

RESEARCH ARTICLE

Open Access



Analysis of metabolic pathways related to fertility restoration and identification of fertility candidate genes associated with *Aegilops kotschyi* cytoplasm in wheat (*Triticum aestivum* L.)

Sha Li, Zihan Liu, Yulin Jia, Jiali Ye, Xuotong Yang, Lingli Zhang and Xiyue Song* 

Abstract

Background: Thermo-sensitive male-sterility based on *Aegilops kotschyi* cytoplasm (K-TCMS) plays an important role in hybrid wheat breeding. This has important possible applications in two-line hybrid wheat breeding but the genetic basis and molecular regulation mechanism related to fertility restoration are poorly understood. In this study, comparative transcriptome profiling based on RNA sequencing was conducted for two near-isogenic lines comprising KTM3315R and its sterile counterpart KTM3315A, a total of six samples (3 repetitions per group), in order to identify fertility restoration genes and their metabolic pathways.

Results: In total, 2642 significant differentially expressed genes (DEGs) were detected, among which 1238 were down-regulated and 1404 were up-regulated in fertile anthers. Functional annotation enrichment analysis identified important pathways related to fertility restoration, such as carbohydrate metabolism, phenylpropanoid metabolism and biosynthesis, as well as candidate genes encoding pectin methylesterase and flavanone 3-hydroxylase. Moreover, transcription factor analysis showed that a large number of DEGs were mainly involved with the WRKY, bHLH, and MYB transcription factor families. Determination of total soluble sugar and flavonoid contents demonstrated that important metabolic pathways and candidate genes are associated with fertility restoration. Twelve DEGs were selected and detected by quantitative reverse-transcribed PCR, and the results indicated that the transcriptome sequencing results were reliable.

Conclusions: Our results indicate that identified DEGs were related to the fertility restoration and they proved to be crucial in *Aegilops kotschyi* cytoplasm. These findings also provide a basis for exploring the molecular regulation mechanism associated with wheat fertility restoration as well as screening and cloning related genes.

Keywords: Comparative transcriptome, Cytoplasmic male sterility, Flavanone 3-hydroxylase, Pectin methylesterase, Wheat (*Triticum aestivum* L.), WRKY transcription factor

* Correspondence: songxiyue@126.com

College of Agronomy, Northwest A&F University, Yangling, Shaanxi, China



Background

Heterosis is one of the most successful strategies for increasing crop yields and it has been exploited widely in plant breeding systems [1]. In this context, cytoplasmic male sterility (CMS) can create sterile male gametophytes without affecting the agronomic performance and this cost-effective system facilitates hybrid seed production [2, 3]. In plants, CMS is due to the maternally inherited inability to establish genome coordination between the organelles (mitochondria) and nuclear genomes, thereby resulting in non-functional pollen production [4–7]. Due to the CMS trait, the requirement for the manual removal of anthers can be eliminated, thereby reducing the needs in terms of labor and material resources, as well as allowing the hybridization technique to produce excellent F_1 generations that are significantly superior to the parent and available popular breeds in terms of their yield, stress resistance, and adaptability [8]. CMS has been applied in crops such as rice [9], maize [10], sorghum [11], soybean [12], and cotton [13] with remarkable results. However, due to its huge genome and long breeding cycle, the use of CMS in wheat production is still a challenge.

The use of CMS in wheat has important implications for simplifying breeding programs and increasing crop yields and quality. At present, the thermo-sensitive CMS wheat lines with *Aegilops kotschy* cytoplasm (K-TCMS) based on the two-line method have advantages as simple restorer and maintenance lines, with no negative cytoplasmic effects [14]. These lines perform well in hybrid wheat breeding and they can produce a large number of hybrid seeds, which is a valuable benefit for the production of hybrid wheat [15, 16].

In recent years, high-throughput sequencing has provided a new approach for transcriptome sequencing and it is especially helpful for conducting comparative analyses of RNA sequencing data obtained from non-model species with limited genomic information. Comparative transcriptome analysis has been successfully employed to study the molecular mechanism of male sterility at different developmental stages in different species. For example, the candidate genes and important metabolic pathways have been studied in cotton [17, 18], *Brassica napus* [19, 20], rice [21], cabbage [22], sesame [23], and other species. However, few comparative transcriptome studies of fertility restoration in wheat have been conducted.

In order to improve the utilization of wheat heterosis and facilitate agricultural development, it is important to elucidate the molecular mechanism responsible for fertility restoration in CMS lines. In the present study, anthers from the binucleate stage of the K-TCMS wheat line KTM3315A and its near-isogenic restoring line KTM3315R were used as experimental materials for RNA sequencing to identify important candidate genes

and the biological pathways related to fertility restoration. In particular, we aimed to: obtain novel insights into the molecular regulation mechanism associated with fertility restoration in wheat; and provide a foundation for screening and further cloning genes related to fertility restoration in subsequent studies.

Results

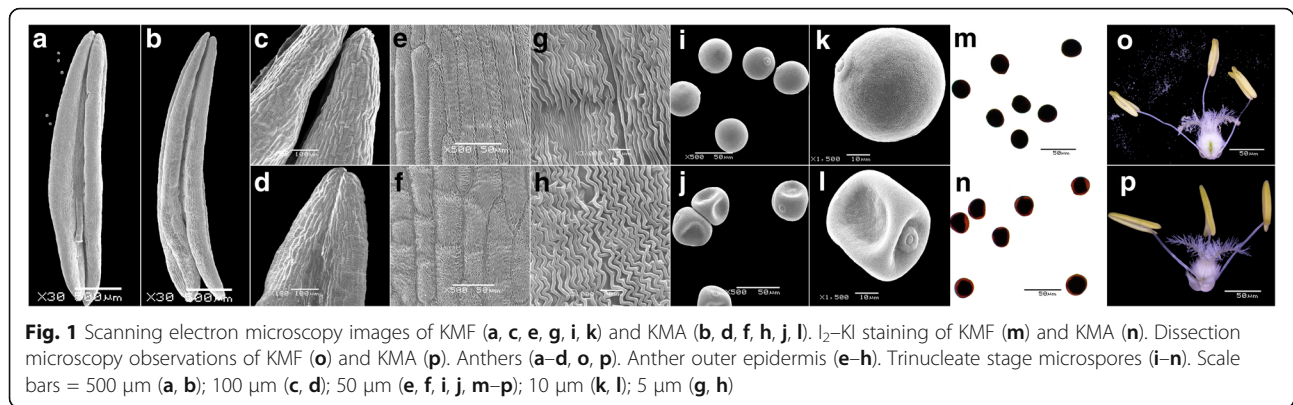
Phenotypic traits and microscopic observations

In order to determine the visual differences between the fertile and sterile wheat lines, we observed the morphology of anthers, anther outer epidermis, and microspores in the fertile wheat line KTM3315R and sterile wheat line KTM3315A during the trinucleate stage (Fig. 1). Scanning electron microscopy showed that, compared with the KTM3315A anthers, the KTM3315R anthers were plump, symmetrical, and regular, and the upper ends of the anthers were cracked and they released mature pollen grains (Fig. 1a–d). Dissection microscopy also clearly demonstrated these features (Fig. 1o, p). Observations of the outer epidermis of the anthers (Fig. 1e–h) showed that the folds formed by the KTM3315R anthers were regular and flat (Fig. 1g), whereas they were irregular in KTM3315A (Fig. 1h). In addition, the mature pollen grains of KTM3315A were shrunken and deformed, whereas those of KTM3315R were plump and spherical (Fig. 1i–l). Cytological observations of the mature pollen grains were also obtained by I_2 -KI staining, which showed that the mature pollen grains stained completely black in KTM3315R, whereas the mature pollen grains at the periphery were not stained in KTM3315A and the staining was uneven, which is the typical staining pattern for abortive pollen (Fig. 1m, n). According to these observations, there were clear differences in the fertility phenotype between KTM3315A and KTM3315R due to the induction of fertility restoration genes.

Transcriptome sequencing and genome mapping

Based on the phenotypic traits observed for the anthers and mature pollen grains, we hypothesized that the gene expression pattern had changed. To validate our hypothesis and further elucidate the molecular mechanisms of regulating fertility, we employed an Illumina HiSeq PE150 sequencer for high-throughput sequencing of the fertile and sterile anthers from KTM3315R (designated as KMF) and KTM3315A (designated as KMA), a total of six samples (3 repetitions per group, including KMA1, KMA2, KMA3; KMF1, KMF2, KMF3, respectively), at the binucleate stage.

In total, 44 GB of clean data were produced. The data were filtered for each sample and an average of 7.3 GB of high quality clean data was obtained where the Q20 (sequencing error rate, 1%) base percentage exceeded 95.69%. All of the bases were identified and the CG



content ranged from 59.15 to 61.62%. The average read length of each sequence was 150 bp. For each sample, the high quality clean reads were filtered using the rRNA database and aligned with the reference genome, where the alignment efficiency was between 63.11 and 68.59% (Additional file 1: Table S1). After directly comparing the abundances and discrete distributions of the expression levels for different sample genes, we found that the sequencing quality and gene expression levels were basically the same (Fig. 2a, b). Thus, the results indicated that the quality of the data obtained by sequencing was reliable and suitable for subsequent analyses.

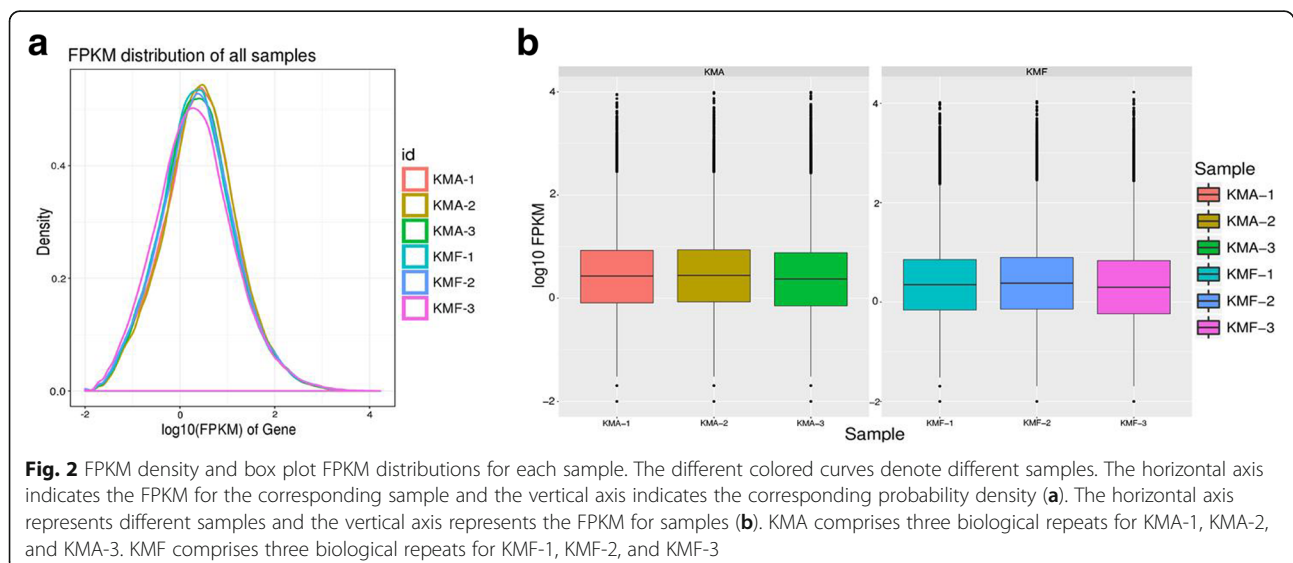
Identification of DEGs

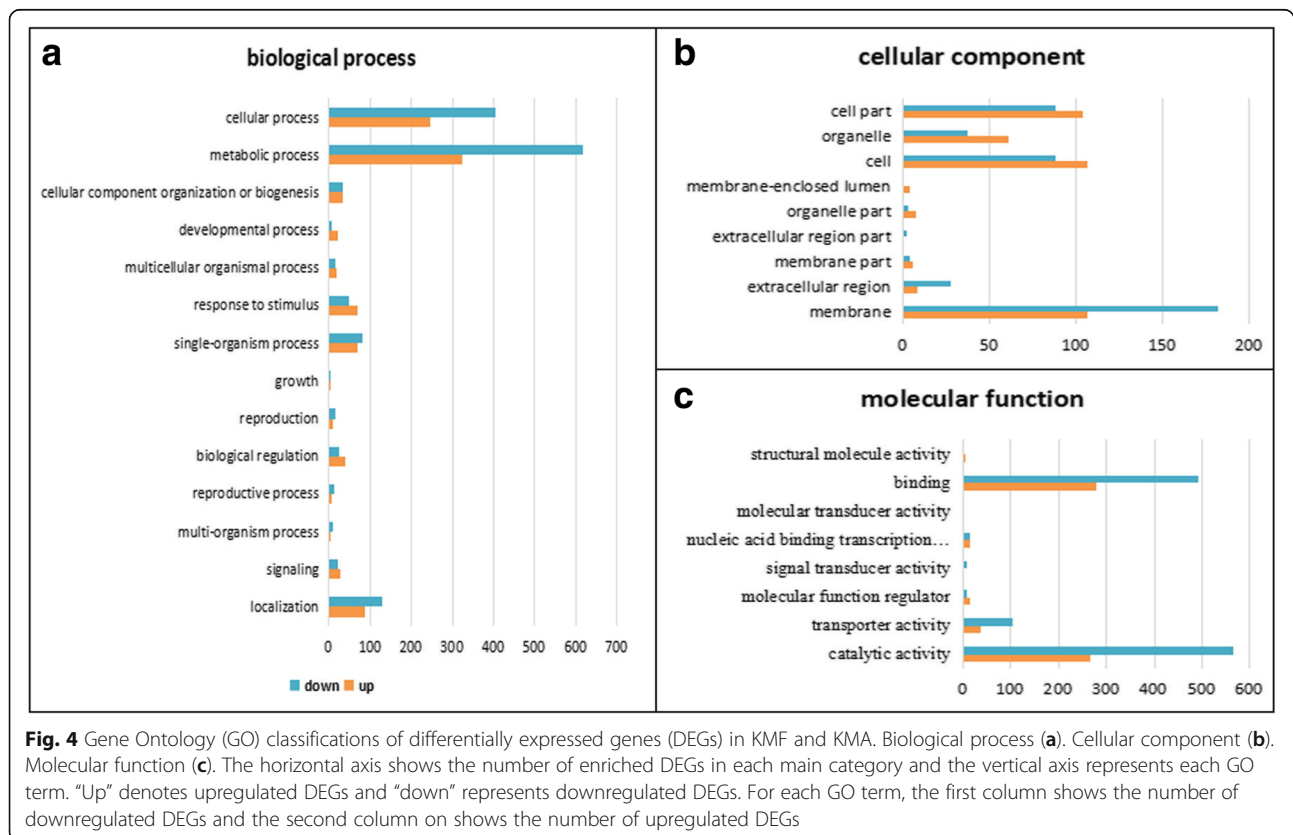
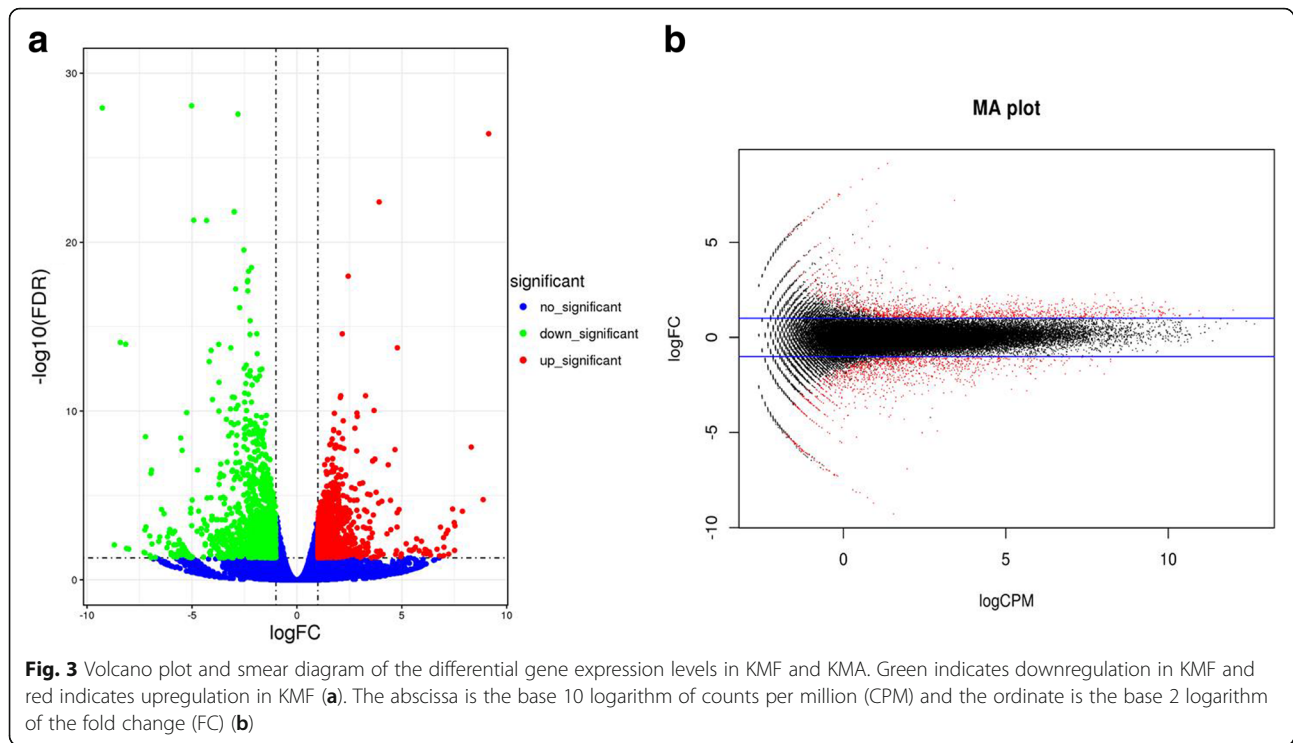
To identify differentially expressed genes across groups, the edgeR package was used [24]. We identified genes with $\log_2(\text{fold change (FC)}) \geq 1$ and the false discovery rate (FDR) < 0.05 in a comparison as significant DEGs. In total, 2642 DEGs were identified in the KMF and KMA groups (Additional file 2: Table S2), where 1238 were down-regulated and 1404 were up-regulated in

KMF compared with KMA. We used a volcano plot and a smear diagram to visualize the significant differences in DEGs (Fig. 3a, b). The subsequent bioinformatics analyses were conducted based on these DEGs.

Functional classification of DEGs

To determine the major biological functions of differentially expressed genes, 2642 DEGs were sequentially aligned into the Gene Ontology (GO), Eukaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Using the hypergeometric test, the $q\text{-value} \leq 0.05$ obtained after the $P\text{-value}$ is corrected by multiple tests is used as the threshold, and the GO term and path which are significantly enriched in the DEGs are obtained. Due to the large numbers and the complex branch structure of GO categories, only the three most abundant functional groups, namely “Biological Process”, “Cellular Component” and “Molecular Function” were presented, as an example (Fig. 4a–c, Additional file 3: Table S3-1). In the sub-category of “Biological Process”, “metabolic process (GO: 0008152)” and “cellular process



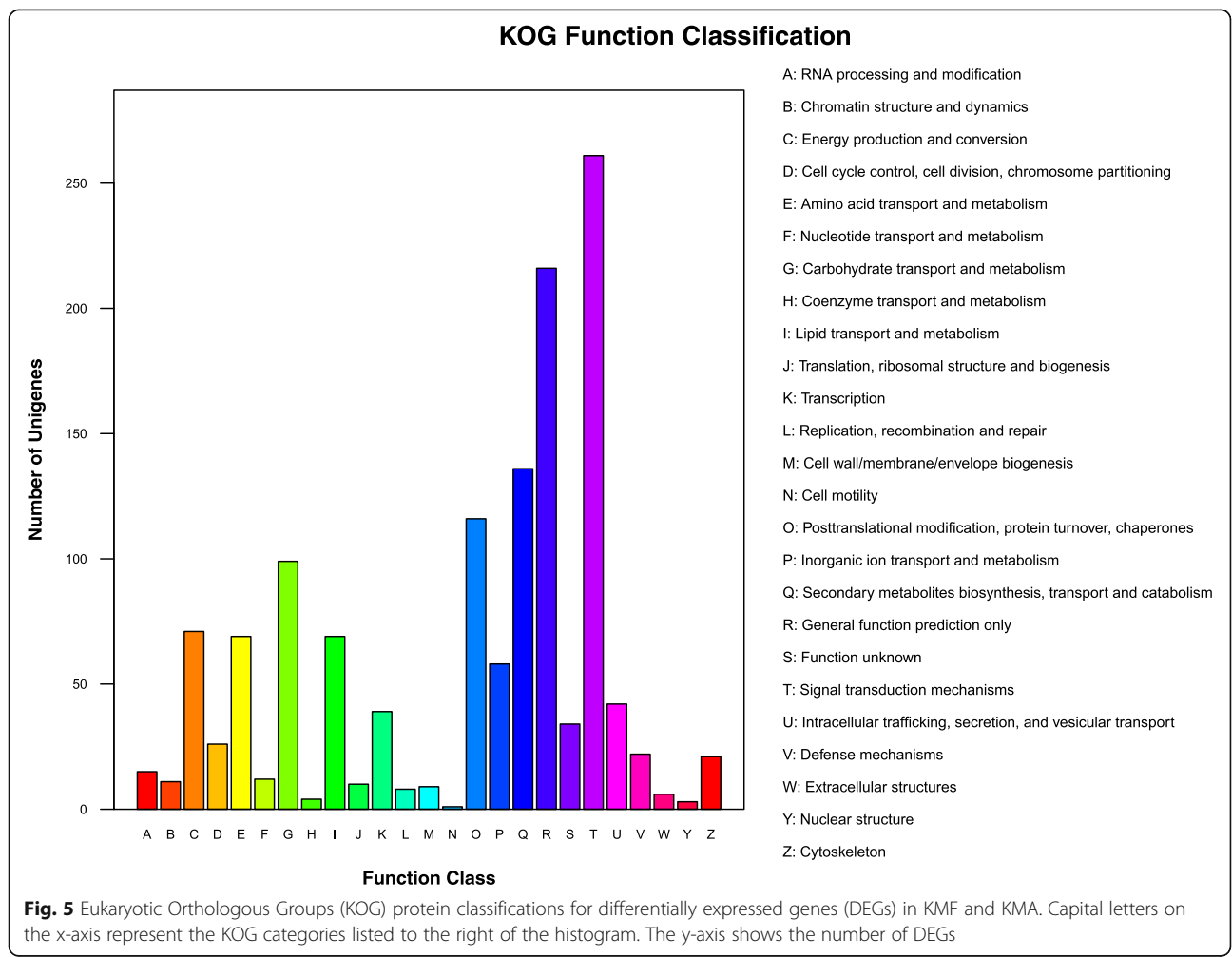


(GO: 0009987)” were the most abundant cascades that have a respective of 942 and 649 DEGs. In these two sub-categories, “catabolic processes (GO: 0009056)”, “secondary metabolic processes (GO: 0019748)” and “cell death (GO: 0008219)” were significantly enriched (Additional file 3: Table S3–2). In the next main sub-category of “Cellular Component”, “membrane (GO: 0016020)” and “cell (GO: 0005623)” were the most abundant cascades that have a respective of 288 and 194 DEGs, respectively. In the “cell” category, DEGs were significantly enriched for “cell wall (GO: 0005618)” (Additional file 3: Table S3–2). Within the last sub-category “Molecular Function”, “catalytic activity (GO:0003824)” and “binding (GO:0005488)” were the most abundant cascades that have 833 and 771 DEGs, respectively, where the number of down-regulated DEGs was almost two times that of up-regulated DEGs. In these two sub-categories, “kinase activity (GO: 0016301)”, “hydrolase activity (GO: 0016787)” and “Carbohydrate binding (GO: 0030246)” were significantly enriched (Additional file 3: Table S3–2). Based on these significantly enriched terms,

we predict that they are closely related to the fertility restoration of wheat.

According to the KOG annotated results, 1198 DEGs were annotated into 25 KOG categories (Fig. 5, Additional file 4: Table S4). “Signal transduction mechanisms” was the largest group (group T with 261 DEGs), followed by “General function prediction only” (group R with 216 DEGs) and “Secondary metabolites biosynthesis, transport and catabolism” (group Q with 136 DEGs). “Post-translational modification, protein turnover, chaperones” (group O with 116 DEGs), “Carbohydrate transport and metabolism” (group G with 99 DEGs), “Energy production and conversion” (group C with 71 DEGs), “Amino acid transport and metabolism” (group E with 69 DEGs), and “Lipid transport and metabolism” (group I, with 69 DEGs) were the other main categories. Therefore, the eight categories highlighted above may be important for understanding fertility.

In the KEGG analysis, a total of 438 DEGs were annotated and classified into 102 pathways. Metabolism (64%,

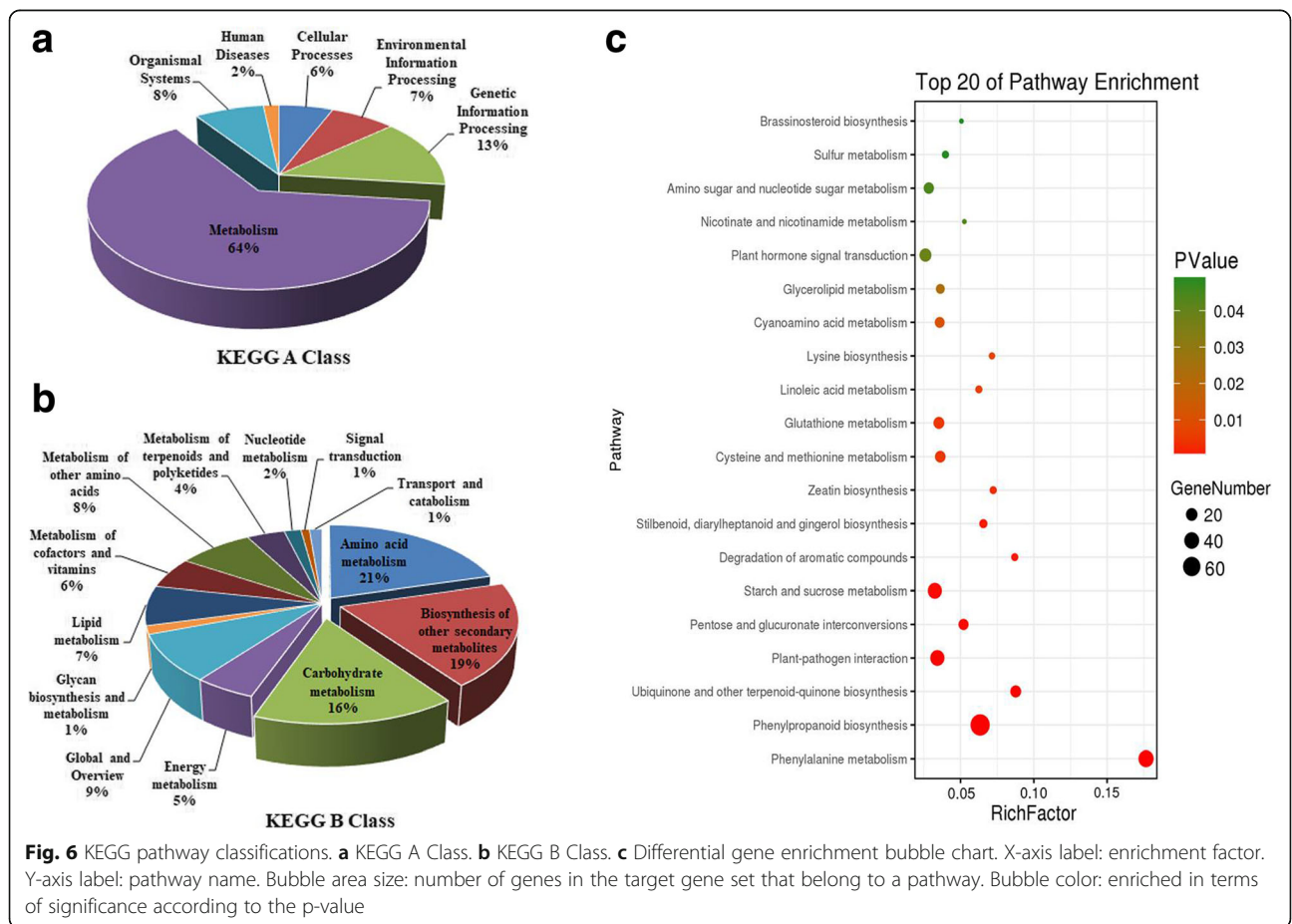


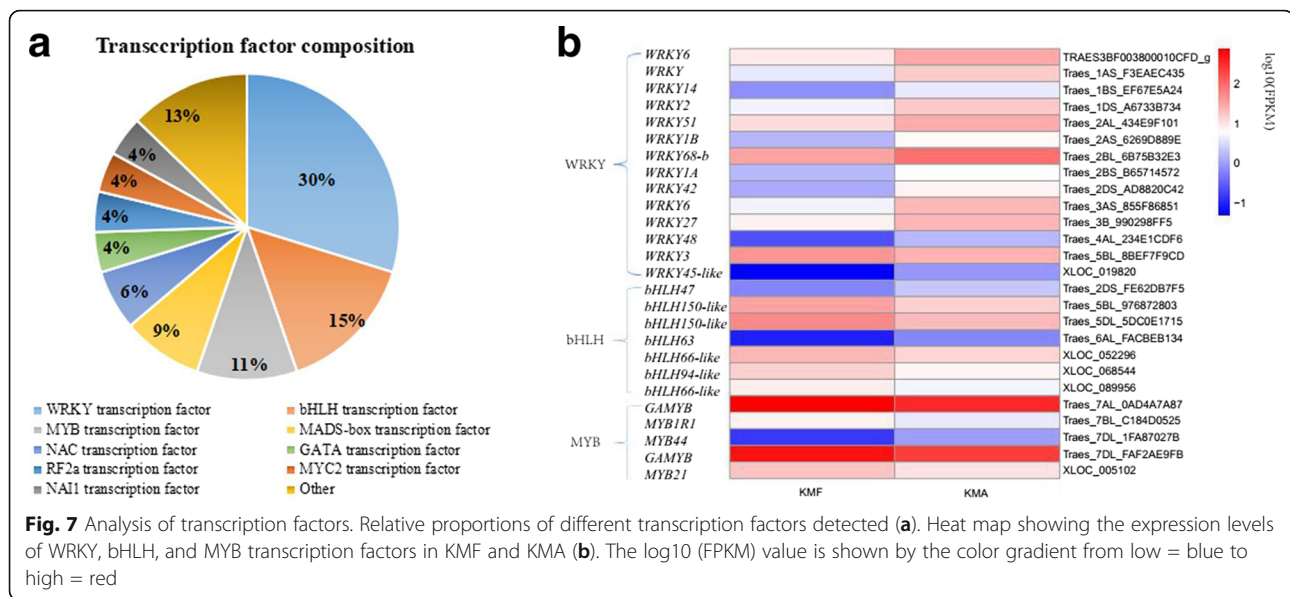
296 DEGs) in classification A categories accounted for most of the DEGs, followed by B categories, where they were mainly enriched in terms of amino acid metabolism (21%, 93 DEGs), biosynthesis of other secondary metabolites (19%, 87 DEGs), and carbohydrate metabolism (16%, 71 DEGs) (Fig. 6a, b). In the amino acid metabolism category, phenylalanine metabolism (ko00360) and cysteine and methionine metabolism (ko00270) were enriched most with 42 and 16 DEGs, respectively. In the biosynthesis of other secondary metabolites category, phenylpropanoid biosynthesis (ko00940) was enriched most with 77 DEGs. In the carbohydrate metabolism category, significant enrichment was determined for starch and sucrose metabolism (ko00500) and pentose and glucuronate interconversion (ko00040) with 34 and 14 DEGs, respectively. A high-level bubble diagram was prepared for the top 20 pathways where $p \leq 0.05$ in order to visualize the significant pathways (Fig. 6c, Additional file 5: Table S5). Based on the KOG annotations and the analysis given above, we identified four major pathways: phenylalanine metabolism (ko00360), phenylpropanoid biosynthesis (ko00940), starch and sucrose metabolism (ko00500), and pentose and glucuronate interconversion (ko00040). The metabolic pathways associated with

these pathways could facilitate further analyses of the genes and molecular mechanisms associated with fertility restoration in K-TCMS lines.

Identification of transcription factors related to fertility restoration

The regulation of plant gene expression comprises a complex network. Transcription factors can regulate structural genes related to specific traits, and thus their effects on plant growth and development may be greater than those of structural genes. Based on previous studies of transcription factors [25–27] and the annotation information obtained from the Nr database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>), 47 DEGs were predicted to encode transcription factors, which were divided into nine families comprising WRKY, bHLH, MYB, MADS-box, NAC, GATA, RF2a, MYC2, and NAI1 transcription factors. The WRKY transcription factors were most common and they were encoded by 14 DEGs, which accounted for 30% of the total, followed by bHLH transcription factors (seven DEGs, 15%) and MYB transcription factors (five DEGs, 11%) (Fig. 7a). Analyses of the expression levels of these three major transcription factor families in KMF and KMA detected similar trends (Fig. 7b, Additional file 6: Table S6).





The heat map showed that excluding the upregulation of *WRKY3* in KMF compared with KMA, the other 13 DEGs were downregulated. *bHLH47*, *bHLH63*, and *MYB44* were downregulated in KMF, but the remaining transcription factors were upregulated, i.e., bHLH and MYB. These findings suggest that WRKY, bHLH, and MYB family transcription factors may affect fertility restoration in K-TCMS lines.

Carbohydrate metabolism and Phenylpropanoid biosynthesis pathways involved with the regulation of pollen development with *Aegilops kotschy* cytoplasm

According to the analyses based on KOG and KEGG, a large number of DEGs were mainly enriched in the carbohydrate metabolism and phenylpropanoid biosynthesis and metabolic pathways. To clarify whether these DEGs regulate pollen development via these pathways, we mapped the metabolic pathways regulated by some of the key DEGs in the four key pathways identified previously (Fig. 8a). In the pollen maturation process, sugars provide the basis for pollen growth and development. In the sugar metabolism process, sucrose is used as a substrate to produce sucrose 6-phosphate and β -D-fructose is produced under catalysis by hydrolase beta-fructofuranosidase (INV, EC: 3.2.1.26). β -D-Fructose is then used to produce β -D-Fructose-6P under catalysis by transferase fructokinase (scrK, EC: 2.7.1.4) and hexokinase (HK, EC: 2.7.1.1), which is catalyzed by a series of enzymes to finally produce uridine diphosphate-glucose (UDP-glucose). UDP-glucose is an important intermediate in monosaccharide interconversion or the formation of uranic acid, and it plays a central role in carbohydrate metabolism. UDP-glucose is catalyzed by a series of enzymes to produce pectin. Pectin is hydrolyzed by pectin methyl-esterase (PME, EC: 3.1.1.11) to form D-galacturonate and

maintain cell wall development. The expression levels of these key enzymes in fertile and sterile anthers were shown in the heat map in Fig. 8b (Additional file 7: Table S7). *Traes_1AL_1C66B5A24*, *Traes_1AL_6F271A0E5*, *Traes_1BL_555D3DF55*, and *Traes_1DL_D313 CC8EE* were downregulated in fertile anthers but the other enzymes encoded by DEGs were upregulated in fertile anthers and downregulated in infertile anthers. Therefore, we hypothesized that the upregulated expression of these enzymes can maintain normal pollen growth and development. It should be noted that many DEGs encoded PME with roles during sugar metabolism. In the phenylpropanoid metabolism and biosynthetic pathway, phenylalanine is used as a substrate to form cinnamic acid under the action of phenylalanine ammonia-lyase (PAL, EC: 4.3.1.24). Cinnamic acid finally produces P-cinnamoyl-CoA under the actions of oxidase trans-cinnamate 4-monooxygenase (CYP73A, EC: 1.14.13.11) and 4-coumarate-CoA ligase (4CL EC: 6.2.1.12). P-Cinnamoyl-CoA and cinnamoyl-CoA produce naringenin, which finally yields flavonoids under catalysis by oxidoreductase flavanone 3-hydroxylase (F3H, EC: 1.14.11.9), which maintains the development of the pollen tubes and pollen walls. The heat map showing the expression levels of these key enzymes is presented in Fig. 8c (Additional file 7: Table S7). *Traes_2BS_88CF42F2E*, *Traes_2DL_10A4DDD75*, and *Traes_2DS_F6307AF21* were downregulated in fertile anthers, but the other key enzymes encoded by these DEGs were upregulated in fertile anthers. These results indicate that the upregulation of carbohydrate metabolism and phenylpropanoid biosynthesis as well as metabolism-related genes in fertile (KMF) anthers may ensure the normal development and fertility of pollen.

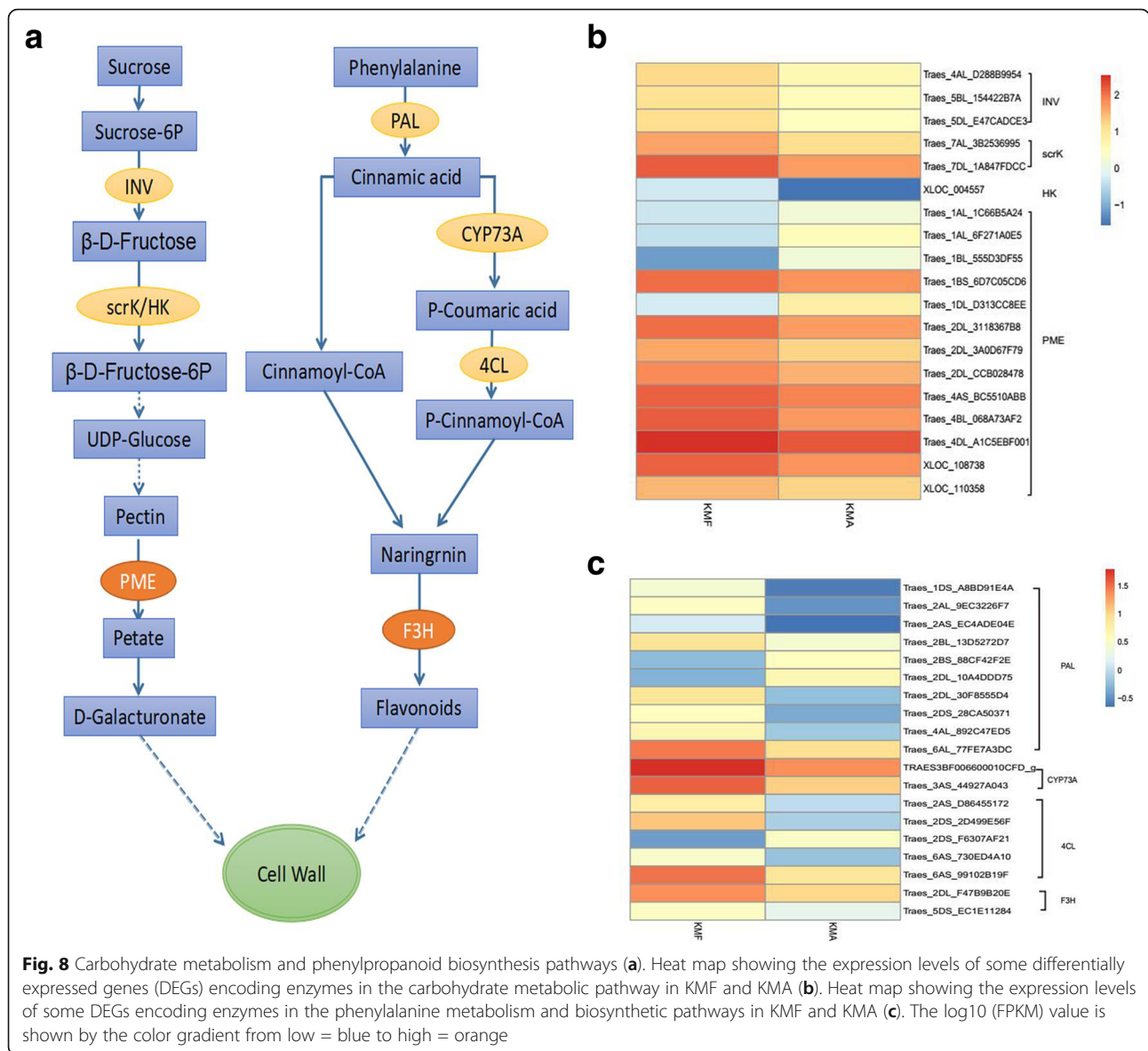


Fig. 8 Carbohydrate metabolism and phenylpropanoid biosynthesis pathways (a). Heat map showing the expression levels of some differentially expressed genes (DEGs) encoding enzymes in the carbohydrate metabolic pathway in KMF and KMA (b). Heat map showing the expression levels of some DEGs encoding enzymes in the phenylalanine metabolism and biosynthetic pathways in KMF and KMA (c). The log₁₀ (FPKM) value is shown by the color gradient from low = blue to high = orange

Total soluble sugar and flavonoid contents

The soluble sugar and flavonoid contents were determined to elucidate the changes in carbohydrate metabolism and phenylpropanoid biosynthesis and metabolism in the fertile anthers (KMF) and sterile anthers (KMA). In the carbohydrate metabolism pathway, most of the DEGs were upregulated in KMF. Thus, we hypothesized that the expression of genes related to fertility may lead to differences in the total soluble sugar contents in fertile (KMF) and sterile (KMA) plant anthers, where the fertile plant anthers may have higher sugar contents. Therefore, we measured the total soluble sugar contents in the anthers from KMF and KMA plants, and the results showed that the soluble sugar contents of the fertile (KMF) anthers

were significantly higher than those of the sterile (KMA) anthers (Fig. 9a, Additional file 8: Table S8). Moreover, previous studies have shown that phenylpropanoid metabolism has important links with flavonoids and it influences anther development [28]. We found that F3H was highly expressed in fertile (KMF) anthers but barely expressed in the sterile (KMA) anthers according to the RNA sequencing results, and thus we measured the flavonoid contents in anthers from KMF and KMA plant. The results showed that the flavonoid contents of the fertile (KMF) anthers were significantly higher than those of the sterile (KMA) anthers (Fig. 9b, Additional file 9: Table S9). These results agreed with our hypothesis and also verified the reliability of the sequencing results.

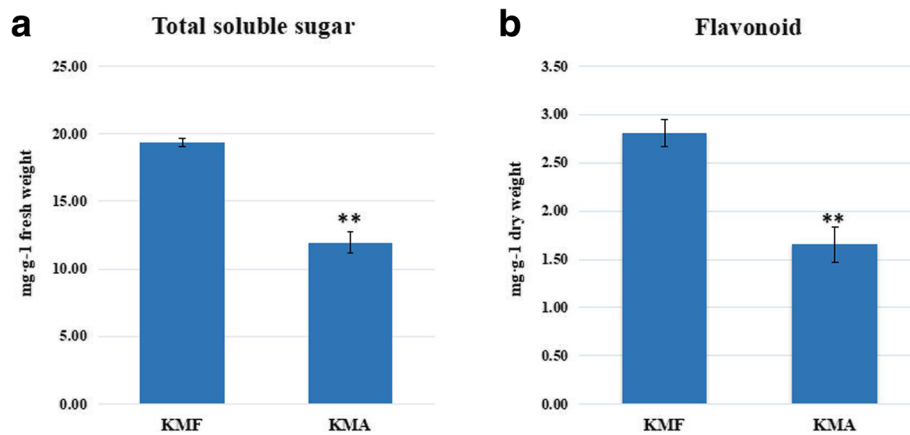


Fig. 9 Total soluble sugar and flavonoid contents in the anthers from KMF and KMA. Total soluble sugar contents (a). Flavonoid contents (b). Data represent the mean and standard deviation based on three replicates. ** $p < 0.01$ according to the Student's t-test

Verification of DEGs by qRT-PCR

In order to validate the differential expression patterns detected for DEGs by RNA sequencing, 12 DEGs were selected for qRT-PCR detection from the significant metabolic pathways, which encoded F3H (Traes_2DL_F47B9B20E), pectinesterase/pectinesterase inhibitor 21 (PME 21, Traes_2DL_3A0D67F79, Traes_4AS_BC5510ABB), 4-coumarate: coenzyme A ligase (4CL, Traes_2DS_2D499E56F), lysosomal beta glucosidase (GBA, Traes_2DL_B87857C02), WRKY27 transcription factor (WRKY 27, Traes_3B_990298FF5), phenylalanine ammonia-lyase 4 (PAL 4, Traes_6AL_

67ED9DB4D), 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (PGAM, Traes_6BL_01DF926F1, Traes_6DL_6286CEC01), fructokinase-2-like (FLN, Traes_7DL_1A847FDCC), trans-cinnamate 4-monooxygenase (CYP73A, Traes_3AS_44927A043), phenylalanine ammonia-lyase (PAL, Traes_6DL_1AEA7B869). The expression patterns of these genes are shown in Fig. 10 (Additional file 10: Table S10), which demonstrated that the gene expression patterns obtained by qRT-PCR and RNA sequencing data exhibited similar trends, thereby confirming the accuracy of the RNA sequencing results obtained in this study.

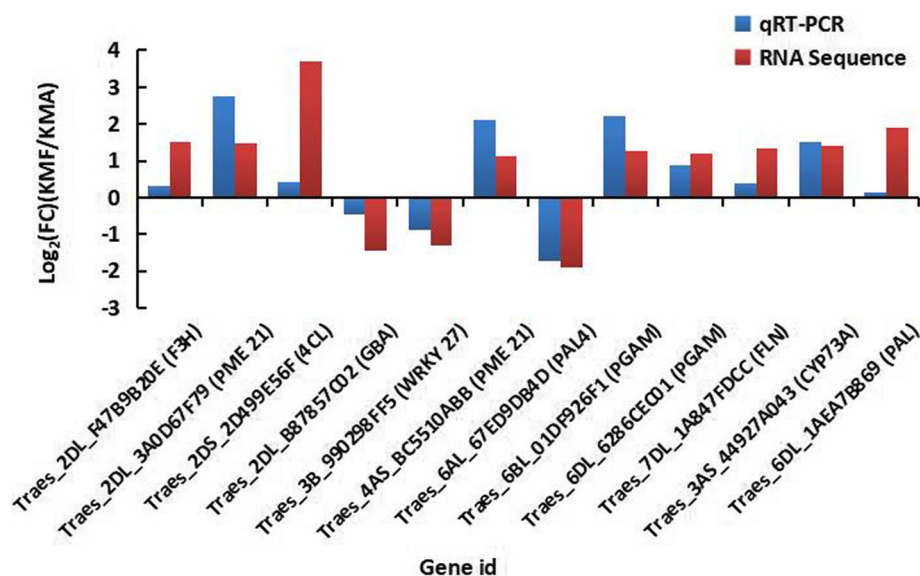


Fig. 10 qRT-PCR validates the RNA sequencing results of some DEGs. Data obtained by qRT-PCR represent means based on three replicates. Log₂ (FC) represents logarithmic value of change in expression for KMF relative to KMA

Discussion

Carbohydrate metabolism affects Cell Wall development, anther growth, and fertility

From the differentiation of the pistil and stamen until the formation of the mature pollen grains, plants synthesize and degrade materials frequently in order to provide energy to support pollen germination, pollen tube growth, and fertilization. Carbohydrates provide energy to maintain the development of the anthers and pollen, but they also act as signals that influence their development [29]. A large number of related genes or proteins are involved in the combined regulatory process. Many studies have shown that the abnormal expression of genes or proteins in anthers interferes with pollen development, thereby affecting pollen fertility. PME (E3.1.1.11) belongs to a large family of carbohydrate esterases and this enzyme regulates the degree of pectin methylation as well as being a major component of plant cell walls. Previous studies of the function of PME indicate that it is involved in cell wall loosening and it also participates in various plant growth processes, such as pollen formation and pollen tube elongation [30]. PME can endogenously regulate the pectin content between plant cell walls and cells. After demethylation by PME, pectic acid binds to Ca^{2+} to form insoluble substances, which inhibit cell separation and organ shedding [31]. The loss of function by VANGUARD1 may reduce the efficiency of the interaction between the pollen tube and the extracellular matrix in the style and transmitting tract, thereby leading to the blockage of the pollen tube in the style and transmitting tract to greatly reduce male fertility [32]. *BcMF23a* and *BcMF23b* are expressed in microspores during the dinuclear pollen and mature pollen stages, and they are highly expressed in fertile flower buds, but silent in genic male sterile lines [33]. In the starch and sucrose metabolism pathway, PME catalyzes the hydrolysis of pectin to produce pectic acid and methanol. In the present study, 13 DEGs were annotated as PMEs and nine were significantly upregulated in fertile anthers but not in infertile anthers (Fig. 8b, Additional file 7: Table S7), and the other four had very low expression levels in both fertile and sterile anthers. These DEGs were annotated in the starch and sucrose metabolism, and pentose and glucuronate interconversion pathways. It is notable that of the 14 DEGs detected in the pentose and glucuronate interconversion pathway, 13 encoded PMEs. Therefore, we suggest that PME enzymes have crucial effects on cell wall development, anther growth, and fertility, and thus the roles of PMEs in wheat fertility merit further study.

Flavonoids are crucial for the development of fertile anthers

Many studies have shown that the phenylalanine metabolism and biosynthetic pathways have important roles in

plant fertility. During the pollen maturation process, phenylalanine is used as a substrate for a series of biochemical enzymes, including PAL, C4H (Cinnamic acid 4 hydroxylase), CHS (Chalcone synthase), CHI (Chalcone isomerase), and FLS (Flavonol synthase), before the formation of phenolic resin and long-chain fatty acids, and sporopollenin is finally synthesized as the main component of the pollen wall [34]. Defects in the pollen wall structure can lead to abnormal pollen development, which affects the fertility of plants [35]. In the present study, numerous DEGs were enriched in the phenylpropanoid metabolism and biosynthetic pathways. The enzymes encoded by these DEGs were upregulated in fertile anthers and we suggest that they may have important relationships with fertility. Flavonoids are synthesized via the phenylpropane pathway [36] and they have key roles in pollen development because they are essential for pollen maturation and pollen tube growth in flowering plants [37]. In addition, F3H is a key enzyme for the formation of flavonoids and it belongs to the family of dioxygenases that depend on 2-ketoglutarate [38]. F3H can regulate metabolism alone but it often co-catalyzes the synthesis of flavonoids with chalcone isomerase and chalcone synthase [39]. In the present study, F3H was upregulated in fertile pollen but its expression level was low in infertile anthers. Flavonoids are the main raw material used for the synthesis of pink pigments as well as being important for the development of normal anthers. We also detected high levels of flavonoids in fertile anthers, thereby supporting the key role of flavonoids in fertility.

Transcription factors involved are in the regulation of pollen development and male fertility

In flowering plants, male gametophytes (or pollen grains) play a vital role in plant fertility, and the development of pollen grains is strictly controlled, where it involves the complex and precise regulation of many transcription factors [40, 41]. Changes in the expression levels of some transcription factors may lead to abnormal pollen development and male sterility during plant growth and development. For example, the overexpression of *WRKY27* in *Arabidopsis* plants leads to pollen cracking defects, growth aberrations, and male sterility [42]. *WRKY2* and its close homolog *WRKY34* are required for male gametogenesis and they play important roles in pollen development and function. Thus, the overexpression of *WRKY 34* in mature pollen from transgenic *Arabidopsis* plants was associated with greatly reduced fertility [40, 43]. Previous studies have also shown that many WRKY transcription factors are linked with resistance to abiotic and biotic stresses. However, the effects of WRKY on the control of pollen and flower growth and development have not been widely investigated, and previous studies have focused mainly on

Arabidopsis. In the present study, we detected 14 genes encoding WRKY transcription factors that were specifically expressed in wheat anthers with differences in fertility, i.e., *WRKY6*, *WRKY*, *WRKY14*, *WRKY2*, *WRKY51*, *WRKY1B*, *WRKY68-b*, *WRKY1A*, *WRKY42*, *WRKY27*, *WRKY48*, *WRKY3*, and *WRKY45-like*. The expression levels of 13 of these WRKY transcription factors were up-regulated in male sterile wheat anthers (Fig. 8b), thereby indicating that their upregulation in sterile wheat has important effects on fertility. Further studies of the regulation of these WRKY transcription factors in wheat pollen grains may enhance our understanding of the molecular mechanisms related to pollen development and fertility.

The MYB transcription factor family is a large class of transcription factors and they are also involved in the regulation of gene expression during plant growth, where they mainly participate in primary and secondary metabolism, including phenylpropanoid metabolism [44] and flavanol biosynthesis [45]. Moreover, some MYB transcription factors are associated with flower development. The R2R3-MYB transcription factor MYB21 plays a dominant role in the elongation of stamen filaments. In *Arabidopsis*, the overexpression of *MYB21* partially restores stamen filament growth and fertility in the *opr3* mutant [46, 47]. Gibberellin MYB is involved with gibberellin acid signal transduction, which strongly affects flower development in angiosperms, but mainly anther and pollen growth [48]. In addition, some bHLH transcription factors interact with MYB transcription factors to regulate diverse plant responses, including secondary metabolic processes, flavonoid pathways, and stamen development [49]. For example, bHLH transcription factors (MYC2, MYC3, MYC4, and MYC5) interact with MYB transcription factors (MYB21 and MYB24) to form the bHLH-MYB complex, and jasmonates repress the bHLH-MYB complex to regulate jasmonic acid-mediated stamen development and fertility [50]. Some modulators of flavonoid biosynthesis that interact with R2R3-MYBs or bHLHs can promote or impair the production of flavonoids [51]. In our study, most of the bHLH and MYB transcription factors were upregulated in fertile wheat and downregulated in sterile wheat (Fig. 8b). The variable expression levels of these MYB and bHLH transcription factors suggest that they have relationships with wheat fertility and they may have important roles in this process. These results provide novel insights into flavonoid biosynthesis and fertility, thereby increasing our understanding of the various mechanisms responsible for regulating fertility.

Mitochondria-associated genes affecting pollen fertility restoration

The normal development of plant pollen is inseparable from the coordinated interaction between nucleus and

cytoplasm, and is inseparable from the selective expression of genes.

Studying mitochondria-related gene expression is of great significance for understanding the temporal and spatial expression of genes and elucidating the molecular mechanism of cytoplasmic male sterility. Mitochondria is a semi-autonomous organelle that encodes its own rRNA, tRNA and a small amount of protein, and it is inseparable from the guidance and regulation of the nuclear genome [52]. Studies have shown that pollen fertility restoration is a process of nucleo-cytoplasmic interaction. When one or more fertility restoration genes are introduced, the structure or expression of CMS-related genes are regulated accordingly, so that the function of mitochondria returns to normal, thus making pollen fertility is restored [53]. Some studies have found that the normal expression of mitochondria-related genes plays an important role in pollen development and fertility. The Mg^{2+} transporter *MRS2/MGT* gene family mainly mediates magnesium transport between cytoplasm and mitochondria [54]. Studies have found that *AtMGT 5* gene is specifically expressed in anthers at early flower development, and male gametes of *AtMGT 5* mutant produce defective pollen [55]. In addition, disruption of *AtMGT 9* gene expression leads to abortion of mature pollen grains [56], indicating *AtMGT 9* and *AtMGT 5* genes are critical in pollen development and male fertility. In this study (Additional file 11: Table S11), two genes annotated as the mitochondrial inner membrane magnesium transporter *mrs2* were up-regulated in fertile anthers, which expressed twice as much as the sterile anthers, indicating normal expression of the *mrs2* gene is associated with fertility restoration of fertile anthers. NADH dehydrogenase is an enzyme that catalyzes the transfer of electrons from NADH to Coenzyme Q in the mitochondrial inner membrane, and its presence makes the plant's respiratory chain more flexible in the process of bioenergy function [57]. Formate dehydrogenase (FDH) is a NAD^+ dependent enzyme widely distributed in organisms, and FDH catalyzes the oxidation of formic acid to carbon dioxide while reducing NAD^+ to NADH in this process [58]. ATP is the main energy currency of cells, which is catalytically decomposed into ADP and inorganic phosphate by ATPase to promote most biological reactions in the cytoplasm. Mohammed Sabar et al. found that the ATPase activity of F1-F0-ATP synthase in the sterile line was significantly lower than that of the fertile line by comparing the gelase activities of several mitochondrial respiratory complexes between the fertile line and the sterile line, and thought that this may be the effect of *orf 522* expression, which makes the sterile line unable to meet the energy requirements needed to maintain anther development, leading to pollen abortion [59]. In our study, two genes encoding NADH dehydrogenase, one encoding formate dehydrogenase and five encoding AAA-ATPase

were down-regulated in sterile wheat anthers. We predicted that the down-regulated expression of this series of related enzymes affects energy transfer in mitochondria associated with cytoplasmic and nuclear interactions, which in turn affects fertility restoration of pollen. These findings provide new insights into the potential molecular regulation mechanisms of mitochondria-associated genes and fertility restoration in cytoplasmic male sterile wheat.

Conclusions

The induction of fertility restoring genes in the presence of *Aegilops kotschy* cytoplasm causes normal pollen development and fertility in KTM3315R, mainly via effects on carbohydrate metabolism, phenylpropanoid metabolism and biosynthesis, and upregulating the expression of candidate genes that encode PME and F3H, which are closely related to pollen development. In addition, the regulation of pollen development and male fertility is affected by the WRKY, bHLH, and MYB transcription factor families, where male fertility is associated with the downregulation of WRKY and the upregulation of bHLH and MYB. Thus, these transcription factors can enhance pollen growth and development, and restore male fertility in wheat with *Aegilops kotschy* cytoplasm. Our results indicate that the genes responsible for regulating fertility restoration play important roles in the presence of *Aegilops kotschy* cytoplasm, thereby providing a basis for exploring the molecular mechanisms that regulate changes in wheat fertility.

Methods

Plant materials

The study materials comprised the near-isogenic lines KTM3315A (designated as KMA, a K-TCMS line with *Aegilops kotschy* cytoplasm) and KTM3315R (designated as KMF, a cytoplasmic male fertile line with *Aegilops kotschy* cytoplasm), which were developed over the course of many years by Northwest A&F University. The sterile line KTM3315A was used as a female parent to hybridize with the homologous maintainer line TM3315B for which anthers are provided to produce the sterile line KTM3315A (sterile line propagation method). The breeding procedure of KTM3315R is as follows: Chinese Spring has a high and stable restoration degree (seed-setting rate, ~ 75.00%) for K-CMS [34], so we selected restorer Chinese Spring as the pollen donor hybridized with KTM3315A to produce F₁ generation of KTM3315A and Chinese Spring. K-CMS lines belong gametophytic male-sterile type [60] and the recessive genes could hardly be transmitted by male gametes (~ 0.54%), so we selected its homologous maintainer TM3315B as the pollen donor and crossed it with the F₁ generation of KTM3315A and Chinese Spring. As a recurrent parent, TM3315B was backcrossed 8 times with the F₁ generation, and the fertile plants were selected

to produce the BC₈F₁ population. The BC₈F₁ population was self-crossed for 5 generations, forming a near-isogenic line KTM3315R that was stably inherited. The breeding procedure is illustrated in Additional file 12: Figure S1.

During October 2016, KMF and KMA were cultivated under natural conditions and regular field management at the Northwest A&F University experimental station in Yangling (34°15'N, 108°08'E), China. During May 2017, three spikes were selected randomly on each plant for self-pollination and bagged at the heading stage until subsequent fertility analyses. The stages and fertility of the other spikes were identified using 1% acetocarmine and 1% I₂-KI. Anthers from the binucleate stage were collected as three biological replicates for 20 fertile plants and 20 sterile plants with equal amounts from KMF and KMA for sequencing. The fertile and sterile anthers were immediately snap frozen in liquid nitrogen and stored at - 80 °C until use.

Phenotypic traits and microscopic observations

Photographs of the sterile and fertile wheat anther phenotypes were obtained using a Nikon E995 digital camera (Nikon, Tokyo, Japan) mounted on a Motic K400 dissection microscope (Preiser Scientific, Louisville, KY, USA). The different anther development stages were identified by staining with 1% acetocarmine [61]. To evaluate the viability of mature pollen grains, anthers in the stage before dehiscence were crushed in 1% I₂-KI, incubated for 15 to 20 min at room temperature with I₂-KI, washed with buffer, and the pollen grains were then observed by microscopy [62]. The anthers, anther outer epidermis, and trinucleate stage microspores were analyzed by scanning electron microscopy as described by [63] with a JSM-6360LV scanning electron microscope (JEOL, Tokyo, Japan).

RNA extraction, cDNA library construction, and Illumina deep sequencing

The total RNA was extracted from six samples of KMF and KMA plant anthers, each with three biological replicates, according to the instruction manual provided with the RNAiso for Polysaccharide-rich Plant Tissue kit (Takara Biological Engineering (Dalian) Co. Ltd., China). The RNA concentration and purity were measured using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). RNA integrity was assessed using an RNA Nano 6000 Assay Kit for the Agilent Bioanalyzer 2100 System (Agilent Technologies, Inc., Santa Clara, CA, USA). After the RNA samples passed through these three steps, eukaryotic mRNA was enriched with Oligo (dT) beads. The enriched mRNA was then fragmented into short fragments using fragmentation buffer and reverse transcribed into first-strand cDNA with random primers. Second-strand cDNA was synthesized using DNA

polymerase I, RNase H, dNTP, and buffer. The cDNA samples were purified with a QIAquick Gel Extraction kit (Beijing Lanbo Kangsi Technology Co. Ltd., China), before end repair, adding poly (A) tails, and ligating to Illumina sequencing adapters. The ligation products were selected by size via agarose gel electrophoresis, before PCR amplification, and sequencing using the Illumina HiSeq™2500 system by Sagene Biotech Co. Ltd. (Guangzhou, China).

RNA sequencing data analysis

The reads obtained by sequencing comprised raw reads containing adapters or low quality bases, which would have affected the subsequent assembly and analysis steps. Thus, the raw reads were filtered to obtain high-quality reads by removing low-quality reads containing adapters, unknown nucleotides ($N > 10\%$), and more than 50% low quality bases (Q-value ≤ 20). The short reads alignment tool Bowtie2 [64] was used to map the reads to the ribosomal RNA (rRNA) database (<ftp://ftp.ncbi.nlm.nih.gov/genbank/>). The reads that mapped to the rRNA database were removed. The remaining reads from each sample were mapped to the Ensembl release 31 IWGSC1.0 + NC_002762.1 reference genome with TopHat2 (version 2.0.3.12) [65]. The transcripts were reconstructed with Cufflinks (version 2.2.1) [66]. Gene abundances were quantified with RSEM (version 1.2.31) [67]. The gene expression levels were normalized using the fragments per kilobase of transcripts per million mapped reads (FPKM) method.

Bioinformatics analysis of DEGs

DEGs were functionally annotated using the non-redundant protein database (Nr; NCBI: <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>). The DEGs were then further aligned with the Clusters of Orthologous Groupss (<https://www.ncbi.nlm.nih.gov/COG/>) database to predict and classify their functions [68], before enrichment analysis according to their Gene Ontology (GO; <http://www.geneontology.org/>) functions and Kyoto Encyclopedia of Genes and Genomes) (KEGG; <http://www.genome.jp/kegg/genes.html>) pathways. In addition, the GO annotations were analyzed to identify the main biological functions of the DEGs where the Blast2GO program [69] (<https://www.blast2go.com/>) was used to obtain GO annotations for all of the DEGs. The results were submitted to WEGO (<http://wego.genomics.org.cn>) to generate a GO classification graph for all of the DEGs. KEGG is the major public database related to pathways [70] and it can identify the major biochemical metabolic pathways and signal transduction pathways for DEGs. According to the GO and pathway enrichment analysis results, q -value ≤ 0.05 was selected as the threshold of significance to determine enrichment in the DEG sets [71]. OmicShare small tools2 (<http://www.omichshare.com/tools/>) was employed to obtain a heat map without rows and column clusters.

Assays of soluble sugar and flavonoid contents

Approximately 0.1 g of the fertile (KMF 1, KMF 2, and KMF 3 represents three biological repeats, respectively, the same as following) or sterile (KMA 1, KMA2, and KMA3) anthers was weighed, before adding 1 mL of distilled water, and grinding into a homogenate, which was then poured into a 2 mL centrifuge tube and the lid was closed. The sample was heated in a water bath at 95 °C for 10 min, cooled, and then subjected to centrifugation at 8000×g and 25 °C for 10 min, before pouring the supernatant into a 10 mL test tube. The volume was adjusted to 10 mL with distilled water, before shaking gently by hand for about 30 s. The procedures and calculations were performed according to the instructions provided with a Plant Soluble Sugar Content Determination kit (Suzhou Keming Biotechnology Co. Ltd., China).

The flavonoid contents were determined by drying samples (same as above, KMF 1, 2, 3 and KMA 1, 2, 3) to a constant weight and crushing after passing through a 40-mesh sieve. Approximately 0.1 g of the crushed sample was weighed and added to 2.5 mL of distilled water, before ultrasonic extraction (ultrasonic power = 300 W, crushing = 5 s, intermittent = 8 s, temperature = 60 °C, extraction = 30 min). The sample was then subjected to centrifugation at 12000 rpm and 25 °C for 10 min, before analyzing the supernatant from the extract. A spectrophotometer was turned on for more than 30 min and the wavelength was adjusted to 502 nm, before zeroing with distilled water and measuring the absorbance of each sample. The operations and calculations were performed according to the instructions provided with a Plant Flavonoid Content Determination kit (Sino Best Biological Technology Co., Ltd., China).

Confirmation of candidate DEGs by quantitative reverse-transcription PCR (qRT-PCR)

To validate the DEGs detected by RNA sequencing, 12 DEGs were selected from the significant metabolic pathways. The primers used for qRT-PCR were designed with Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and synthesized by Xi'an Qingke Zexi Biotechnology Co. Ltd., China. The actin gene (GenBank: GQ339766.1) was used as a reference to normalize the gene expression levels where it was set to 1 [72]. qRT-PCR analysis was performed as described by Ye [35]. The sequence-specific primer pairs used for qRT-PCR are listed in Additional file 13: Table S12, with those for the actin gene and the 12 selected DEGs. The analysis was conducted using three technical replicates for each sample. We calculated the expression level of the actin gene using the $2^{-\Delta\Delta C_t}$ method and normalized the relative expression levels with respect to the expression level of the actin gene [73].

Additional files

- Additional file 1:** Table S1. Summary of sequencing results. (XLSX 10 kb)
- Additional file 2:** Table S2. Differentially expressed genes. (XLSX 387 kb)
- Additional file 3:** Table S3. Gene Ontology annotations. (XLSX 66 kb)
- Additional file 4:** Table S4. KOG class annotations. (XLSX 30 kb)
- Additional file 5:** Table S5. Significantly enriched KEGG pathways. (XLSX 14 kb)
- Additional file 6:** Table S6. WRKY, bHLH, and MYB transcription factors. (XLSX 12 kb)
- Additional file 7:** Table S7. DEGs involved in pathways. (XLSX 17 kb)
- Additional file 8:** Table S8. Soluble sugar contents. (XLSX 11 kb)
- Additional file 9:** Table S9. Flavonoid contents. (XLSX 11 kb)
- Additional file 10:** Table S10. The gene expression of qRT-PCR and RNA sequencing. (XLSX 15 kb)
- Additional file 11:** Table S11. Mitochondrial function-related genes. (XLSX 14 kb)
- Additional file 12:** Figure S1. Breeding procedure of near-isogenic lines. (DOCX 1939 kb)
- Additional file 13:** Table S12. Sequence-specific primers used for qRT-PCR. (XLSX 11 kb)

Abbreviations

4CL: 4-Coumarate-CoA ligase; CMS: Cytoplasmic male sterility; CYP73A: Trans-cinnamate 4-monooxygenase; DEGs: Differentially expressed genes; F3H: Flavanone 3-hydroxylase; FC: Fold Change; FDR: False Discovery Rate; FPKM: Fragments Per Kilobase of transcript per Million mapped reads; GBA: lysosomal beta glucosidase; HK: Hexokinase; INV: Hydrolase beta-fructofuranosidase; KEGG: Kyoto Encyclopedia of Genes and Genomes; KMA: KTM3315A; KMF: KTM3315R; KOG: Eukaryotic Orthologous Groups; K-TCMS: Thermo-sensitive male-sterility; PAL: Phenylalanine ammonia-lyase; PGAM: 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase; PME: Pectin methylesterase/pectinesterase; scrK: Fructokinase; UDP-glucose: Uridine diphosphate-glucose

Acknowledgements

We thank Guangzhou Sagene Biotech Company for the help with high throughput sequencing.

Funding

The high throughput sequencing was financially supported by grants from the National Natural Science Foundation of China (31771874). The editing and publishing fee were supported by the Program in Science and Technology of Yangling State Demonstration Zone of Agricultural High-tech Industries (2018NY-19).

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

XS and LZ conceived and designed the study. SL, YJ, and JY performed the experiments. SL and ZL analyzed the data. XY provided advice and assistance. SL wrote the article with contributions from all the authors. XS critically revised the manuscript. All of the authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's note

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

Received: 3 November 2018 Accepted: 9 May 2019

Published online: 11 June 2019

References

- Liu YJ, Gao SQ, Tang YM, Gong J, Zhang X, Wang YB, Zhang LP, Sun RW, Zhang Q, Chen ZB, et al. Transcriptome analysis of wheat seedling and spike tissues in the hybrid Jingmai 8 uncovered genes involved in heterosis. *Planta*. 2018;247(6):1307–21.
- Lin C, Zhang C, Zhao H, Xing S, Wang Y, Liu X, Yuan C, Zhao L, Dong Y. Sequencing of the chloroplast genomes of cytoplasmic male-sterile and male-fertile lines of soybean and identification of polymorphic markers. *Plant Sci*. 2014;229:208–14.
- Bohra A, Jha UC, Adhimoalam P, Bisht D, Singh NP. Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep*. 2016;35(5):967–93.
- Ivanov MK, Dymshits GM. Cytoplasmic male sterility and restoration of pollen fertility in higher plants. *Genetika*. 2007;43(4):451–68.
- Touzet P, Budar F. Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? *Trends Plant Sci*. 2004;9(12):568–70.
- Fujii S, Toriyama K. Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol*. 2008;49(10):1484–94.
- Touzet P, Meyer EH. Cytoplasmic male sterility and mitochondrial metabolism in plants. *Mitochondrion*. 2014;19(Pt B):166–71.
- Saxena KB, Kumar RV, Tikle AN, Saxena MK, Gautam VS, Rao SK, Khare DK, Chauhan YS, Saxena RK, Reddy BVS, et al. ICPH 2671 - the world's first commercial food legume hybrid. *Plant Breed*. 2013;132:479–85.
- Huang JZ, E ZG, Zhang HL, Shu QY. Workable male sterility systems for hybrid rice: genetics, biochemistry, molecular biology, and utilization. *Rice (N Y)*. 2014;7(1):13.
- Yongming L, Zhuofan Z, Yanli L, Chuan L, Jing W, Boxiao D, Bing L, Tao Q, Wenbing Z, Moju C. A preliminary identification of Rf^{*}-A619, a novel restorer gene for CMS-C in maize (*Zea mays* L.). *PeerJ*. 2016;4:e2719.
- Elkonin LA, Tsvetova MI. Heritable effect of plant water availability conditions on restoration of male fertility in the "9E" CMS-inducing cytoplasm of sorghum. *Front Plant Sci*. 2012;3:91.
- Dong DK, Li Z, Yuan FJ, Zhu SL, Chen P, Yu W, Yang QH, Fu XJ, Yu XM, Li BQ, et al. Inheritance and fine mapping of a restorer-of-fertility (Rf) gene for the cytoplasmic male sterility in soybean. *Plant Sci*. 2012;188-189:36–40.
- Yan R, Liang C, Meng Z, Malik W, Zhu T, Zong X, Guo S, Zhang R: Progress in genome sequencing will accelerate molecular breeding in cotton (*Gossypium spp.*). *3 Biotech* 2016, 6(2):217.
- Song X, Fang P, Ma L, Xi Y, Hu Y, Liu S, He B. A comparison of wheat CMS lines of *Ae. kotschyi* cytoplasm of no 1B/1R type and 1B/1R type. *Jour of Northw est Sci-Tech Univ of Agri and For (Nat Sci Ed)*. 2002. <https://doi.org/10.13207/j.cnki.jnwafu.2002.01.001>.
- Song X, Zhang L, Zeng J, Qian H, Li H, He B. Development of thermo-sensitive cytoplasmic male sterile (TCMS) lines of wheat characterized by complete male sterility at lower-temperatures and partially restored fertility at higher temperatures. *Euphytica*. 2013;192:393–9.
- Meng L, Liu Z, Zhang L, Hu G, Song X. Cytological characterization of a thermo-sensitive cytoplasmic male-sterile wheat line having K-type cytoplasm of *Aegilops kotschyi*. *Breed Sci*. 2016;66:752–61.
- Suzuki H, Rodriguez-Urbe L, Xu J, Zhang J. Transcriptome analysis of cytoplasmic male sterility and restoration in CMS-D8 cotton. *Plant Cell Rep*. 2013;32(10):1531–42.
- Yang P, Han J, Huang J. Transcriptome sequencing and de novo analysis of cytoplasmic male sterility and maintenance in JA-CMS cotton. *PLoS One*. 2014;9(11):e112320.
- Li Z, Cheng Y, Cui J, Zhang P, Zhao H, Hu S. Comparative transcriptome analysis reveals carbohydrate and lipid metabolism blocks in *Brassica napus* L. male sterility induced by the chemical hybridization agent monosulfuron ester sodium. *BMC Genomics*. 2015;16(1):206.

20. M-x S, Qu C, Fu F, Liu M, Zhao H, Liu C, Li J, Tang Z, Xu X, Qiu X, et al. Comparative transcriptome analysis of recessive male sterility (RGMS) in sterile and fertile *Brassica napus* lines. *PLoS One*. 2015;10(12):e0144118.
21. Ishiguro S, Ogasawara K, Fujino K, Sato Y, Kishima Y. Low temperature-responsive changes in the anther transcriptome's repeat sequences are indicative of stress sensitivity and pollen sterility in rice strains. *Plant Physiol*. 2014;164(2):671–82.
22. Wang S, Wang C, Zhang XX, Chen X, Liu JJ, Jia XF, Jia SQ. Transcriptome de novo assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage. *Plant Physiol Biochem*. 2016;105:224–32.
23. Liu H, Tan M, Yu H, Li L, Zhou F, Yang M, Zhou T, Zhao Y. Comparative transcriptome profiling of the fertile and sterile flower buds of a dominant genic male sterile line in sesame (*Sesamum indicum* L.). *BMC Plant Biol*. 2016;16(1):250.
24. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(11):139–40.
25. Dong X, Feng H, Xu M, Lee J, Kim YK, Lim YP, Piao Z, Park YD, Ma H, Hur Y. Comprehensive analysis of genic male sterility-related genes in *Brassica rapa* using a newly developed Br300K oligomeric chip. *PLoS One*. 2013;8(9):e72178.
26. Zhou X, Liu Z, Ji R, Feng H. Comparative transcript profiling of fertile and sterile flower buds from multiple-allele-inherited male sterility in Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*). *Mol Gen Genomics*. 2017;292(5):967–90.
27. Xie F, Yuan JL, Li YX, Wang CJ, Tang HY, Xia JH, Yang QY, Wan ZJ. Transcriptome analysis reveals candidate genes associated with leaf etiolation of a cytoplasmic male sterility line in Chinese cabbage (*Brassica Rapa* L. ssp. *Pekinensis*). *Int J Mol Sci*. 2018;19(4):922.
28. Zhang G, Ye J, Jia Y, Zhang L, Song X. iTRAQ-based proteomics analyses of Ssterile/FFertile anthers from a thermo-sensitive cytoplasmic male-sterile wheat with *Aegilops kotschy* cytoplasm. *Int J Mol Sci*. 2018;19:1344.
29. Clément C, Burrus M, Audran JC. Floral organ growth and carbohydrate content during pollen development in Liliium. *Am J Bot*. 1996;83(4):459–69.
30. Tian GW, Chen MH, Zaltsman A, Citovsky V. Pollen-specific pectin methyltransferase involved in pollen tube growth. *Dev Biol*. 2006;294(1):83–91.
31. Micheli F. Pectin methyltransferases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci*. 2001;6(9):414–9.
32. Jiang L, Yang SL, Xie LF, Puah CS, Zhang XQ, Yang WC, Sundaresan V, Ye D. VANGUARD1 encodes a pectin methyltransferase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. *Plant Cell*. 2005;17(2):584–96.
33. Lin S, Huang L, Yu X, Xiong X, Yue X, Liu T, Liang Y, Lv M, Cao J. Characterization of *BcMF23a* and *BcMF23b*, two putative pectin methyltransferase genes related to pollen development in *Brassica campestris* ssp. *chinensis*. *Mol Biol Rep*. 2017;44(1):139–48.
34. Herdt E, Sutfeld R, Wiermann R. The occurrence of enzymes involved in phenylpropanoid metabolism in the tapetum fraction of anthers. *Cytobiologie*. 1978;17(2):433–41.
35. Ye J, Duan Y, Hu G, Geng X, Zhang G, Yan P, Liu Z, Zhang L, Song X. Identification of candidate genes and biosynthesis pathways related to fertility conversion by wheat KTM3315A transcriptome profiling. *Front Plant Sci*. 2017;8:449.
36. Kang K, Park S, Natsagdorj U, Kim YS, Back K. Methanol is an endogenous elicitor molecule for the synthesis of tryptophan and tryptophan-derived secondary metabolites upon senescence of detached rice leaves. *Plant J*. 2011;66(2):247–57.
37. Wani TA, Pandith SA, Gupta AP, Chandra S, Sharma N, Lattoo SK. Molecular and functional characterization of two isoforms of chalcone synthase and their expression analysis in relation to flavonoid constituents in *Grewia asiatica* L. *PLoS One*. 2017;12(6):e0179155.
38. Pelletier MK, Shirley BW. Analysis of flavanone 3-hydroxylase in Arabidopsis seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase. *Plant Physiol*. 1996;111(1):339–45.
39. Owens DK, Crosby KC, Runac J, Howard BA, Winkel BSJ. Biochemical and genetic characterization of Arabidopsis flavanone 3b-hydroxylase; 2008.
40. Guan Y, Meng X, Khanna R, LaMontagne E, Liu Y, Zhang S. Phosphorylation of a WRKY transcription factor by MAPKs is required for pollen development and function in Arabidopsis. *PLoS Genet*. 2014;10:e1004384.
41. Lei R, Li X, Ma Z, Lv Y, Hu Y, Yu D. Arabidopsis WRKY2 and WRKY34 transcription factors interact with VQ20 protein to modulate pollen development and function. *Plant J*. 2017;91(6):962–76.
42. Mukhtar MS, Liu X, Somssich IE. Elucidating the role of WRKY27 in male sterility in Arabidopsis. *Plant Signal Behav*. 2017;12(9):e1363945.
43. Zou C, Jiang W, Yu D. Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in Arabidopsis. *J Exp Bot*. 2010;61(14):3901–14.
44. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of Phenylpropanoid metabolism in plants. *Mol Plant*. 2015;8(5):689–708.
45. Stracke R, Ishihara H, Hupé G, Barsch A, Mehrrens F, Niehaus K, Weisshaar B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J*. 2007;50(4):660–77.
46. Cheng H, Song S, Xiao L, Soo HM, Cheng Z, Xie D, Peng J. Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in Arabidopsis. *PLoS Genet*. 2009;5(3):e1000440.
47. Song S, Qi T, Huang H, Ren Q, Wu D, Chang C, Peng W, Liu Y, Peng J, Xie D. The Jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect Jasmonate-regulated stamen development in Arabidopsis. *Plant Cell*. 2011;23(3):1000–13.
48. Kwon CT, Kim SH, Kim D, Paek NC. The Rice floral repressor early flowering1 affects spikelet fertility by modulating gibberellin signaling. *Rice (N Y)*. 2015;8(1):58.
49. Chezem WR, Clay NK. Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry*. 2016;131:26–43.
50. Chen X, Huang H, Qi T, Liu B, Song S. New perspective of the bHLH-MYB complex in jasmonate-regulated plant fertility in Arabidopsis. *Plant Signal Behav*. 2016;11(2):e1135280.
51. Li S. Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal Behav*. 2014;9(1):e27522.
52. Van Dingenen J, Blomme J, Gonzalez N, Inze D. Plants grow with a little help from their organelle friends. *J Exp Bot*. 2016;67(22):6267–81.
53. Hanson MR, Bentolila S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell*. 2004;16 Suppl:S154–69.
54. Zhang L, Wen A, Wu X, Pan X, Wu N, Chen X, Chen Y, Mao D, Chen L, Luan S. Molecular identification of the magnesium transport gene family in *Brassica napus*. *Plant Physiol Biochem*. 2019;136:204–14.
55. Li LG, Sokolov LN, Yang YH, Li DP, Ting J, Pandey GK, Luan S. A mitochondrial magnesium transporter functions in Arabidopsis pollen development. *Mol Plant*. 2008;1(4):675–85.
56. Chen J, Li LG, Liu ZH, Yuan YJ, Guo LL, Mao DD, Tian LF, Chen LB, Luan S, Li DP. Magnesium transporter AtMGT9 is essential for pollen development in Arabidopsis. *Cell Res*. 2009;19(7):887–98.
57. Brugiere S, Kowalski S, Ferro M, Seigneurin-Berny D, Miras S, Salvi D, Ravel S, d'Herin P, Garin J, Bourguignon J, et al. The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. *Phytochemistry*. 2004;65(12):1693–707.
58. Choi DS, Kim NH, Hwang BK. Pepper mitochondrial FORMATE DEHYDROGENASE1 regulates cell death and defense responses against bacterial pathogens. *Plant Physiol*. 2014;166(3):1298–311.
59. Sabar M, Gagliardi D, Balk J, Leaver CJ. ORFB is a subunit of F1F(O)-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Rep*. 2003;4(4):381–6.
60. Qi Z, Shi X, Meng L, Hu G, Song X. Genetic characteristics and stability of fertility restoration genes of K-cytoplasmic male sterile wheat. *J Chin Agr Univ*. 2015;04:02.
61. Yao M, Ye J, Yang Z, Duan Y, Meng L, Yan P, Liu Z, Zhang L, Song X. Abortion feature and fertility restoration of five kinds of cytoplasmic male sterile wheat lines. *J Triticeae Crops*. 2015;35(12):1676–84.
62. Pu CX, Han YF, Zhu S, Song FY, Zhao Y, Wang CY, Zhang YC, Yang Q, Wang J, Bu SL, et al. The Rice receptor-like kinases Dwarf and Runtish SPIKELET1 and 2 repress cell death and affect sugar utilization during reproductive development. *Plant Cell*. 2017;29(1):70–89.
63. Singh A, Ansari MW, Rani V, Singh CP, Shukla A, Pant RC, Tuteja N, Bains G. First evidence of putrescine involvement in mitigating the floral malformation in mangoes: a scanning electron microscope study. *Protoplasma*. 2014;251(5):1255–61.
64. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9.
65. Kim D, Perteau 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(4):R36.

66. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc.* 2012;7(3):562–78.
67. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics.* 2011;12:323.
68. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000;28(1):33–6.
69. Conesa A, Gotz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008;2008:619832.
70. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36(Database issue):D480–4.
71. Gao Y, Xu H, Shen Y, Wang J. Transcriptomic analysis of rice (*Oryza sativa*) endosperm using the RNA-Seq technique. *Plant Mol Biol.* 2013;81(4–5):363–78.
72. Geng X, Ye J, Yang X, Li S, Zhang L, Song X. Identification of proteins involved in carbohydrate metabolism and energy metabolism pathways and their regulation of cytoplasmic male sterility in wheat. *Int J Mol Sci.* 2018;19:324.
73. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method. *Methods (San Diego, Calif.)*. 2001;25(4):402–8.

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

