also known that there is a calcium dependent PLC



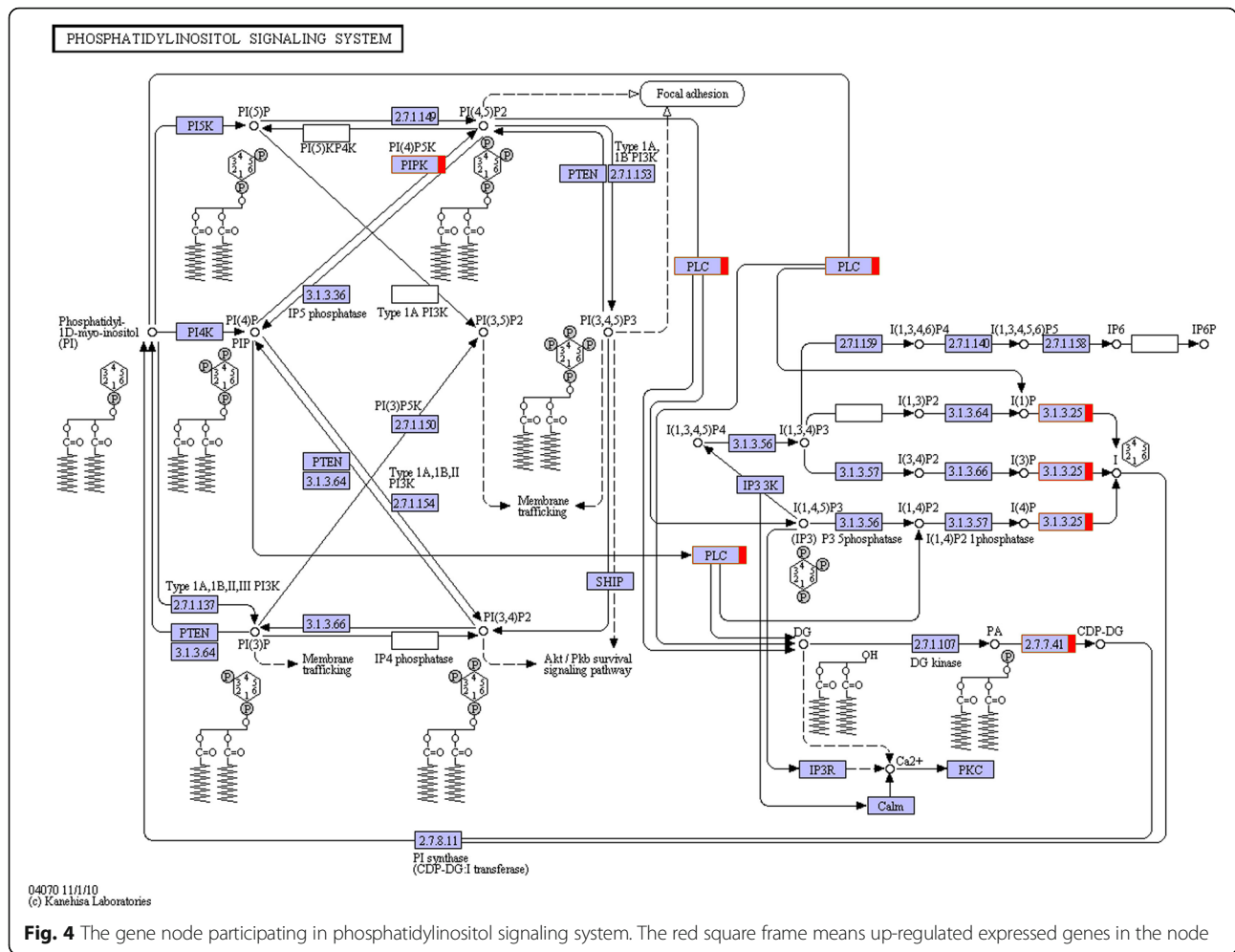
activity in pollen tubes, and PLC and IP_3 are involved in the germination and growth of pollen tubes [66]. PLC can monitor the growth of the pollen tube not only by adjusting the content of $PI(4,5)P_2$, but also by regulating the concentration of Ca^{2+} and membrane secretion.

The development of pollen walls plays a key role in the development of pollen, and it leads to male sterility in plants if the development of the pollen outer wall is blocked. Sporopollenin is the main substance that constitutes the outer wall of pollen, and the sporopollenins are mainly composed of polymerized phenols and long-chain fatty acid derivatives, which are difficult to decompose and can effectively protect pollen from degradation from the outside. There is also a thick oil-bearing layer outside of the pollen grain, including hydrophobic lipids and proteins, composed of long-chain fatty acid derivatives [67, 68]. Previous studies showed that sporopollenin biosynthesis is closely associated with fatty acid metabolism [69, 70]. Therefore, lipid accumulation plays an important role in pollen morphogenesis. In this study, some up-regulated genes related to lipid metabolism pathway were discovered, and these metabolic genes may be involved in the synthesis of lipids in the tapetum, which are then transported to the pollen wall surface (Additional file 1: Table S1). It had also been reported in

cabbage that in the protein interaction network around CYP704B1, more than 1/3 of the protein species were involved in fatty acid metabolism, and all of them were significantly reduced in 83121A [71].

Vesicle transport and lipid carrier transport are the two main modes of sporopollen transportation. Adenosine triphosphate binding cassette (ABC) is a protein family in which ABC members were involved in the transmembrane transport of substances. In *Arabidopsis thaliana*, more than 120 ABC proteins have been identified. They are involved in the transport of ions, hormones, secondary metabolites, lipids, and proteins. The ABCG26/GBC27 belongs to the G subfamily of ABC protein and is expressed in microspore and tapetum, which may be involved in the transport of sporopollen [72, 73]. In this study, two up-regulated ABC transport genes were discovered that may be involved in the transportation of sporopollen (Additional file 1: Table S1).

In addition, the synthesis of fatty acid needs ATP and acetyl-CoA that come from cell respiration and energy metabolism. Plant energy metabolism mainly includes the tricarboxylic acid cycle, glycolysis, pentose phosphate pathway, and the electron transport and oxidative phosphorylation. This energy metabolism provides ATP and acetyl-CoA for the synthesis of fatty acids. Previous



studies have shown that the impairment of respiratory chain enzymes and enzyme complexes can lead to CMS [74, 75]. Similar to the results in eggplant [76], this study found many up-regulated genes in the restorer line to participate in energy pathways, such as oxidative phosphorylation, glycolysis, and citrate cycle (Additional file 1: Table S1).

The relative expression of NewGene11661 was down-regulated in the F₁ while up-regulated in 8A, and the expression increased with the development of flower buds. However, it did not improve more dramatically than other candidate genes. NewGene11661 was annotated as UDP-glycosyltransferase gene. The UDP glycosyltransferase could

Table 7 The selected genes related to the restorer-of-fertility and their expression and function annotation

Gene ID	FPKM				log ₂ FC		Nr_annotation
	8A	R ₁	SP	RP	R ₁ /8A	RP/SP	
Capana00g002348	0.086539	0.264261	1.76E-06	0.611852	8.59032	9.6092	putative fructokinase-5-like [<i>Solanum tuberosum</i>]
Capana00g004424	0	1.56466	0	2.64384	7.27376	7.7515	phosphatidylinositol 4-phosphate 5-kinase 4-like [<i>Solanum tuberosum</i>]
Capana02g000930	0.280479	150.854	1.74488	322.03	8.97438	7.50312	pectate lyase-like [<i>Nicotiana tomentosiformis</i>]
Capana00g003267	0.112148	1.39687	0	3.16586	10.2657	6.57239	expolygalacturonase clone GBGE184-like [<i>Nicotiana tomentosiformis</i>]
Capana01g002849	7.51E-05	1.2932	6.12E-06	21.9292	5.49358	11.4473	pectinesterase4-like [<i>Solanum lycopersicum</i>]
Capana05g002270	0	2.81676	2.06E-05	10.1792	8.25692	9.93493	cellulose synthase-like protein D1-like [<i>Solanum tuberosum</i>]
Capana01g004065	0.148665	55.2292	0.128079	158.337	8.39073	9.87305	fasciclin-likearabinogalactan protein 3 [<i>Solanum lycopersicum</i>]
Capana06g002131	5.21E-05	1.04156	0.203871	0.65287	4.64828	4.27352	PLC

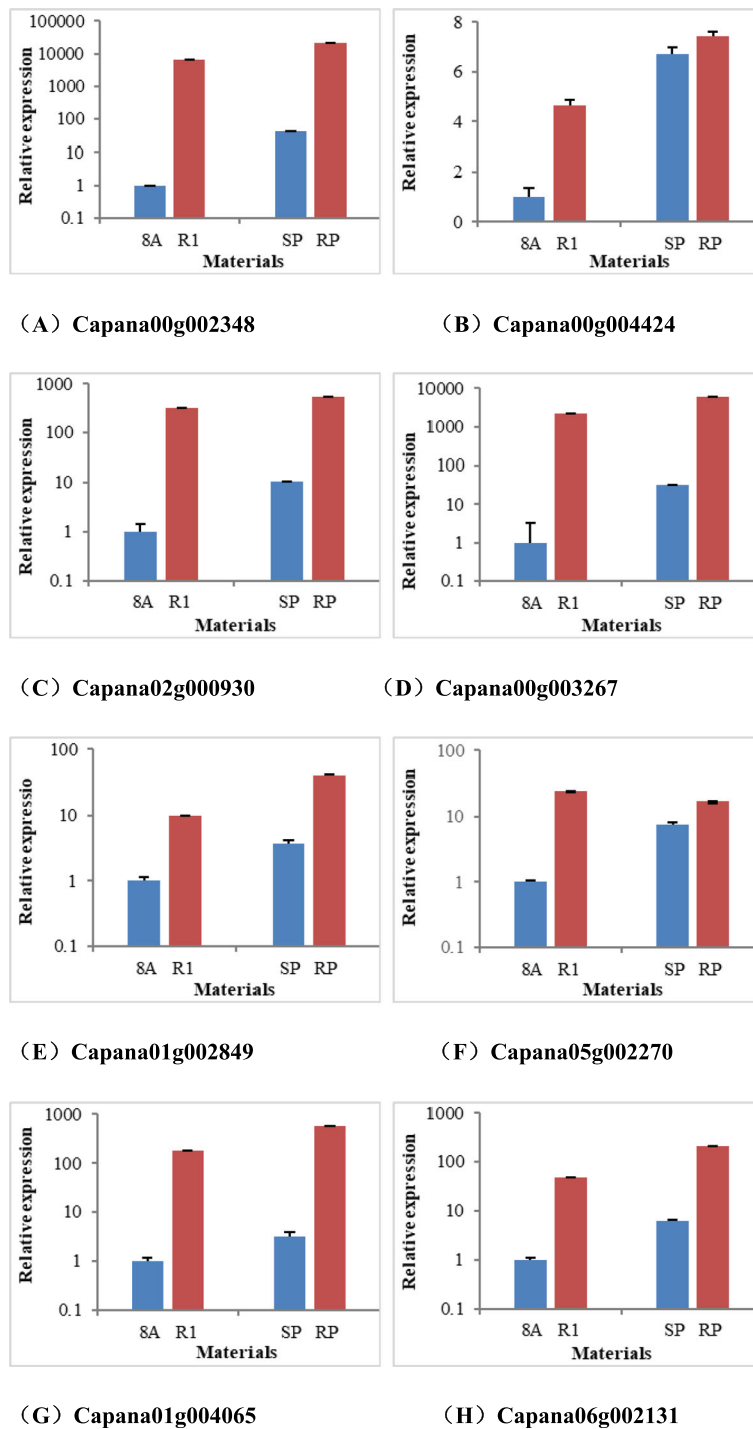


Fig. 5 The qRT-PCR validation of eight genes related to fertility restorer between 8A and R1, as well as SP and RP. **(a)** The relative expression of Capana00g002348 in four accessions. **(b)** The relative expression of Capana00g004424 in four accessions. **(c)** The relative expression of Capana02g000930 in four accessions. **(d)** The relative expression of Capana00g003267 in four accessions. **(e)** The relative expression of Capana01g002849 in four accessions. **(f)** The relative expression of Capana05g002270 in four accessions. **(g)** The relative expression of Capana01g004065 in four accessions. **(h)** The relative expression of Capana06g002131 in four accessions

catalyze conjugation of lipophilic chemicals with sugar donated by UDP-glycoside, generating water-soluble products that can be easily excreted and resulting in the detoxification

and elimination of their substrate [77, 78]. However, the roles of this enzyme in the development pollen or CMS have been rarely reported and needs further study.

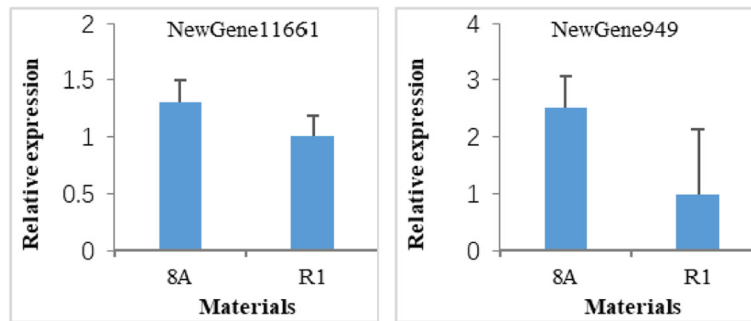


Fig. 6 The qRT-PCR validation of two obviously down-regulated genes between 8A and R1

Conclusion

By RNA sequencing, in this study 891 DEGs were up-regulated and 53 DEGs were down-regulated in restorer accessions as compared to the CMS accessions. The 891 up-regulated DEGs in restorer accessions were assigned to three main functional categories: cellular component, molecular function, and biological process, all of which were composed of 53 functional groups using GO assignment. The up-regulated DEGs were annotated to 49 KEGG metabolic pathways in five categories including genetic information processing, metabolism, organismal systems, cellular processes, and environmental information processing. As in Fig. 3, two metabolic pathways, phosphatidylinositol signaling system and inositol phosphate metabolism, were statistically the most likely pathways for affecting fertility restoration. Finally, eight DEGs were selected that encoded fructose kinase, phosphatidylinositol phosphokinase, pectin lyase, extragalacturonase, pectin esterase, cellulose synthase, and bundle arabinogalactan protein. The qRT-PCR results showed that the expression of these genes had a lower and relatively stable level between four developmental stages of the flower bud in 8A. But in the F_1 generation, with the development of flower buds, the relative expression had a tendency to significant

increase, especially at the stage III, where the expression increased rapidly and peaked at stage IV. The tendency to significantly increase the eight candidate genes supports RNA differential expression and are likely to be involved in pepper fertility restoration.

Materials and Methods

Plant sample collection and preparation

The CMS 8A line was completely sterile, and the fertility restorer line R1 completely restored the fertility of 8A (Fig. 10). The F_2 population from the self-pollination of the F_1 individuals showed segregation for fertility. The restore fertility pool (RP) was constructed from 30 polar fertile plants and the sterile pool (SP) was constructed from 30 polar sterile plants. The flower buds of 8A, R1, SP, and RP at the tetrad stage were used for total RNA extraction. Four different developmental stages of flower buds of 8A and F_1 analyzed the relative expression of fertility restoration related genes.

RNA extraction and library construction and RNA Sequencing

Total RNA was extracted separately from the flower buds of 8A, R1, SP, and RP according to the instructions



Fig. 7 The four developmental stages of flower buds used for qRT-PCR analysis

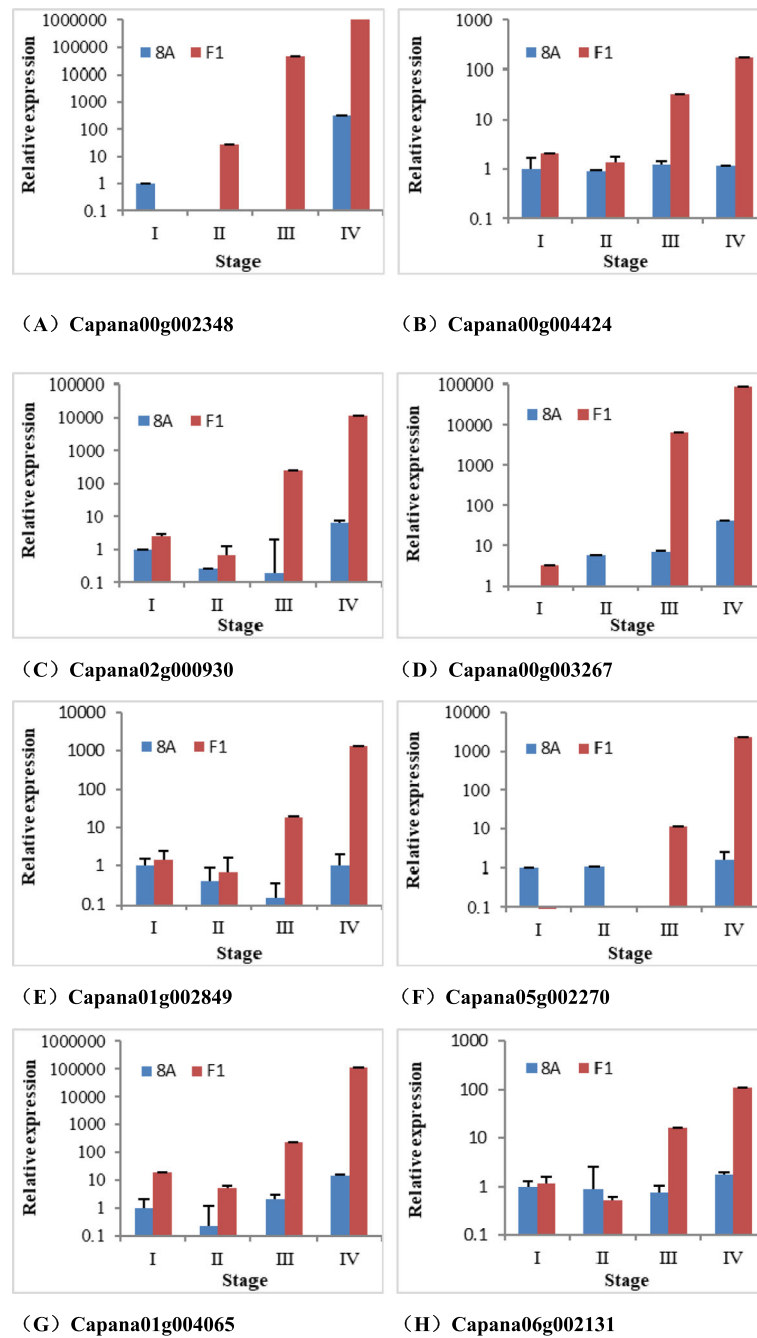


Fig. 8 The expression of genes related to fertility restorer in different developmental stage of buds between 8A and F₁. **(a)** The relative expression of Capana00g002348. **(b)** The relative expression of Capana00g004424. **(c)** The relative expression of Capana02g000930. **(d)** The relative expression of Capana00g003267 in four accessions. **(e)** The relative expression of Capana01g002849. **(f)** The relative expression of Capana05g002270. **(g)** The relative expression of Capana01g004065. **(h)** The relative expression of Capana06g002131

of Trizol Reagent (Life technologies, California, USA). The RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). The RNA concentration was measured using Qubit® RNA Assay Kit in Qubit®2.0 Fluorometer (Life Technologies, CA, USA). The RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer

2100 system (Agilent Technologies, CA, USA). Then the mRNA was enriched using magnetic beads with Oligo (dT) for reverse transcription. The cDNA library was constructed following the manufacturer’s instructions for NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530). Finally, the cDNA libraries of the four flower buds from 8A, R1, SP, and RP have been

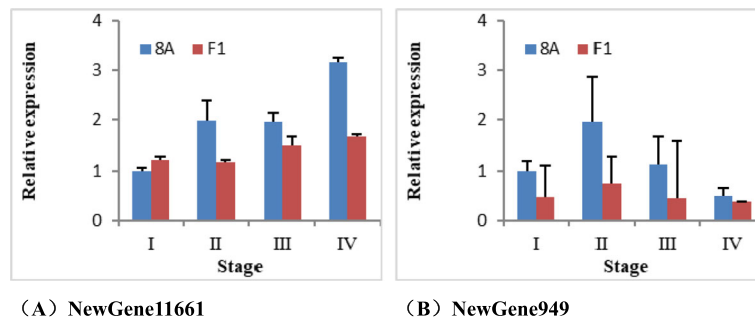


Fig. 9 The relative expression of two down-regulated genes in different developmental stage of buds between 8A and F₁. (a) The relative expression of NewGene11661. (b) The relative expression of NewGene949

sequenced on a flow cell using the Illumina HiSeq™ 2500 sequencing platform.

Transcriptome analysis using reference genome-based reads mapping

Low quality reads, such as adaptor sequences, unknown nucleotides greater than 5%, or low quality score (Q-value) ≤ 20 bases, were removed by Perl script. The aligned reads that were filtered from the raw reads were mapped to the [Pepper Genome Database \(release 2.0\)](#) [79] using TopHat2 software [80]. The aligned records in BAM/SAM format examined to remove potential duplicate molecules. Gene expression levels were estimated using FPKM values by the Cufflinks software [81]. Identification of new genes was based on transcript discovery with Cufflinks, and then filtering the genes encoding for a short peptide chain (less than 50 amino acid residues). By comparing the raw gene model and the result of transcriptome assembly, gene structure was optimized. SpliceGrapher integrated splice graphs generated from Cufflinks output and read alignments to produce comprehensive alternative splicing predictions [82].

Identification of DEGs

To evaluate DEGs among lines 8A, R1, SP, and RP, the EBSeq [83] was employed with screening criteria $FC \geq 2$ / $\log_2FC \geq 1$ and False Discovery Rate (FDR) < 0.01 . The differential multiple / FC represents the ratio of the expression between the two groups. By adjusting the difference between P-values, the FDR is obtained. Because the differential expression analysis of transcriptome sequencing is an independent statistical hypothesis test for a large number of gene expression values, there is a false positive problem. To address this, in the process of differential expression analysis, the accepted Benjamini-Hochberg correction method corrected the significant P-value obtained from the original hypothesis test, and finally FDR was used as the key index for screening DEGs.

GO classification and enrichment analysis

Perl script plotted GO functional classification for the unigenes with a GO term hit to view the distribution of gene functions. GO enrichment analysis of the DEGs was implemented by the Goseq R packages based Wallenius non-central hyper-geometric distribution [84], which can adjust for gene length bias in DEGs.



Fig. 10 The flower pictures of CMS line 8A and fertility restorer line R1

KEGG pathway and enrichment analysis

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism, and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (<http://www.genome.jp/kegg/>) [85]. This study used KOBAS software to test the statistical enrichment of differential expression genes in KEGG pathways [86].

Screening of fertility restoration related genes

Genes related to fertility were selected according to comprehensive estimation of gene \log_2FC value and gene function. Firstly, genes that were dramatically up-regulated ($\log_2FC \geq 7$) in restorer accessions as compared to sterile accessions were selected. Genes were selected on the basis of having FPKM values higher in restorer accessions and having FPKM values close to zero in sterile accessions (Table 7). Secondly, this group was filtered according to function, choosing genes playing an important role in the normal development of pollen as found in previous

studies, or are involved mostly in the important or enriching pathways in this study to further evaluate.

Expression analysis of fertility restoration related genes

Extraction of RNA was performed under the conditions described above. The first strand cDNA synthesis was carried out using the 1st strand cDNA synthesis kit (Revert Aid Premium Reverse Transcriptase) (Thermo Scientific, EP0733). A total of eight fertility restorer related genes and two down-regulated genes were tested for their relative expression and the *CaActin* (GenBank Accession: GQ339766.1) gene was the internal control (Table 8). The total volume of each reaction is 20 μ L. Using SG Fast qPCR Master Mix (High Rox) (2 \times), qRT-PCR was carried out (BBI, B639273). The qRT-PCR was performed on an ABI Stepone plus Real-Time PCR System (Applied Biosystems, USA) with the following cycling parameters: 95°C for 3 m, followed by 45 cycles of: 95 °C for 7 s, 57°C for 10 s, 72°C for 15 s. The relative expression was calculated by using the $2^{-\Delta\Delta Ct}$ method described by Livak and Schmittgen [87]. All reactions were performed with at least three replicates.

Table 8 The test genes and their primers for qRT-PCR

Gene	Forward Primer		Reverse Primer		length
	Sequence	Tm	Sequence	Tm	
Capana00g002348	AGGGTCCAAA ATGGGCTGT	59.0	CTCCCCAAA CATACGACGA	58.3	190
Capana00g004424	CAGGTG AATGGAAG AATGGTG	57.7	ACATAAAAGC TCCCATCA GTCC	57.8	139
Capana02g000930	TGCAAAGCTA CAAA ACCATCG	59.2	ATCACCTCA TCCCCAGTCTT	59.6	191
Capana00g003267	GTGGTGGTAC TCTAAACG GACAA	58.7	TGAATCTGGT GCTGTAAT CTTGAG	59.1	215
Capana01g002849	TATGGCGATG GATGATTGC	57.2	GGTAGGCGTA GACCTTTGAA AG	58	149
Capana05g002270	CCCTTGAGGA ACTATGGCGTA	59.5	CTTCATCATC ACCAGCAG ACTTT	58.7	150
Capana01g004065	TGTTGTAGGC TCTGTGCGTG	57.7	GGCTGAGATA GGTGCCATTG	57.9	133
Capana06g002131	TCAGGGCTAA TGTTGGTTGT	57.7	TTCTTACTT CATCAGCAGG GA	57.1	238
NewGene11661	GCCCATCGCT GTTCTTATTG	58.3	TGAAGACGGG CGTGAGTGT	59.2	116
NewGene949	TTTGAGGTGG GGTGGTTGT	57.9	AGTGGTTGCC ATGAAGTTG	58.1	166
CaActin	TGCCTGATGG ACAAGTTA TTACC	58.8	TGAGCACAAT GTTACCGT AGAGG	59.7	174

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6210-3>.

Additional file 1: Table S1. The KEGG Pathway related to the pepper restoration of fertility or pollen development.

Abbreviations

ABC: triphosphate binding cassette; BSA: Bulked segregate analysis; BSR: BSA RNA-seq; CGMS: Cytoplasmic Genic Male Sterility; CMS: Cytoplasmic male sterility; CSCs: Cellulose synthase complex; DEG: Differentially expressed gene; DG: Diacylglycerol; FC: Fold Change; FDR: Discovery Rate; FPKM: Fragments per kilobase of exon per million fragments mapped; GO: Gene Ontology; IP₃: Inositol triphosphate; KEGG: Kyoto Encyclopedia of Genes and Genomes; ORF: Open reading frames; PCD: Programmed cell death; PG: Polygalacturonase; PI (4) P: Phosphatidylinositol-4-phosphate; PI (4) P5K: Phosphatidylinositol-4-phosphate 5-kinase; PI (4,5) P₂: phosphatidylinositol-4,5-bisphosphate; Pl: Phosphatidylinositol; PL: Pectin lyase; PLC: Phospholipase C; PMC: Pollen mother cell; PME: Methylsterase; PPR: Pentatricopeptide Repeat; qRT-PCR: quantitative real-time polymerase chain reaction; QTL: Quantitative trait loci; Rf: Fertile restorer / restoration of fertility; RP: Restorer pool; SP: Sterile pool

Acknowledgements

Special thanks to Franchesca Ortega from the Chile pepper breeding and genetics program at New Mexico State University for editing the manuscript.

Authors' contributions

BW designed the experiments, developed the sterile pool and fertile pool, contributed to the bioinformatics analysis and drafted the initial manuscript. LW provided the parental materials. PWB revised and edited the manuscript. GZ and RZ made the qRT-PCR analysis. All authors read and approved the final manuscript.

Funding

This research was supported by the National Natural Sciences Foundation of China (31560555, 31760572). The part of high throughput sequencing and data analysis were supported by the National Natural Sciences Foundation of

China (31560555), and the qRT-PCR were commonly supported by the National Natural Sciences Foundation of China (31560555, 31760572).

Availability of data and materials

All data generated and analyses done within this study are included in this published article and its supplementary information files. The raw data can be found in NCBI (<https://www.ncbi.nlm.nih.gov/>) with an accession number of SRA895207.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Horticulture, Gansu Agricultural University, Lanzhou 730070, China. ²Vegetable Institute, Gansu Academy of Agricultural Sciences, Lanzhou 730070, China. ³College of Agriculture, Consumer, and Environmental Sciences, New Mexico State University, Las Cruces 88001, USA.

Received: 14 March 2019 Accepted: 22 October 2019

Published online: 11 November 2019

References

- Bosland PW, Votava EJ. Peppers. In: Vegetable and spice capsicums. 2nd ed. Wallingford: CAB International; 2012. p. 230.
- Colombo N, Galmarini CR. The use of genetic, manual and chemical methods to control pollination in vegetable hybrid seed production: a review. *Plant Breeding*. 2017;136:287–99.
- Kaul MLH. Male Sterility in Higher Plants. Berlin, Heidelberg: Springer-Verlag; 1988.
- Ji JJ, Huang W, Yin YX, Li Z, Gong ZH. Development of a SCAR marker for early identification of S-cytoplasm based on mitochondrial SRAP analysis in pepper (*Capsicum annuum* L.). *Mol Breeding*. 2014;33:679–90.
- Havey MJ. The Use of Cytoplasmic Male Sterility for Hybrid Seed Production. In: Daniell H, Chase C, editors. *Molecular Biology and Biotechnology of Plant Organelles*. Dordrecht: Springer; 2004. p. 623–34.
- Kumar S, Singh V, Singh M, Rai S, Kumar S, Rai SK, Rai M. Genetics and distribution of fertility restoration associated RAPD markers in pepper (*Capsicum annuum* L.). *Sci Hortic*. 2007;111:197–202.
- Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell, Supplement*. 2004;16:S154–S69.
- Chase CD. Cytoplasmic male sterility: A window to the world of plant mitochondrial-nuclear interactions. *Trends Genet*. 2007;23:81–90.
- Dieterich JH, Braun HP, Schmitz UK. Alloplasmic male sterility in *Brassica napus* (CMS 'Tournfortii-Stiewe') is associated with a special gene arrangement around a novel *atp9* gene. *Mol Genet Genomics*. 2003;269:723–31.
- Jing B, Heng S, Tong D, Wan Z, Fu T, Tu J, Ma CZ, Yi B, Wen J, Shen JX. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *J Exp Bot*. 2012;63:1285–95.
- Peng XJ, Li FH, Li SQ. Expression of a mitochondrial gene orfH 79 from the CMS-HongLian rice inhibits *Saccharomyces cerevisiae* growth and causes excessive ROS accumulation and decrease in ATP. *Biotechnol Lett*. 2009;31:409–14.
- Tuteja R, Saxena RK, Davila J, Shah T, Chen W, Xiao YL, Fan G, Saxena KB, Alverson AJ, Spillane C, Town C, Varshney RK. Cytoplasmic male sterility-associated chimeric open reading frames identified by mitochondrial genome sequencing of four cajanus genotypes. *DNA Res*. 2013;20:485–95.
- Peterson PA. Cytoplasmically inherited male sterility in *Capsicum*. *Amer Natl*. 1958;92:111–9.
- Swamy BN, Hedau NK, Chaudhari GV, Kant L, Pattanayak A. CMS system and its stimulation in hybrid seed production of *Capsicum annuum* L. *Sci Hortic*. 2017;222:175–9.
- Kim DH, Kang JG, Kim BD. Isolation and characterization of cytoplasmic male sterility-associated orf456 gene of chili pepper (*Capsicum annuum* L.). *Plant Mol Biol*. 2007;63:519–32.
- Kim DS, Kim BD. The organization of mitochondrial *atp6* gene region in male fertile and CMS lines of pepper (*Capsicum annuum* L.). *Curr Genet*. 2006;49:59–67.
- Gulyas G, Shin Y, Kim H, Lee JS, Hirata Y. Altered transcript reveals an orf507 sterility-related gene in chili pepper (*Capsicum annuum* L.). *Plant Mol Biol Rpt*. 2010;28:605–12.
- Chen LT, Liu YG. Male sterility and fertility restoration in crops. *Annu Rev Plant Biol*. 2014;65:579–606.
- Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu Y. 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*. 2006;18:676–87.
- Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, Hiard S, Gherbi H, Bendahmane A, Budar F, Mireau H. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by a fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell*. 2008;20:3331–45.
- Cui X, Wise RP, Schnable PS. The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Sci*. 1996;272:1334–6.
- Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male sterile plants. *Proc Natl Acad Sci USA*. 2002;99:10887–192.
- Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J*. 2003;35:262–72.
- Deslorie S, Gherbi H, Laloui W, Maradour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small ID, Caboche M, Delourme R, Bendahmane A. Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rpt*. 2003;4:588–94.
- Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J. Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosenada radish. *Plant J*. 2003;34:407–15.
- Wang Z, Wang C, Mei S, Gao L, Zhou Y, Wang T. An insertion-deletion at a pentatricopeptide repeat locus linked to fertility transition to cytoplasmic male sterility in radish (*Raphanus sativus* L.). *Mol Breeding*. 2015;25:1–5.
- Matsuhira H, Kagami H, Kurata M, Kitazaki K, Matsunaga M, Hamaguchi Y, Hagihara E, Ueda M, Harada M, Muramatsu A, Yui-Kurino R, Taguchi K, Tamagake H, Mikami T, Kubo T. Unusual and typical features of a novel restorer-of-fertility gene in sugar beet (*Beta vulgaris* L.). *Genet*. 2012;192:1347–58.
- Kitazaki K, Arakawa T, Matsunaga M, Yui-Kurino R, Matsuhira H, Mikami T, Kubo T. Post-translational mechanisms are associated with fertility restoration of cytoplasmic male sterility in sugar beet. *Plant J*. 2015;83:290–9.
- Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the collinear region of rice chromosome 12. *Theor Appl Genet*. 2005;111:994–1012.
- Hu J, Huang W, Huang Q, Qin X, Yu C, Wang L, Li S, Zhu R, Zhu Y. Mitochondria and cytoplasmic male sterility in plants. *Mitochondrion*. 2014;19:282–8.
- Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, Takahashi H, Mori K, Fujimura T. Positional cloning of the rice *rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeted PPR protein. *Theor Appl Genet*. 2004;108:1449–57.
- Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *Plant J*. 2004;37:315–25.
- Fujii S, Toriyama K. Suppressed expression of RETROGRADE_REGULATED MALE STERILITY restores pollen fertility in cytoplasmic male sterile rice plants. *Proc Natl Acad Sci USA*. 2009;106:9513–8.
- Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K. The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *Plant J*. 2011;65:359–67.
- Fujii S, Bond CS, Small ID. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci USA*. 2011;108:1723–8.

36. Dahan J, Mireau H. The *Rf* and *Rf-like* PPR in higher plants, a fast evolving subclass of PPR genes. *RNA Biol.* 2013;10:1286–93.
37. Jo YD, Ha Y, Lee JH, Park M, Bergsma AC, Choi HI, Goritschnig S, Kloosterman B, van Dijk PJ. Fine mapping of Restorer-of-fertility in pepper (*Capsicum annuum* L.) identified a candidate gene encoding a pentatricopeptide repeat (PPR)-containing protein. *Theor Appl Genet.* 2016;129: 2003–17.
38. Barchenger DW, Said JI, Zhang Y, Song M, Ortega FA, Ha Y, Kang BC, Bosland PW. Genome-wide identification of chile pepper pentatricopeptide repeat domains provides insight into fertility restoration. *J Amer Soc Hort Sci.* 2018;143:418–29.
39. Novak F, Betlach J, Dubovsky J. Cytoplasmic malesterility in sweet pepper (*Capsicum annuum* L.) I. Phenotype and inheritance of male sterile character. *Z Pflanzenzüchtung.* 1971;65:129–40.
40. Wei B, Wang L, Chen L, Zhang R. Genetic analysis on the restoration of cytoplasmic male sterility with mixed model of major gene plus polygene in pepper. *Acta Hort Sinica.* 2013;40:2263–8.
41. Wei B, Wang L, Zhang R, Zhang J. Identification of two major quantitative trait loci restoring the fertility of cytoplasmic male sterility in *Capsicum annuum*. *J Agric Biot.* 2017;25:43–9.
42. Wang LH, Zhang BX, Lefebvre V, Huang SW, Daubèze AM, Palloix A. QTL analysis of fertility restoration in cytoplasmic male sterile pepper. *Theor Appl Genet.* 2004;109:1058–63.
43. Lee J, Yoon JB, Park HG. Linkage analysis between the partial restoration (*pr*) and the restorer-of-fertility (*Rf*) loci in pepper cytoplasmic male sterility. *Theor Appl Genet.* 2008;117:383–9.
44. Shiffriss C. Male sterility in pepper (*Capsicum annuum* L.). *Euphytica.* 1997;93:83–8.
45. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10:57–63.
46. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregate analysis – a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA.* 1991;88:9828–32.
47. Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, Song L, Feng J, Liu S, Liu Z. Bulk segregate RNA-seq reveals expression and positional candidate genes and allele specific expression for disease resistance against enteric septicemia of catfish. *BMC Genomics.* 2013;14:929.
48. Du H, Zhu J, Su H, Huang M, Wang H, Ding S, Zhang B, Luo A, Wei S, Tian X, Xu Y. Bulk segregate RNA-seq reveals differential expression and SNPs of candidate genes associated with water logging tolerance in maize. *Front Plant Sci.* 2017;8:1022.
49. Castro AJ, Clement C. Sucrose and starch catabolism in the anther of *Lilium* during its development: a comparative study among the anther wall, locular fluid and microspore/pollen fractions. *Planta.* 2007;225:1573–82.
50. Rolland F, Moore B, Sheen J. Sugar sensing and signaling in plants. *The Plant Cell.* 2002;14:5185–205.
51. Preuss D, Lemieux B, Yen G, Davis R. A conditional sterile mutation eliminates surface components from *Arabidopsis* pollen and disrupts cell signaling during fertilization. *Genes Dev.* 1993;7:974–85.
52. Rhee SY, Osborne E, Poindexter PD, Somerville CR. Microspore separation in the quartet mutant of *Arabidopsis* is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. *Plant Physiol.* 2003;133:1170–80.
53. Micheli F. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* 2001;6:414–9.
54. Palusa SG, Golovkin M, Shin SB, Richardson DN, Reddy ASN. Organ-specific, developmental, hormonal and stress regulation of expression of putative pectate lyase genes in *Arabidopsis*. *New Phytol.* 2007;174:537–50.
55. Francis KE, Lam SY, Copenhaver GP. Separation of *Arabidopsis* pollen tetrads is regulated by QUARTET1, a pectin methylesterase gene. *Plant Physiol.* 2006;142:1004–13.
56. Hamid R, Tomar RS, Marashi H, Shafaroudi SM, Golakiya BA, Mohsenpour M. Transcriptome profiling and cataloging differential gene expression in floral buds of fertile and sterile lines of cotton (*Gossypium hirsutum* L.). *Gene.* 2018;660:80–91.
57. Persson S, Paredes A, Carroll A, Palsdottir H, Doblin M, Poindexter P, Khitrov N, Auer M, Somerville CR. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in *Arabidopsis*. *Proc Natl Acad Sci USA.* 2007;104:15566–71.
58. Wang W, Wang L, Chen C, Xiong G, Tan XY, Yang KZ, Wang ZC, Zhou Y, Ye D, Chen LQ. *Arabidopsis* CSLD1 and CSLD4 are required for cellulose deposition and normal growth of pollen tubes. *J Exp Bot.* 2011;62:5161–77.
59. Xue H, Chen X, Mei Y. Function and regulation of phospholipid signaling in plants. *Biochem J.* 2009;421:145–56.
60. Hamamura Y, Nagahara S, Higashiyama T. Double fertilization on the move. *Curr Opin Plant Biol.* 2012;15:70–7.
61. Chapman LA, Goring DR. Misregulation of phosphoinositides in *Arabidopsis thaliana* decreases pollen hydration and maternal fertility. *Sex Plant Reprod.* 2011;24:319–26.
62. Kost B, Lemichez E, Spielhofer P, Hong Y, Tolia K, Carpenter C, Chua NH. Rac homologues and compartmentalized phosphatidylinositol 4,5-bisphosphate act in a common pathway to regulate polar pollen tube growth. *J Cell Biol.* 1999;145:317–30.
63. Ischebeck T, Stenzel I, Heilmann I. Type B phosphatidylinositol-4-phosphate 5-kinases mediate *Arabidopsis* and *Nicotiana tabacum* pollen tube growth by regulating apical pectin secretion. *Plt Cell.* 2008;20:3312–30.
64. Sousa E, Kost B, Malho R. *Arabidopsis* phosphatidylinositol-4-monophosphate 5-kinase 4 regulates pollen tube growth and polarity by modulating membrane recycling. *Plant Cell.* 2008;20:3050–64.
65. Ugalde JM, Rodriguez-Furlán C, Rycke R, Norambuena L, Friml J, León G, Tejos R. Phosphatidylinositol 4-phosphate 5-kinases 1 and 2 are involved in the regulation of vacuole morphology during *Arabidopsis thaliana* pollen development. *Plant Sci.* 2016;250:10–9.
66. Franklin-Tong VE. Signaling and the modulation of pollen tube growth. *Plant Cell.* 1999;11:727–38.
67. Paxson-Sowders DM, Owen HA, Makaroff CA. A comparative ultrastructural analysis of exine pattern development in wild-type *Arabidopsis* and a mutant defective in pattern formation. *Protoplasma.* 1996;198:53–65.
68. Ahlers F, Thom I, Lambert J, Kuckuk R, Wiermann R. 1H NMR analysis of sporopollenin from *Typha angustifolia*. *Phytochem.* 1999;50:1095–8.
69. Jiang J, Zhang Z, Cao J. Pollen wall development: the associated enzymes and metabolic pathways. *Plant Biol.* 2013;15:249–63.
70. Liu L, Fan XD. Tapetum: regulation and role in sporopollenin biosynthesis in *Arabidopsis*. *Plant Mol Biol.* 2013;83:165–75.
71. Ji J, Yang L, Fang Z, Zhuang M, Zhang Y, Lv H, Liu Y, Li Z. Complementary transcriptome and proteome profiling in cabbage buds of a recessive male sterile mutant provides new insights into male reproductive development. *J Proteomics.* 2018;179:80–91.
72. hoi H, Jin JY, Choi S, Hwang JU, Kim YY, Suh MC, Lee Y. An ABCG/WBC-type ABC transporter is essential for transport of sporopollenin precursors for exine formation in developing pollen. *Plant J.* 2011;65:181–93.
73. Quilichini TD, Friedmann MC, Samuels AL, Douglas CJ. ATP-binding cassette transporter G26 is required for male fertility and pollen exine formation in *Arabidopsis*. *Plant Physiol.* 2010;154:678–90.
74. Wang K, Gao F, Ji Y, Liu Y, Dan Z, Yang P, Zhu Y, Li S. ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. *New Phytol.* 2013;198:408–18.
75. Nie Z, Zhao T, Yang S, Gai J. Development of a cytoplasmic male-sterile line NJCMS4A for hybrid soybean production. *Plant Breeding.* 2017;136:516–25.
76. Yang Y, Bao S, Zhou X, Liu J, Zhuang Y. The key genes and pathways related to male sterility of eggplant revealed by comparative transcriptome analysis. *BMC Plant Biol.* 2018;18:209–12.
77. Meech R, Mubarakah N, Shivasami A, Rogers A, Nair PC, Hu DG, McKinnon RA, Mackenzie PI. A novel function for UDP Glycosyltransferase 8: Galactosidation of bile acids. *Mol Pharmacol.* 2015;87:442–50.
78. Li X, Zhu B, Gao X, Liang P. Over-expression of UDP-glycosyltransferase gene *UGT2B17* is involved in chlorantraniliprole resistance in *Plutella xylostella* (L.). *Pest Manag Sci.* 2017;73:1402–9.
79. Qin C, Yub C, Shena Y, Fang X, Chen L, Mind J, Cheng J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc Natl Acad Sci USA.* 2014; 111:5135–40.
80. Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14:R36.
81. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28:511–5.
82. Rogers MF1, Thomas J, Reddy AS, Ben-Hur A. SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. *Genome Biology.* 2012;13:R4.

83. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*. 2013;29:1035–43.
84. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11:R14.
85. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36:480–4.
86. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics*. 2005;21:3787–93.
87. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*. 2001;25:402–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

