RESEARCH ARTICLE

Integrative analysis of transcriptomic and proteomic changes related to male sterility in *Tagetes erecta*

Nan Tang^{1,2,3} · Wei Liu⁴ · Wuhua Zhang¹ · Daocheng Tang¹

Received: 7 May 2020 / Revised: 24 August 2020 / Accepted: 18 September 2020 / Published online: 25 September 2020
© Prof. H.S. Srivastava Foundation for Science and Society 2020

Abstract Male sterile and male fertile two-type lines are important in heterosis utilization and breeding in *Tagetes erecta*, but the genes and pathways involved in male sterility are poorly understood. To explore these topics, transcriptome data (by RNA-seq) and proteome data (by iTRAQ) were gathered from flower buds of the male sterile line ‘MS2-2’ and male fertile line ‘MF2-2’ and integrated for a better understanding of the underlying molecular mechanisms of male sterility in *T. erecta*. The RNA-seq procedure generated 285,139,740 clean reads and 63359 unigenes and 6640 differentially expressed genes (DEGs) were identified, of which 4136 were downregulated and 2504 were upregulated in ‘MS2-2’. DEGs related to flower development, pollen development, pollen wall assembly, endogenous hormones and transcription factors were identified. The iTRAQ analysis identified 3950 proteins in total; 789 were differentially expressed proteins (381

upregulated, 408 downregulated), which were mainly annotated to the Ribosome, Carbon metabolism and Biosynthesis of amino acids pathways. An association analysis revealed strong correlation (r Pearson = 0.6019) between the transcriptomic and proteomic data, and 256 and 34 proteins showed the same and opposite expression patterns with regard to their transcripts, respectively. Pathways such as photosynthesis, fatty acid biosynthesis and phenylpropanoid biosynthesis which influence tapetum and pollen development in male sterile plants, were significantly enriched at the transcript and protein levels. Most genes involved in these pathways were downregulated in ‘MS2-2’. The low expression of these genes or functional loss of proteins could be associated with flower development, pollen development and related to changes in fertility in *T. erecta*. This study provided transcriptomic and proteomic information for *T. erecta* that could illuminate the mechanism of male sterility.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12298-020-00886-z>) contains supplementary material, which is available to authorized users.

✉ Daocheng Tang
tangdaocheng6333@163.com

- ¹ Plateau Flower Research Centre, Department of Agriculture and Husbandry, Qinghai University, Xining 810016, People’s Republic of China
- ² State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, People’s Republic of China
- ³ Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, People’s Republic of China
- ⁴ College of Agriculture, Henan University of Science and Technology, Luoyang 471003, People’s Republic of China

Keywords Marigold · Transcriptome sequencing · Proteome · Male sterility

Introduction

Marigold (*Tagetes erecta* L.) is a species of the genus *Tagetes* native to Mexico. It is cultivated worldwide as a landscape plant and cut flower due to its beautiful flowers. And it is also valuable for its medicinal and industrial uses because its flowers are rich in carotenoid pigments (Siri-amornpun et al. 2012). Male sterile two-type lines, which contain 50% male fertile plants and 50% male sterile plants, are the most commonly used male sterile type for heterosis utilization in *T. erecta* breeding. Several pollination control systems have been reported, including genic

male sterility (GMS) and cytoplasmic male sterility (CMS) (Zhao et al. 2013). Genetic analysis showed that male sterility of *T. erecta* is controlled by a pair of recessive genic genes (Hou et al. 2016). Agronomic and cytological studies were performed for floral organ of *T. erecta*. Male sterile flowers have normal and fertile pistils but completely abortive stamens, which makes them convenient for the pollination of sterile plants with pollen from restorer lines (Sreekala and Raghava 2003).

Although there were reports about agronomic traits of male sterile lines, genetic analysis of male sterility and cytological studies of abortive anther, less information about male sterility related genes and pathways in *T. erecta* is known at present. Our previous study constructed suppression subtractive hybridization (SSH) cDNA libraries between male fertile and male sterile flower buds of *T. erecta* to isolate differentially expressed genes. Fifteen candidate genes were identified in which *CDKB2;1*, *AMS*, *LAP3*, *ACOS5* and *CYP703A2* were significantly up-regulated in male fertile plants (Hou et al. 2016). Using bulked segregant analysis, 1 SRAP and 4 AFLP markers which were closely related to male sterility were screened and 1 AFLP-SCAR marker was obtained to distinguish male sterile and male fertile plants (He et al. 2010). *APETALA2* is a gene involved in flower development which was characterized in *Arabidopsis thaliana*. Molecular analysis showed that *APETALA2* ortholog has played important role during flower development of *T. erecta* (Alegría-Mundo et al. 2012). Ye et al. (2016) supposed that male sterility in *T. erecta* is caused by the homeotic conversion of stamens into other floral organs. It was proposed that this homeotic conversion of floral organs was due to mutation in MADS-box genes, such as *AP1*, *AP2*, *AP3*, *PI*, *AG*, *AGL11*, and the *AGL2*-like genes.

As the rapid development of sequencing technology proceeds, genomics, transcriptomics, proteomics, and metabolomics are frequently used to investigate the mechanisms of plant male sterility. RNA-seq is a high-throughput deep-sequencing technology used for transcriptome profiling that provides a more precise measurement of transcript levels and isoforms than other approaches (Wang et al. 2009). It provides a large amount of information and does not require a genomic background, making it very helpful for non-model plants whose genome sequences are not available. Currently, transcriptome analysis of male sterile and male fertile types has been used to study the mechanisms of male sterility in many plants, such as cybrid pummelo (Zheng et al. 2012), cotton (Suzuki et al. 2013; Wei et al. 2013; Zhang et al. 2015), chili pepper (*Capsicum annuum*) (Liu et al. 2013), wheat (Dong et al. 2013; Zhu et al. 2015), cabbage (Guo et al. 2016; Wang et al. 2016), and rapeseed (*Brassica napus*) (Qu et al. 2015). A large number of differentially expressed

genes (DEGs) have been detected, and candidate genes related to male sterility in these species have been discussed. However, the expression of genes is not a simple one-way process from the transcriptome to the proteome but reflects their complex interactions. Therefore, at present, an increasing number of studies are opting to detect the expression levels of the transcriptome and proteome simultaneously (Nakaminami et al. 2014; Xing et al. 2018; Zhuang et al. 2013). Differentially expressed genes and proteins were identified by transcriptomic and proteomic analysis, and candidate genes associated with male sterility were found in different crops (Ji et al. 2018; Xing et al. 2018; Shemeshmayer et al. 2015). iTRAQ (isobaric tags for relative and absolute quantification) is a novel MS-based approach for the relative and absolute quantification of proteins using isobaric tags (Sebastian et al. 2007). It has advantages such as high throughput, high sensitivity, good quantitation and high accuracy, and it is widely used in the quantitative study of proteomics in the fields of botany, microbiology, food science and medicine. In this study, DEGs and DEPs between ‘MF2-2’ and ‘MS2-2’ were identified in the transcriptome and proteome, respectively. And an integrated analysis was performed to gather information that is not accessible through single approach. The results provided important information for identifying genes related to male sterility and further exploring the regulation mechanism of the male sterility in *T. erecta*.

Materials and methods

Plant material and RNA extraction

The genic male sterile two-type line ‘2-2’ of *T. erecta* was used for this study. Male sterile two-type lines have been screened from a number of cultivars and breeding lines since 2006 by the Plateau Flower Research Center of Qinghai University, China. Male fertile plants showed normal floral organs, while male sterile plants showed degenerated petals and stamens. Genetically stable sterile two-type lines were obtained through years of selection on plant height, flower color, flower shape and petal shape. Line ‘2-2’ was the most stable male sterile two-type line and was maintained by sib-mating at Qinghai University, Xining, China (N 36°43′32″, E 101°45′2″, altitude 2310 m). An F1 population was obtained by the self-pollination of a single plant in 2014. The segregation ratio of male sterile and male fertile plants in the population was 1:1. The plants were grown in the experimental field of Qinghai University. Flower buds (5.0 mm in length) of male sterile (‘MS2-2’) and male fertile (‘MF2-2’) plants were collected separately. The sepals of each flower bud were removed, immediately frozen in liquid nitrogen and

stored at -80°C until RNA isolation. Three replicates were performed for all experiments.

The total RNA from above samples was extracted using the TRIzol Kit (Promega, USA) according to the manufacturer's instructions and then purified using RNase-free DNase I (Takara Biotech Incorporation, Otsu, Japan). The quality and purity of RNA samples were verified using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and RNase-free agarose gel electrophoresis.

cDNA library construction and RNA-seq

A total of 5 μg RNA per sample was used for cDNA library preparation. First, mRNA was enriched by oligo (dT) magnetic beads and then cut into short fragments using fragmentation buffer. The short fragments were used as a template to synthesize the first strand of cDNA. Based on the first strand, the second strand of cDNA was synthesized by adding buffer, dNTPs, RNase H and DNA polymerase I. The first- and second-strand cDNA was purified using QIAquick PCR purification kit (Qiagen, Shanghai, China) and eluted with EB buffer for end repair and A tailing. The cDNA fragments were then connected with sequencing adapters and suitable fragments were purified by agarose gel electrophoresis for PCR amplification. The constructed cDNA libraries were sequenced using an Illumina HiSeqTM 2000 at Gene DENOVO (Guangzhou, China).

De novo transcriptome assembly, annotation and differential expression analysis

The raw reads were filtered by trimming adaptors and removing reads containing more than 10% 'N' bases and low-quality reads (reads containing more than 50% bases with a Q-value ≤ 20). The clean short reads were assembled into contigs using the program Trinity (Grabherr et al. 2011) and then clustered into unigenes. To obtain a reference transcriptome sequence for *Tagetes erecta* L., the reads from the male sterile and male fertile flower buds were assembled and annotated.

To determine the protein annotation information of the unigenes, all unigenes were compared to sequences in the protein databases Nr (<http://www.ncbi.nlm.nih.gov/>), Swiss-Prot (<http://www.expasy.ch/sprot>), Kyoto Encyclopedia of Genes and Genomes database (KEGG, <http://www.genome.jp/kegg>) and Cluster of Orthologous Groups (COG, <http://www.ncbi.nlm.nih.gov/COG/>) using BLASTx with an E-value threshold $< 1 \times 10^{-5}$. Gene annotation was conducted using BLAST2GO based on the Nr annotation. The unigenes were then annotated to one of the three functional categories, biological process, molecular function, or cellular component, using WEGO.

Differential expression analysis of RNA-seq expression profiles with three biological replications was performed with edgeR based on the negative binomial distribution. Genes with a p value < 0.05 (adjusted by FDR) were considered differentially expressed.

Protein extraction, digestion and iTRAQ labelling

Proteins were extracted from male fertile and male sterile flower buds according to the cold acetone method. Flower buds were ground to powder in liquid nitrogen and dissolved in 2 mL lysis buffer (8 M Urea, 2% SDS, 1 $\times$ Protease inhibitor cocktail). Solution was ultrasonic processed for 30 min and then centrifuged at 13,000 rpm for 30 min at 4°C . The supernatant was transferred to a new tube and proteins were precipitated using ice-cold acetone at -20°C overnight. Proteins were cleaned with acetone three times and re-dissolved in 8 M Urea. Quality of proteins was examined by SDS-PAGE. A BCA protein assay was performed to detect and quantify the protein concentration of the supernatant. For each sample, 100 μg of protein was transferred into a new tube and adjusted to 100 μL with 8 M urea. Next, 11 μL of 1 M DTT was added to the samples, which was incubated at 37°C for 1 h and centrifuged at 14,000g for 10 min. Then, 120 μL of 55 mM iodoacetamide was added and incubated in the dark for 20 min. The iodoacetamide was removed by centrifuging and the proteins were digested with 1:50 trypsin (Promega) overnight and then centrifuged at 13,500g for 12 min. The proteins were lyophilized and redissolved in 50 μL 500 mM TEAB. Each sample was labeled by iTRAQ Reagents-8Plex (SCIEX) for 2 h under room temperature. Add 100 μL UP water and stop the reaction for 15 min.

LC-MS/MS, protein identification and data analysis

After labeling, the peptide mixtures were pooled and vacuum dried. Sample fractionation was performed using an Ultimate 3000 system (ThermoFisher scientific, MA, USA). The peptide mixture was redissolved in 100 μL buffer A (20 mM ammonium formate, pH 10.0) and loaded onto an XBridge C18 column (4.6 mm $\times$ 250 mm, 5 μm , Waters Corporation, MA, USA). The column was equilibrated for 10 min and the flow rate was maintained at 1 mL/min. The peptides were eluted with 5–45% buffer B (20 mM ammonium formate, 80% CAN, pH 10.0) for 40 min. Fractions were collected every 100 s and pooled into 12 fractions according to collection time.

The peptides were segregated by an Easy-nLC 1000 system (Thermo Fisher Scientific, MA, USA) followed by MS/MS in the Orbitrap Fusion Tribrid (Thermo Fisher Scientific, MA, USA). The column was equilibrated with

buffer C for 10 min. Ten microliters of peptide sample was loaded at 10 μ l/min onto a C18 column (Thermo Scientific Acclaim PepMap C18, 100 μ m $\times$ 2 cm) and then separated by an analytical C18 column (Acclaim PepMap C18, 75 μ m $\times$ 15 cm) for 70 min with a linear gradient from 2 to 40% buffer D (0.1% formic acid acetonitrile).

The raw data were transformed to mgf format using Mascot Distiller. Mascot 2.3.02 was applied to identify and quantify the proteins and searches were made against the database (38,532 sequences). The search parameters were as follows: the type of search was MS/MS Ion; trypsin was chosen as digestion enzyme; the fragment mass tolerance was 0.05 DA and peptide mass tolerance was 20 ppm; carbamidomethylation of cysteine and iTRAQ 8-plex were set as fixed modifications; pyrophosphorylation of glutamine and oxidation of methionine were set as variable modifications. These experimental procedures were performed by Guangzhou Gene denovo Biotechnology Co., Ltd. (Guangzhou, China).

The protein identification filters were set as follows: significance threshold $p < 0.05$ (95% confidence) and ion score or expected cutoff < 0.05 (95% confidence). The filters for protein quantification were set as follows: “median” was chosen as the protein ratio type; the minimum precursor charge was set to 2 and unique spectra ≥ 2 ; median intensities were set for normalization, and outliers were removed automatically. The peptide identification threshold was set as described above. Based on the protein abundance level, proteins were considered differentially expressed between male fertile and male sterile samples when the fold change was larger than 1.2 ($p < 0.05$). All of DEPs were analyzed by GO and KEGG pathway analysis.

Quantitative real-time PCR validation

First-strand cDNA synthesis was performed using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara). The primers (Supplementary Table 1) were designed using Primer Premier. The housekeeping gene *Tub* was selected as a reference gene to normalize the amplification efficiency. The expression levels of 29 DEGs were identified by real-time PCR (LightCycler480, Roche, USA) using 2 $\times$ SuperReal Color PreMix (TianGen) with the following amplification program: 95 $^{\circ}$ C for 5 min, followed by 40 cycles of 95 $^{\circ}$ C for 10 s, 55 $^{\circ}$ C for 20 s and 72 $^{\circ}$ C for 30 s. Each 20 μ l reaction mixture contained 1 μ l of template (100 ng), 10 μ l of 2 $\times$ SuperReal Color PreMix, 0.5 μ l of each primer (10 μ M) and 8 μ l of RNase-free H₂O. Melting curves were recorded by heating from 65 $^{\circ}$ C to 95 $^{\circ}$ C to determine the specificity of the reactions. The Ct value of each reaction was automatically calculated by software LightCycler480 SW version 1.5. The relative

expression levels of the genes were calculated using the $2^{-\Delta\Delta Ct}$ method.

Results

Morphological characteristics of floral organs

A phenotype comparison showed that the flowers and floral organs were normal in ‘MF2-2’ but abnormal in ‘MS2-2’. The flower buds of ‘MF2-2’ are cylindrical (Fig. 1a), and ‘MF2-2’ has a typical capitulum that consists of ray florets in the periphery and disk florets in the center (Fig. 1b, c). Ray florets have sepals, petals and pistils, while disk florets have sepals, petals, stamens and pistils (Fig. 1d). The flower buds of ‘MS2-2’ are conical (Fig. 1e). Its ray florets and disk florets develop into sepal-like structures and its stamens develop into filaments without pollen (Fig. 1f–h).

Overall statistics for transcriptome sequencing

In this study, we obtained 285,139,740 clean reads with a mean length of 125 bp. The Q20 percentage was 94.7%, the N percentage was 0.00% and the GC percentage was 43.73%. A total of 63,359 unigenes were obtained, with an average length of 998.82 bp and an N50 of 1572 bp. The lengths of the unigenes ranged from 228 bp to 12,846 bp, of which 41,679 (65.8%) were in the range of 201 to 1000 bp, 13,874 (21.9%) ranged from 1001 to 2000 bp, and 7806 (12.3%) of the unigenes were longer than 2000 bp (Supplementary Fig. 1). These results showed that the throughput and sequencing quality were high enough for the following analysis. All of the assembled unigenes were BLAST searched against the Nr, Swiss-Prot, KEGG and euKaryotic Orthologous Groups (KOG) databases (Table 1). A total of 36,799 unigenes (58.80%) were annotated, of which 8341 unigenes were annotated in all four databases. In addition, 6424 unique unigenes were annotated in only the Nr database, 29 unigenes in only the KOG database, 12 unigenes in only the Swiss-Prot database and 14 unigenes in only the KEGG database.

Gene Ontology (GO) analysis was carried out based on the Nr annotations. A total of 36,730 unigenes were annotated in Nr, of which 18,966 (51.64%) unigenes were sorted into one or more GO terms. In total, 50,693, 10,362 and 8110 unigenes were assigned to the biological process, cellular component and molecular function groups, respectively (supplementary Fig. 2). Under biological process, unigenes classified into metabolic process (10,846 unigenes) and cellular process (10,362 unigenes) was prominently represented. For cellular component, cell (7680 unigenes) and cell part (7475 unigenes) were highly

Fig. 1 Flower phenotype characterization of the male fertile and male sterile two-type line '2-2' of *T. erecta*. **a** flower bud of 'MF2-2'; **b** and **c** capitulum of 'MF2-2'; **d** disc floret and ray floret of 'MF2-2'; **e** flower bud of 'MS2-2'; **f** and **g** capitulum of 'MS2-2'; **h** disc floret

and ray floret of 'MS2-2', in which petals of ray and disc floret developed into sepal-like structure and the stamens developed into filaments

Table 1 Annotation results for unigenes in four databases

	Total unigenes	Annotated unigenes	Annotated in different database			
			Nr	Swissprot	KOG	KEGG
Number	63,359	36,799	36,730	28,127	22,730	10,296
Percentage (%)		58.80	57.97	44.39	35.87	16.25

represented GO terms, followed by organelle (6010 unigenes). For the molecular function category, the most frequently represented terms were catalytic activity (9097 unigenes) and binding (8108 unigenes).

In addition, all unigenes were BLAST against the KOG database for functional prediction and classification. A total of 38,844 unigenes were grouped into 25 functional classifications (supplementary Fig. 3), in which 7962 unigenes (20.50%) belonged to the largest cluster 'General function prediction only', followed by 'Posttranslational modification, protein turnover, chaperones' (4352, 11.20%) and 'Signal transduction mechanisms' (4344, 11.18%). The smallest cluster was 'Cell motility', and only 28 (0.07%) unigenes were in this group.

To further demonstrate the usefulness of unigenes generated in this study, pathway annotation of unigenes was performed based on the KEGG annotation information. In the transcriptome, a total of 10,296 (16.25%) unigenes were annotated in 123 typical KEGG pathways. Many unigenes were mapped to the metabolic pathways (2741, 19.53%), Biosynthesis of secondary metabolites (1308, 9.32%) and Ribosome pathway (466, 3.32%). Few genes

were annotated in the pathways 'Caffeine metabolism' (5), 'Anthocyanin biosynthesis' (3) and 'Other types of O-glycan biosynthesis' (3).

Identification of DEGs between male sterile and male fertile 'MS2-2'

In total, we obtained a set of 6640 DEGs ('MF2-2' vs. 'MS2-2', 4136 downregulated and 2504 upregulated) for the subsequent identification of genes related to male sterility. GO enrichment analysis was performed to find significantly enriched GO terms among the DEGs. There were 41 significantly enriched GO terms, including 19 in biological process, 14 in cellular component and 8 in molecular function (Fig. 2). In biological process, the significantly overrepresented GO terms were 'metabolic process', 'cellular process' and 'single-organism process'. In cellular component, the major categories included 'cell', 'cell part' and 'organelle'. 'catalytic activity' and 'binding' were major categories annotated for the GO term molecular function. The downregulated DEGs were associated with 38 GO terms, among which 5 GO terms (biological

Fig. 2 GO term enrichment analysis for DEGs in ‘MF2-2’ and ‘MS2-2’. The left y-axis represents the percentage of genes in a category. The right y-axis shows the number of genes in a category

process: ‘rhythmic process’; cellular component: ‘extracellular region’, ‘virion’ and ‘nucleoid’; molecular function: ‘antioxidant activity’) were significantly enriched in only the downregulated DEGs. Three GO terms in the cellular component category, ‘membrane-enclosed lumen’ (2 DEGs), ‘extracellular matrix’ (1 DEG) and ‘extracellular matrix component’ (1 DEG), were significantly enriched in only the upregulated DEGs.

In organisms, different genes intercoordinate with each other to perform their biological functions. Furthermore, DEGs were subjected to KEGG pathway enrichment analysis to determine the biological functions. A total of 113 pathways were enriched in this study and the top 20 enriched pathways are shown in Fig. 3 (Supplementary Table 2). The most significantly enriched pathways were ‘DNA replication’ (Rich factor = 0.39, Qvalue = 0, 28 DEGs) and ‘Stilbenoid, diarylheptanoid and gingerol biosynthesis’ (Rich factor = 0.39, Qvalue = 0, 21 DEGs). These pathways represented the main biochemical pathways and signal transduction pathways which the DEGs were involved in. Among the 113 pathways, 86 pathways were enriched in both upregulated and downregulated DEGs. Four pathways were enriched in only upregulated DEGs, among which the pathway ‘Base excision repair’ involved 7 upregulated DEGs and showed significant enrichment. Twenty-three pathways were enriched in only downregulated DEGs, among which ‘Photosynthesis’ (31 DEGs) and ‘Monoterpenoid biosynthesis’ (4 DEGs) were significantly enriched.

DEGs were further screened corresponding to their biological functions, such as DEGs probably related to flower development, and pollen development, including pollen wall assembly, endogenous hormones and transcription factor-related genes. Six DEGs (Unigene0002325, Unigene0016140, Unigene0061661, Unigene0003810, Unigene0010398 and Unigene0014309) were found to be involved in the process of flower development or floral organ development (Supplementary Table 3). Unigene0002325 was described as *PI* (PIS-TILLATA) gene which determines petal formation, and it was significantly downregulated in ‘MS2-2’. Unigene0016140 and Unigene0061661 were predicted as *EMF* (embryonic flower) and *ELF4* (EARLY FLOWERING 4-like), respectively. They were upregulated in ‘MS2-2’ and showed negative regulation of flower development. There were 3 DEGs assigned to floral organ development. Unigene0003810 was predicted as *RDR6* (RNA-dependent RNA polymerase 6-like), which related to carpel and gynoecium development. Unigene0010398 was predicted as *CRC* (CRABS CLAW-like) which related to carpel development, floral meristem determinacy and style development. Unigene0014309 was suppressor of overexpression of *CONSTANS 1*.

Ten DEGs (Supplementary Table 3) involved in pollen development, pollen wall assembly were downregulated in ‘MS2-2’. Tapetal oleosins are important components of the tapetosomes. There were 4 DEGs (Unigene0011183, Unigene0019462, Unigene0048683, and Unigene0055108) were identified as related to oleosin. Unigene0048683 was

Fig. 3 Bubble chart of the top 20 pathways in KEGG analysis of DEGs

downregulated in ‘MS2-2’ at both gene and protein levels. Unigene0018299 was described as *rf2* (restorer of fertility 2) gene which was downregulated in ‘MS2-2’.

It was noticed that Unigene0020032 was predicted as *AMS* (ABORTED MICROSPORES) gene, which was assigned to developmental process involved in reproduction and was downregulated in ‘MS2-2’. A total of 9 DEGs were annotated as WRKY transcription factors, of which 7 DEGs were downregulated in ‘MS2-2’. Three DEGs were annotated as NAC family, and 2 of which were downregulated. MYB is an important transcription factor family. In total, 6 DEGs were annotated in this family, 4 of which were downregulated in ‘MS2-2’. One DEG (Unigene0018600) was annotated as a bHLH transcription factor that could be involved in the pollen development process. It was significantly downregulated in the male sterile buds. Two WUS transcription factors were significantly downregulated in ‘MS2-2’ (Supplementary Table 4). Several genes were implicated in plant hormone signal transduction (Supplementary Table 5). There were 23 DEGs described as hormone related genes, including 7

IAA-related DEGs, 2 ABA related DEGs, 2 GA related DEGs and 12 ethylene related DEGs. IAA related DEGs showed distinctly downregulation in ‘MS2-2’. While DEGs involved in ABA, GA and ethylene were upregulated in ‘MS2-2’. It was supposed that the accumulation of IAA occurred during late metaphase, therefore, weakening of the auxin signal, which was probably related to male sterility. The differential expression of these DEGs in male sterile and male fertile plants showed that the metabolism of endogenous hormones changed, leading to pollen abortion in male sterile line of *T. erecta*.

Identification of DEPs between male sterile and male fertile ‘MS2-2’

A comparative proteome analysis was performed for ‘MF2-2’ and ‘MS2-2’ by iTRAQ technology to complement the transcriptome study. A total of 188,882 spectra, 39,395 spectra, 27,493 unique spectra, 11,120 peptides, 8929 unique peptides and 3950 proteins were identified. Among the detected proteins, 2311 (58.5%) were related to at least

two unique peptides. The mass of proteins ranged from 10 to 100 kDa, in which proteins in the range of 90–100 kDa accounted for 2.86% (113 proteins). A total of 789 proteins were identified as DEPs between ‘MF2-2’ and ‘MS2-2’ (fold change > 1.2, p value < 0.05), of which 381 DEPs were upregulated and 408 DEPs were downregulated. Consistent with the RNA-seq data, the number of upregulated DEPs was less than the number of downregulated DEPs.

The 789 DEPs were enriched in 41 GO terms, 18 terms in Biological process, 14 terms in Cellular Component and 9 terms in Molecular Function (Fig. 4). For biological process, the DEPs were mainly related to “metabolic process”, “cellular process” and “single-organism process”. Among the cellular component functions, the DEPs were mainly associated with “cell”, “cell part”, “organelle” and “macromolecular complex”. In the molecular function category, the DEPs were mainly associated with “catalytic activity”, “binding” and “structural molecule activity”. Among biological processes, 7 DEPs were enriched in “pollen wall assembly” and “pollen development”. These DEPs included chalcone and stilbene synthase family protein (Unigene0031769), spermidine hydroxycinnamoyl transferase (Unigene0004513), spermidine hydroxycinnamoyl transferase-like (Unigene0053878), short-chain dehydrogenase reductase (Unigene0031860), fatty acyl-CoA reductase (Unigene0059033), cytochrome P450 (Unigene0046185, Unigene0032859). Their corresponding transcripts showed the same expression pattern in the transcriptome data.

KEGG analysis showed that the DEPs were assigned to 95 pathways, of which 76 pathways belonged to metabolism. The most enriched pathways included ribosome (57, 11.22%), glycosphingolipid biosynthesis-glybo series (3, 0.59%), glyoxylate and dicarboxylate metabolism (13, 2.56%), glycerolipid metabolism (8, 1.57%), amino sugar and nucleotide sugar metabolism (15, 2.95%), sulfur metabolism (5, 0.98%), protein processing in endoplasmic reticulum (21, 4.13%), starch and sucrose metabolism (13, 2.56%), nitrogen metabolism (4, 0.79%), protein export (6, 1.18%), regulation of autophagy (3, 0.59%) and sphingolipid metabolism (3, 0.59%) (Fig. 5). This pathway enrichment analysis suggested that the occurrence of male sterility is regulated at both the mRNA and protein levels.

Association analysis of transcriptome and proteome

A global correlation analysis was performed between the proteins and their corresponding transcripts to evaluate the congruence between the transcriptome and proteome. Association analysis (Fig. 6) revealed strong correlation (r Pearson = 0.6019), with 2600 proteins matched to transcripts, of which 651 were DEPs. A total of 439 DEPs included 235 upregulated proteins and 204 downregulated proteins that were consistent with their transcripts. As shown in Fig. 6, the proteins/transcripts (1319) concentrated at the center of the plot (quadrant 5) were not DEPs/DEGs. The proteins in quadrants 4 (459) and 6 (395) were DEPs whose corresponding genes were not differentially expressed. The proteins of quadrants 2 (53) and 8 (84) were

Fig. 4 GO categories for DEPs in the proteome. The y-axis indicates the number of DEPs in a category

Fig. 5 Bubble chart of the top 20 pathways in KEGG analysis of DEPs

not differentially expressed but their corresponding genes were differentially expressed. The proteins in quadrants 1 (18) and 9 (16) had the opposite expression patterns from their transcripts, while the proteins of quadrants 3 (94) and 7 (162) showed the same expression patterns as their transcripts. It suggested that expression of proteins in quadrant 3 and 7 is regulated at the transcriptional level. However, for proteins in quadrants 1, 2, 4, 6, 8 and 9, there was regulatory events at the posttranscriptional level and therefore gene expression cannot fully represent the abundance of these proteins.

Most of the DEPs that showed the same expression pattern as their transcripts were annotated in catalytic activity and metabolic processes. From GO analysis of the DEGs/DEPs downregulated in 'MS2-2', it was found that the GO terms pollen development (GO: 000,955), pollen wall assembly (GO: 0,010,208) and gametophyte development (GO: 0,048,229) were significantly enriched. It suggested that pollen formation and development was defective in male sterile plant. Ten DEGs (except Unigene0009712) involved in pollen development and pollen

wall assembly were downregulated in 'MS2-2'. It is consistent with result of transcriptome analysis.

KEGG pathway analysis was performed to explore which metabolic pathways were associated with DEGs/DEPs. It was found that the proteins were mainly involved in metabolism processes, such as carbohydrate metabolism, lipid metabolism, glycan biosynthesis and metabolism, biosynthesis of other secondary metabolites, energy metabolism, metabolism of terpenoids and polyketides, metabolism of cofactors and vitamins, etc. Upregulated DEG/DEP pairs were mainly enriched in starch and sucrose metabolism, DNA replication, and Amino sugar and nucleotide sugar metabolism. Downregulated DEG/DEP pairs were mainly enriched in the photosynthesis and phenylpropanoid biosynthesis pathways. In quadrant 3, 94 unigenes showed a positive relationship between mRNA enrichment and protein abundances. Pathways Starch and sucrose metabolism, DNA replication, amino sugar and nucleotide sugar metabolism and pentose, and glucuronate interconversions were significantly enriched. Pollen development process is a metabolic process that involves

Fig. 6 Correlation between proteins and transcripts in ‘MS2-2’ and ‘MF2-2’. Note: black dots represent proteins/genes that were not differentially expressed; blue dots represent differentially expressed proteins with corresponding genes that were not differentially expressed; green dots represent differentially expressed genes with corresponding proteins that were not differentially expressed; red dots represent DEP/DEG pair

starch and sugar accumulation. Therefore, it is supposed that DEGs/DEPs involved in starch and sucrose metabolism pathway are related to pollen development and probably influence fertility of marigold. In quadrant 7, 162 DEGs/DEPs which downregulated in ‘MS2-2’ showed that pathways photosynthesis, phenylpropanoid biosynthesis, carotenoid biosynthesis, glycosphingolipid biosynthesis-ganglio series, peroxisome, carbon fixation in photosynthetic organisms, fatty acid biosynthesis and fatty acid degradation were significantly enriched (Table 2). These pathways are sporopollenin biosynthesis related pathways and key enzymes were significantly downregulated in “MS2-2”. In the Phenylpropanoid biosynthesis, PAL converts phenylalanine to cinnamic acid and then functions upstream of this pathway. And lignins were produced under the catalysis of POD at the end. In this study, 2 PAL genes, 1 POD gene and 1 REF gene were expressed at a very low level in “MS2-2”. Unigenes involved in energy metabolism in plants such as photosynthesis and carbon fixation in photosynthetic organisms showed low expression level in male sterile plants, such as key genes *psbS* (Unigene0014436), *psaD* (Unigene0019585), *PetE* (Unigene0024525), *ATPFIG* (Unigene0027109), *pasN* (Unigene0028519) and *psaG* (Unigene0042466). The expression pattern of DEPs was the same. It suggested that plant need more energy during anther and pollen development. And these genes may be important for resulting in male sterility in marigold.

Validation of RNA-seq-based DEGs by qRT-PCR

Twenty-nine DEGs were selected for validation of sequencing results by qRT-PCR. The expression levels of each selected gene in both ‘MF2-2’ and ‘MS2-2’ were measured and compared with the sequencing data. The RNA-seq data (FPKM value) and qRT-PCR data for each selected DEGs were positively correlated at 0.6692 (Fig. 7, correlation is significant at the 0.01 level).

Discussion

Male sterility is an effective and economical system for pollination control, and it plays an important role in hybrid seed production. Although male sterile and male fertile two-type lines have been utilized in F1 hybrid production in *T. erecta* for many years, less information about male sterility related genes and pathways is known. We have previously identified 15 candidate genes, which were classified into protein kinase genes and hormone-related genes, transcription factors and flower transition genes, cell cycle genes, and pollen wall formation-related genes (Hou et al. 2016). In order to further explore the potential regulation mechanism of male sterility in *T. erecta*, we have performed transcriptomic and proteomic analysis for male sterile and male fertile flower buds in this study. A total of 6640 DEGs and 789 DEPs were identified through transcriptome and proteome, respectively. Both the transcriptomic and proteomic data are important in understanding of molecular processes involved in fertility conversion. A reasonably high correlation between the transcriptome and proteome ($r = 0.6019$) was found in this study. A total of 256 proteins showed the same expression patterns as their transcripts and 34 proteins presented the opposite expression to their transcripts. The integrative analysis highlighted a set of genes/proteins that were possibly involved in the male sterile trait formation of *T. erecta*.

Flower development related genes

Flower development proceeds through floral induction, floral meristem initiation, and the formation of floral organ primordia as well as the differentiation of the primordia into floral organs. Many genes/proteins interact with each other and function in a complex network to control flower development. Previous studies have shown that the spontaneous homeotic conversion of floral organs was the underlying cause of male sterility in marigold (He et al. 2010). In our study, ray florets and disk florets of ‘MS2-2’ develop into sepal-like structures and its stamens develop into filaments without pollen. The defects of stamens provide biological basis for male sterility, including defects

Table 2 DEGs/DEPs that downregulated in ‘MS2-2’ and significantly enriched pathways

ID	Pathway	DEGs/DEPs in the pathways
ko00195	Photosynthesis	psbS (Unigene0014436) psaD (Unigene0019585) PetE (Unigene0024525) ATPF1G (Unigene0027109) pasN (Unigene0028519) psaG (Unigene0042466)
ko00940	Phenylpropanoid biosynthesis	POD (Unigene0007638) REF1 (Unigene0032810) PAL(Unigene0038746, Unigene0039337)
ko00906	Carotenoid biosynthesis	PDS (Unigene0017197) LUT1 (Unigene0020147)
ko00604	Glycosphingolipid biosynthesis-ganglio series	HEXA_B (Unigene0044926)
ko04146	Peroxisome	AGXT (Unigene0024356) HAO (Unigene0044931) ACSL (Unigene0045649)
ko00710	Carbon fixation in photosynthetic organisms	GPT (Unigene0002618) GAPA (Unigene0021663, Unigene0009851) rbcL (Unigene0060550)
ko00061	Fatty acid biosynthesis	FabH (Unigene0025852) ACSL (Unigene0045649) ACACA (Unigene0047774)
ko00071	Fatty acid degradation	ADH (Unigene0000015) ACSL (Unigene0045649) ALDH (Unigene0045852)

Fig. 7 Correlation between RNA-seq and qRT-PCR results for selected genes

in anther and pollen development. In the transcriptome data of this study, many genes have been identified related to

flower development, including *PI* (Unigene0002325), *EMF* (Unigene0016140), *ELF4* (Unigene0061661). *PI*

determines petal formation and its expression level in ‘MS2-2’ was significantly lower than that in ‘MF2-2’ (validated by qRT-PCR). In *Arabidopsis thaliana*, *PI* genes belong to class B in the ABCDE model of flower development. Class B genes control the petal identity in the second whorl (Goto and Meyerowitz 1994). Mutations in the *PI* gene cause homeotic conversion of petals to sepals and of stamens to carpels (Vandenbussche et al. 2004). Meanwhile, *EMF* and *ELF4* were found upregulated in ‘MS2-2’. It was known that *EMF* had negative regulation of flower development. During the process of flower induction, *EMF* is required for the inhibition of reproductive growth. In *Arabidopsis*, the *EMF* genes are required to maintain vegetative development and repress flower development (Moon et al. 2003). Other genes such as *CRC* (Unigene0010398), *RDR6* (Unigene0003810) and *CON-STANS1* (Unigene0014309) also showed significant different expression between ‘MS2-2’ and ‘MF2-2’. These genes are regulators of carpel development, floral meristem determinacy and style development, and therefore play important role during flower development (Basherudin et al. 2013; Thomas et al. 2018). Besides, FLOWERING LOCUS T (*FT*)/TERMINAL FLOWER 1 (*TFL1*) gene family also plays an important role in the flower development process. Unigene0007106 was annotated as an anti-florigenic *FT/TFL1* family in *T. erecta* and presented significantly higher expression in male fertile plant. *FT/TFL1* family genes encode proteins with high similarity to phosphatidyl ethanolamine-binding proteins, which function as flowering promoters or repressors (Li et al. 2015). The differential expression of these genes in flower bud of *T. erecta* may inhibit normal development of floral organ, which is related to formation of sterility.

Pollen development related Genes

The development of pollen is much influenced by surrounding tissues such as tapetum of anther wall. The tapetum cells surround the pollen and secrete sporopollenin during pollen development. Abnormal development of tapetum affects microspores development and ultimately causes male sterility (Ye et al. 2017). In the transcriptome data, DEGs involved in pollen development and pollen wall assembly were identified and most of them were downregulated in ‘MS2-2’. It was noticed that *AMS* gene (Unigene0020032) showed significant low expression in ‘MS2-2’. A previous study showed that *AMS*, encoding a bHLH protein, plays an important role in tapetum development and microspore development (Sorensen et al. 2003). It regulates pollen wall development and sporopollenin biosynthesis (Ferguson et al. 2017; Xu et al. 2010). In *Arabidopsis thaliana*, *AMS* regulates 23 genes involved in sporopollenin precursor synthesis and its mutants released defective microspores (Xu et al.

2014). Therefore, it is deduced that the low expression of *AMS* in the male sterile line ‘MS2-2’ may inhibit the synthesis of sporopollenin, regulate pollen development, and cause male sterility in *T. erecta*. Besides, 7 WRKY, 2 NAC, 4 MYB, 2 WUS and 1 bHLH transcription factors were significantly downregulated in ‘MS2-2’. These transcription factors could be involved in the pollen development process. In model plants, it has been reported that transcription factors play important role in regulating the development and function of tapetum, pollen or stamen (Liu et al. 2015; Lei et al. 2017; Yang et al. 2017). In our proteomic data, 7 proteins have been identified for pollen development and pollen wall assembly and these DEPs showed significant low expression in ‘MS2-2’. Their corresponding transcripts presented the same expression pattern in transcriptome data. The identification of these differentially expressed genes/proteins provided a lot of information to reveal the mechanism of male sterility in *T. erecta*.

By combining the transcriptome and proteome, the GO terms pollen development, pollen wall assembly and gametophyte development were significantly enriched. Sporopollenin is the main component of the pollen exine and that its biosynthesis is closely related to fatty acid and phenylpropanoid metabolism (Liu and Fan 2013; Shi et al. 2015). Basic metabolic pathways such as fatty acid and phenylpropanoid pathways are essential for anther development. The pathway fatty acid metabolism, fatty acid biosynthesis and fatty acid degradation were enriched in this study. Key genes and proteins (Table 2) that participated in fatty acid metabolism were significantly downregulated in ‘MS2-2’. The low expression of these genes or functional loss of proteins could be associated with sporopollenin biosynthesis and cause exine defects or male sterility. Three key genes (*POD*, *REF1* and *PAL*) related to phenylpropanoid metabolism were identified in this study and they were downregulated in ‘MS2-2’. *POD*, *REF1* and *PAL* are enzymes that are active in the phenylpropanoid biosynthesis pathway and may regulate metabolic pathways related to cell wall formation in pollen grains. Many studies showed that phenylpropanoid is very important for cell wall formation in pollen grains and is crucial for normal anther development (Matsuda et al. 1996). Low expression of these genes could be related to abnormal development of pollen wall and the formation of male sterility (Wu et al. 2015). In addition, enzymes such as pectinesterase, polygalacturonase, pectate lyase and Beta-galactosidase are considered to be involved in the pollen wall assembly process and its regulation (Huang et al. 2009). A total of 41 enzyme genes were identified in this study, including 10 *PME* (pectinesterase), 17 *PG* (polygalacturonase), 5 *PL* (pectate lyase) and 9 *βGAL* (beta-galactosidase) genes. Most of these identified DEGs showed low expression in ‘MS2-2’. Among these enzymes, PMEs are cell wall associated proteins which play important roles

in microspore development and pollen tube elongation (Tian et al. 2006). PLs have significant effect in germination and elongation of pollen tube (Jiang et al. 2014). PGs also influence the formation of pollen grain and elongation of pollen tube (Ogawa et al. 2009). These DEGs/DEPs play key roles in development of pollen and may be related to changes in fertility in *T. erecta*. It provides information for further investigation of mechanisms of pollen development, pollen tube elongation and fertility conversion.

Conclusions

To explore the molecular mechanism of male sterility in *T. erecta*, flower buds of male sterile line “MS2-2” and its isogenic maintainer line “MF2-2” were collected for transcriptomic and proteomic analysis. A total of 6640 DEGs and 789 DEPs were identified in the transcriptome and proteome, respectively. A reasonably strong correlation (r Pearson = 0.6019) was found between the transcriptomic and proteomic data. Six hundred and fifty one DEPs matched to their transcripts, in which 256 and 34 DEPs showed the same and opposite expression patterns with regard to their transcripts, respectively. Function annotation analyses suggested that these DEGs/DEPs are involved in flower development, pollen development and probably related to the conversion of fertility in *T. erecta*. The transcriptomic and proteomic information provided in this study will improve our understanding of genes and pathways associated with male sterility in *T. erecta* and contribute to illuminating the mechanism of male sterility.

Acknowledgements We are thankful for the support from National Natural Science Foundation of China (31660582) and the foundation from Science and Technology Department of Qinghai Province (2018-NK-102). The study was also supported by the Key Laboratory of Landscape Plants of Qinghai Province and High level innovative thousand talents program of Qinghai Province (2016).

Authors' contributions Dr. Nan Tang and Professor Daocheng Tang conceived and designed the study. Nan Tang performed the experiments, data analysis and wrote the manuscript. Wuhua Zhang performed qPCR validation. Daocheng Tang reviewed and edited the manuscript. Dr. Wei Liu contributed to the revision of the manuscript.

Funding The research was supported by National Natural Science Foundation of China (31660582) and the foundation from Science and Technology Department of Qinghai Province (2018-NK-102). The study was also supported by the Key Laboratory of Landscape Plants of Qinghai Province and High level innovative thousand talents program of Qinghai Province (2016).

Availability of data and material We declared that materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes, without breaching participant confidentiality.

Compliance with ethical standards

Conflict of interest We have no conflicts of interest associated with this publication.

Consent to participate Not applicable.

Consent for publication The manuscript has been read and approved by all named authors.

Code availability Not applicable.

References

- Alegria-Mundo H, Yong L, Cruz-Ramírez A, Herrera-Estrella L, Cruz-Hernandez A (2012) Molecular analysis of marigold (*Tagetes erecta*) *apetala2* in flower development. *Acta Hort* 929:293–298
- Basherudin N, Haron MR, Adnan N, Muhammad N (2013) Isolation and expression of CONSTANS-like gene in flowering tissues of teak (*Tectona grandis*). *Asia Pac J Mol Biol Biotechnol* 21(3):89–96
- Dong J, Liu Y, Liu HW, Wang XL, Song YZ, Wang Q, Niu N, Ma SC, Zhang GS, Xu H, Wang JW (2013) Transcriptomic profiling of physiological male sterility induced by chemical hybridizing agent SQ-1 in wheat (*Triticum aestivum*). *Chin J Biochem Mol Biol* 29:948–955
- Ferguson AC, Pearce S, Band LR, Yang C, Ferjentsikova I, King J, Yuan Z, Zhang D, Wilson ZA (2017) Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in Arabidopsis. *New Phytol* 213:778–790
- Goto K, Meyerowitz EM (1994) Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev* 8:1548
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Xian A, Lin F, Raychowdhury R, Zeng Q (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* 29:644–652
- Guo J, Zhang Y, Hui M, Ya Cheng, Zhang E, Xu Z (2016) Transcriptome sequencing and de novo analysis of a recessive genic male sterile line in cabbage (*Brassica oleracea* L. var. *capitata*). *Mol Breed* 36:117
- He YH, Ning GG, Sun YL, Hu Y, Zhao XY, Bao MZ (2010) Cytological and mapping analysis of a novel male sterile type resulting from spontaneous floral organ homeotic conversion in marigold (*Tagetes erecta* L.). *Mol Breed* 26:19–29
- Hou Z, Tang D, Tang N (2016) Isolation and analysis of differentially expressed genes between male fertile and male sterile flower buds of marigold (*Tagetes erecta* L.). *Pak J Bot* 48:2423–2431
- Huang L, Cao J, Zhang A, Ye Y, Zhang Y, Liu T (2009) The polygalacturonase gene *BcMF2* from *Brassica campestris* is associated with intine development. *J Exp Bot* 60:301–313
- Ji JL, Yang L, Fang Z, Zhuang M, Zhang Y, Lv H, Liu Y, Li Z (2018) Complementary transcriptome and proteome profiling in cabbage buds of a recessive male sterile mutant provides new insights into male reproductive development. *J Proteomics* 179:80–91
- Jiang J, Yao L, Yu Y, Lv M, Miao Y, Cao J (2014) PECTATE LYASE-LIKE10 is associated with pollen wall development in *Brassica campestris*. *J Integr Plant Biol* 56(11):1095–1105
- Lei R, Li X, Ma Z, Lv Y, Hu Y, Yu D (2017) Arabidopsis WRKY₂ and WRKY₃₄ transcription factors interact with VQ₂₀ protein to modulate pollen development and function. *Plant J* 91(6):962–976

- Li C, Luo L, Fu Q, Niu L, Xu Z (2015) Identification and characterization of the *FT/TFL1* gene family in the biofuel plant *Jatropha curcas*. *Plant Mol Biol Rep* 33:326–333
- Liu L, Fan XD (2013) Tapetum: regulation and role in sporopollenin biosynthesis in *Arabidopsis*. *Plant Mol Biol* 83:165–175
- Liu C, Ma N, Wang PY, Nan F, Shen HL (2013) Transcriptome sequencing and De Novo analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). *PLoS ONE* 8:e65209
- Liu YM, Zhang L, Zhou JY, Cao MJ (2015) Research progress of the bHLH transcription factors involved in genic male sterility in plants. *Hereditas* 37(12):1194–1203
- Matsuda N, Tsuchiya T, Kishitani S, Tanaka Y, Toriyama K (1996) Partial male sterility in transgenic tobacco carrying antisense and sense *PAL* cDNA under the control of a tapetum-specific promoter. *Plant Cell Physiol* 37:215–222
- Moon YH, Chen L, Pan RL, Chang HS, Zhu T, Maffeo DM, Sunga ZR (2003) EMF genes maintain vegetative development by repressing the flower program in *Arabidopsis*. *Plant Cell* 15:681–693
- Nakaminami K, Matsui A, Nakagami H, Minami A, Nomura Y, Tanaka M, Morosawa T, Ishida J, Takahashi S, Uemura M, Shirasu K, Motoaki S (2014) Analysis of differential expression patterns of mRNA and protein during cold-acclimation and de-acclimation in *Arabidopsis*. *Mol Cell Proteomics* 13:3602
- Ogawa M, Kay P, Wilson S, Swain SM (2009) ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (*ADPG1*), *ADPG2*, and *QUARTET2* are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant Cell* 21:216–233
- Qu C, Fu F, Liu M, Zhao H, Liu C, Li J, Tang Z, Xu X, Qiu X, Wang R (2015) Comparative transcriptome analysis of recessive male sterility (RGMS) in sterile and fertile *Brassica napus* lines. *PLoS ONE* 10:e0144118
- Sebastian W, Kai AR, Meyer HE, Bettina W (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7:340–350
- Shemeshmayer E, Benmichael T, Rotem N, Rabinowitch HD, Doronfaigenboim A, Kosmala A, Perlikowski D, Sherman A, Kamenetsky R (2015) Garlic (*Allium sativum* L.) fertility: transcriptome and proteome analyses provide insight into flower and pollen development. *Front Plant Sci* 6:271
- Shi J, Cui M, Yang L, Kim YJ, Zhang D (2015) Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci* 20(11):741–753
- Siriamornpun S, Kaisoon O, Meeso N (2012) Changes in colour, antioxidant activities and carotenoids (lycopene, β -carotene, lutein) of marigold flower (*Tagetes erecta* L.) resulting from different drying processes. *J Funct Foods* 4:757–766
- Sorensen AM, Kröber S, Unte US, Huijser P, Saedler H (2003) The *Arabidopsis* aborted microspores (*ams*) gene encodes a MYC class transcription factor. *Plant J Cell Mol Biol* 33:413–423
- Sreekala C, Raghava SPS (2003) Exploitation of heterosis for carotenoid content in African marigold (*Tagetes erecta* L.) and its correlation with esterase polymorphism. *Theor Appl Genet* 106:771–776
- Suzuki H, Rodriguezuribe L, Xu J, Zhang J (2013) Transcriptome analysis of cytoplasmic male sterility and restoration in CMS-D8 cotton. *Plant Cell Rep* 32:1531–1542
- Thomas G, Suvi B, Annette B (2018) CRABS CLAW acts as a bifunctional transcription factor in flower development. *Front Plant Sci* 9:835
- Tian GW, Chen MH, Zaltsman A, Citovsky V (2006) Pollen-specific pectin methylesterase involved in pollen tube growth. *Dev Biol* 294:83–91
- Vandenbussche M, Zethof J, Royaert S, Weterings K, Gerats T (2004) The duplicated b-class heterodimer model: whorl-specific effects and complex genetic interactions in petunia hybrida flower development. *Plant Cell* 16:741–754
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wang S, Wang C, Zhang XX, Chen X, Liu JJ, Jia XF, Jia SQ (2016) Transcriptome de novo assembly and analysis of differentially expressed genes related to cytoplasmic male sterility in cabbage. *Plant Physiol Biochem* 105:224–232
- Wei M, Song M, Fan S, Yu S (2013) Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. *BMC Genom* 14:97
- Wu Y, Liu M, Wu Z, Yang L, Yang X, Yuan D, Guo X, Zhang X (2015) Defective pollen wall contributes to male sterility in the male sterile line 1355a of cotton. *Sci Rep* 5:9608
- Xing M, Sun C, Li H, Hu S, Kang J (2018) Integrated analysis of transcriptome and proteome changes related to the Ogura cytoplasmic male sterility in cabbage. *PLoS ONE* 13:e0193462
- Xu J, Yang C, Yuan Z, Zhang D, Gondwe MY, Ding Z, Liang W, Zhang D, Wilson ZA (2010) The ABORTED MICROSPORES regulatory network is required for postmeiotic male reproductive development in *Arabidopsis thaliana*. *Plant Cell* 22:91–107
- Xu J, Ding Z, Vizcay-Barrena G, Shi J, Liang W, Yuan Z, Werck-Reichhart D, Schreiber L, Wilson ZA, Zhang D (2014) ABORTED MICROSPORES acts as a master regulator of pollen wall formation in *Arabidopsis*. *Plant Cell* 26:1544–1556
- Yang C, Song J, Ferguson A, Klisch D, Simpson K, Mo R, Taylor B, Mitsuda N, Wilson ZA (2017) Transcription factor MYB26 is key to spatial specificity in anther secondary thickening formation. *Plant Physiol* 175(1):333–350
- Ye A, Zhang Q, Wang W, Zhang C, Zhe C, Bao M, He Y (2016) Transcriptomic analysis of differentially expressed genes during flower organ development in genetic male sterile and male fertile *Tagetes erecta* by digital gene-expression profiling. *PLoS ONE* 11:e0150892
- Ye J, Duan Y, Hu G, Geng X, Zhang G, Yan P, Liu Z, Zhang L, Song X (2017) Identification of candidate genes and biosynthesis pathways related to fertility conversion by wheat KTM3315A transcriptome profiling. *Front Plant Sci* 8:449
- Zhang Y, Chen J, Liu J, Xia M, Wang W, Shen F (2015) Transcriptome analysis of early anther development of cotton revealed male sterility genes for major metabolic pathways. *J Plant Growth Regul* 34:223–232
- Zhao D, Xiao Y, Hai Y, Sheng W (2013) Cytological investigation of anther development in DGMS Line Shaan-GMS in *Brassica napus* L. *Czech J Genet Plant Breed* 49:16–23
- Zheng B, Wu X, Ge X, Deng X, Grosser J, Guo W (2012) Comparative transcript profiling of a male sterile *Cybid pummelo* and its fertile type revealed altered gene expression related to flower development. *PLoS ONE* 7:e43758
- Zhu Q, Song Y, Zhang G, Ju L, Zhang J, Yu Y, Niu N, Wang J, Ma S (2015) De novo assembly and transcriptome analysis of wheat with male sterility induced by the chemical hybridizing agent SQ-1. *PLoS ONE* 10:e0123556
- Zhuang W, Gao Z, Wang L, Zhong W, Ni Z, Zhang Z (2013) Comparative proteomic and transcriptomic approaches to address the active role of GA4 in Japanese apricot flower bud dormancy release. *J Exp Bot* 64:4953–4966

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