

RESEARCH ARTICLE

Transcriptomic Analysis of Differentially Expressed Genes during Flower Organ Development in Genetic Male Sterile and Male Fertile *Tagetes erecta* by Digital Gene-Expression Profiling

Ye Ai^{1,2}, Qinghua Zhang³, Weining Wang⁴, Chunling Zhang¹, Zhe Cao⁴, Manzhu Bao^{1*}, Yanhong He^{1*}

1 Key Laboratory of Horticultural Plant Biology, Ministry of Education, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, Hubei, China, **2** College of Landscape Architecture, Fujian Agriculture and Forestry University, 15# Shangxiadian Road, Cangshan District, Fuzhou 350002, Fujian, China, **3** College of Forestry, Fujian Agriculture and Forestry University, 15# Shangxiadian Road, Cangshan District, Fuzhou 350002, Fujian, China, **4** Gulf Coast Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, Wimauma, Florida 33598, United States of America

* mzbaoy@mail.hzau.edu.cn (MB); hyh2010@mail.hzau.edu.cn (YHH)



CrossMark
click for updates

OPEN ACCESS

Citation: Ai Y, Zhang Q, Wang W, Zhang C, Cao Z, Bao M, et al. (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(3): e0150892. doi:10.1371/journal.pone.0150892

Editor: Serena Aceto, University of Naples Federico II, ITALY

Received: November 14, 2015

Accepted: February 19, 2016

Published: March 3, 2016

Copyright: © 2016 Ai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. The raw reads of our transcriptome (1 record) and DGE (11 records) data were deposited in the Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra/>) under accession number SRP066084.

Funding: This research was supported by grants from the Fundamental Research Funds for the Central Universities (2013PY081), National Natural Science Foundation of China (31201647) (<http://www.>

Abstract

Tagetes erecta is an important commercial plant of Asteraceae family. The male sterile (MS) and male fertile (MF) two-type lines of *T. erecta* have been utilized in F₁ hybrid production for many years, but no report has been made to identify the genes that specify its male sterility that is caused by homeotic conversion of floral organs. In this study, transcriptome assembly and digital gene expression profiling were performed to generate expression profiles of MS and MF plants. A cDNA library was generated from an equal mixture of RNA isolated from MS and MF flower buds (1 mm and 4 mm in diameter). Totally, 87,473,431 clean tags were obtained and assembled into 128,937 transcripts among which 65,857 unigenes were identified with an average length of 1,188 bp. About 52% of unigenes (34,176) were annotated in Nr, Nt, Pfam, KOG/COG, Swiss-Prot, KO (KEGG Ortholog database) and/or GO. Taking the above transcriptome as reference, 125 differentially expressed genes were detected in both developmental stages of MS and MF flower buds. MADS-box genes were presumed to be highly related to male sterility in *T. erecta* based on histological and cytological observations. Twelve MADS-box genes showed significantly different expression levels in flower buds 4 mm in diameter, whereas only one gene expressed significantly different in flower buds 1 mm in diameter between MS and MF plants. This is the first transcriptome analysis in *T. erecta* and will provide a valuable resource for future genomic studies, especially in flower organ development and/or differentiation.

nsfc.gov.cn/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Introduction

Plants with male sterility have been applied effectively and economically in plant breeding for pollination control, especially in Asteraceae family, which has the unique structure of terminal capitulum that contains hundreds of florets of two different types, ray florets in the periphery and disk florets in the center. Breeders are looking for the male sterile (MS) plants with defective anthers, and degenerated petals of ray and disc florets to save the expense on manual emasculation [1, 2]. *Tagetes erecta*, a member of the Asteraceae family, is an important commercial plant used for ornamental, industrial and medicinal purposes [3–5]. Fortunately, MS line of *T. erecta* was found in nature, in which the petals of florets developed into filament-like structures and the stamens became yellow filaments with no pollen formed [6]. The degeneration of petals and stamens seems to be a perfect trait for pollination control and the MS lines of *T. erecta* have been utilized successfully in F₁ hybrid production [7, 8].

The associated phenotypic manifestations of male sterility include the absence or abnormality of male organs, failure to form normal sporogenous tissues, pollen abortion, failure of stamen dehiscence, and the inability of mature pollen to germinate on compatible stigma [9, 10]. The previous histological and cytological analysis found that, in *T. erecta*, the petals of the ray and disc florets of the MS plant developed into sepal-like, while the stamens were partially converted to styles [11]. It indicated that the male sterility in *T. erecta* is probably caused by the homeotic conversion of stamens into other floral organ structures, i.e. corresponding to the category of male organ abnormality. Based on the ABCDE model of floral organ development, the homeotic conversion of floral organs is due to the mutation of MADS-box A-, B-, C-, D- and E-class genes [12]. The homeotic conversion in *T. erecta* might be, at least in part, the result of mutation of MADS-box genes [11]. However, this suggestion needs to be further investigated and validated. And more studies are needed to elucidate the molecular mechanism of male sterility in *T. erecta*.

Next generation sequencing techniques had improved the efficiency and reduced the cost of sequencing, hence accelerated gene expression profile comparison and gene discovery [13]. Transcriptome assembly is a valuable tool to study transcriptomics, in which the expressed genes can almost cover the entire transcriptome when assembled together [14, 15]. Digital gene expression (DGE) analysis, on the other hand, is a powerful tool to identify and quantify gene expression on the whole genome level, in which differentially expressed genes and their related pathways can be analyzed comprehensively [16–19]. Combining transcriptome assembly and DGE approaches has facilitated the identification of candidate genes in non-model plants, as it takes the advantages of both, not only enabling large scale gene functional assignment via large sequenced transcriptome library assembly, but making it possible to easily perform quantitative gene expression comparisons without potential biases, thus allowing for a more sensitive and accurate profiling of the transcriptome that more closely resembles cell activity [20–22].

There were many reports about the use of transcriptome assembly and DGE techniques to study the mechanism of male sterility. In sterile *Cybrid Pummelo* (Rutaceae family), a large number of differentially expressed genes were identified at both petal primordia and stamen primordia stages [23]. In *Capsicum annuum* (Solanaceae family), a set of potential candidate genes were found to associate with the formation or abortion of pollen between a cytoplasmic MS line and its near-isogenic restorer line [24]. In sterile *Brassica napus* (Brassicaceae family), many genes were identified to be involved in pollen tube development and growth, pollen wall assembly and modification, pollen exine formation and pollination [25]. In *Gossypium hirsutum* (Malvaceae family), thousands of genes were differentially expressed at the meiosis, tetrad, and uninucleate microspore stages of anthers [26, 27]. These findings provided a better understanding of the regulatory network involved in stamen, anther and pollen development. To our



Fig 1. . Morphological characteristics of the male sterile and male fertile two-type line M525AB of *T. erecta*. (a) Plant morphology of male sterile plant M525A (right) and male fertile plant M525B (left); (b) Inflorescence morphology of male fertile plant M525B; (c) Inflorescence morphology of male sterile plant M525A. RF: ray floret, DF: disc floret, SF: sterile floret.

doi:10.1371/journal.pone.0150892.g001

knowledge, in Asteraceae family, there has been no transcriptomic analysis of differentially expressed genes related to spontaneous male sterility caused by homeotic conversion.

To generate more complete observations of transcriptome content and find out candidate genes associated with male sterility in *T. erecta*, we constructed a reference transcriptome for flower buds of *T. erecta* using Illumina Sequencing. Further, we used DGE analysis to compare the gene expression level between the MS and male fertile (MF) flower buds when they grew to 1 mm and 4 mm in diameter. This is the first genome-wide gene expression profiling of male sterility in *T. erecta*. The data will provide an invaluable resource for identifying genes involved in flower development and provide insights into the molecular mechanisms of male sterility in *T. erecta*.

Materials and Methods

Materials

The genic MS and MF two-type line M525AB of *T. erecta*, derived from an individual natural mutant found in 2004, was maintained by sib-mating [11]. The MS plant named as M525A displayed degenerated petals and stamens (Fig 1A and 1C), while the MF plant labelled as M525B exhibited normal floral organs (Fig 1A and 1B). An F₁ segregation population was obtained by self-pollination of a single plant M525B in 2013. When two pairs of true leaves emerged, the homozygous MS and homozygous MF plants were identified by the SCAR maker SC4 [11]. Plants were grown in the experimental field of Huazhong Agricultural University (located at 30°28'36.5" North latitude and 114°21'59.4" East longitude), Wuhan, Hubei Province, China.

Morphology observation

When the plants came into bloom, floret organs and different sizes (0.5 mm, 1 mm, 2 mm, 3 mm, 4 mm, 5 mm, 6 mm, 7 mm and 8 mm) of flower buds from the homozygous MS and homozygous MF plants were collected for morphology observation under a microscope (BX61, Olympus). The floret organs were also examined under a JEOL (JSM-6390LV) scanning

electron microscope (SEM) in the Electron Microscopy Laboratory of Huazhong Agricultural University. The operation steps of SEM have been described in detail by Ai et al [28].

RNA isolation

Based on the morphological analysis, flower buds (1 mm and 4 mm in diameter) were collected from ten homozygous MS plants and ten homozygous MF plants, respectively. Collected buds were frozen immediately in liquid nitrogen and stored at -80°C for RNA extraction. Flower buds were sampled three times (representing three replications) with an interval of ten days in May 2014. Total RNA from each sample was isolated by the Trizol Reagent (Invitrogen), and RNA quality and quantity were determined by a Nano Photometer spectrophotometer (IMPLEN, CA, USA), a Qubit RNA Assay Kit in a Qubit 2.0 Fluorometer (Life Technologies, CA, USA) and a Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total of 12 μg RNA, 1 μg from each sample, was used as input for transcriptome library construction and 3 μg RNA per sample was used to construct the DGE library.

Library preparation and sequencing

RNA Samples were sent to Novogene Bioinformatics Technology Co. Ltd (Beijing), where the libraries were constructed and sequenced using Illumina HiSeq 2000 platform. Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's protocols and index codes were added to attribute sequences to each sample. Short fragments ranging from 270 bp to 340 bp in length were selected by gel purification and amplified through PCR to create the final sequencing library. Then transcriptome sequencing was carried out on an Illumina HiSeq 2000 platform that generated 100 bp paired-end raw reads, while DGE sequencing generated 100 bp single-end raw reads.

Transcriptome assembly and gene functional annotation

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts where clean data (clean reads) were obtained by filtering out reads containing adapter, reads with unknown base 'N' (where the 'N' ratio was more than 10%), and other low quality reads (where the quality score was lower than 5) from raw data. Meanwhile, Q20 and Q30 (proportion of nucleotides with quality value larger than 20 and 30), and GC-content (proportion of guanine and cytosine nucleotides among total nucleotides) were calculated. All the downstream analyses were based on clean data of high quality. Transcriptome assembly was accomplished by Trinity (Release 2012-10-05) with min_kmer_cov set to 2 by default and all other parameters set to default [29].

The longest transcript of each gene was selected as an unigene, and the function of all assembled unigenes was annotated based on the following databases: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG (euKaryotic Ortholog Groups), Swiss-Prot (A manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), and GO (Gene Ontology). The unigenes were annotated in public NR, NT, Swiss-Prot and KOG databases using NCBI blast 2.2.28+ [30], and the Nr, Nt and Swiss-Prot databases had a cut-off E-value of 10^{-5} , while KOG database had a cut-off E-value of 10^{-3} .

Analysis of DGE tags and bioinformatics

The clean data of DGE were mapped back onto the assembled transcriptome and read count of each gene was obtained from the mapping results by RSEM-1.2.0 [31] for each sample. The bowtie parameter was set at mismatch 2. All read counts were normalized to FPKM (expected

number of fragments per kilobase of transcript per million mapped reads) value, representing gene expression level [32]. To examine the reliability of data between replications, the Pearson's correlation analysis of gene expression among these samples were carried out by the SPSS software.

Differential expression analysis of the samples with three biological replications (two replications for S1) was performed using the DESeq 1.10.1 via the negative binomial distribution. The input values were based on the read counts. The obtained P values were adjusted using the Benjamini and Hochberg's approach to control false discovery rate [33]. Genes with an adjusted P value < 0.05 calculated by DESeq were regarded differentially expressed [34].

GO and KEGG pathway enrichment analysis

GO term enrichment analysis of differentially expressed genes (DEGs) was performed using Goseq 1.10.0 based on the wallenius non-central hyper-geometric distribution which could adjust for gene length bias in DEGs [35]. The GO term with P value < 0.05 was defined as significantly enriched GO term. KEGG (Kyoto encyclopedia of genes and genomes) pathway enrichment analysis was performed based on a FDR cut-off value of 0.05 using KOBAS (version 2.0) after the unigenes were mapped to KEGG pathways [36].

Quantitative real-time PCR analysis

Quantitative real-time PCR (qRT-PCR) analysis was used to verify the expression levels of genes identified in DGE sequencing. The RNA samples used for qRT-PCR assays were the same as used for the DGE experiments. Reverse cDNA for each sample was generated via the PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa Biotechnology, Dalian, China). Real-time PCR was performed with specific primers that were designed based on the selected unigene sequences with Primer 5.0 software. Housekeeping gene β -actin was used as the control gene. All primers are listed in S1 Table. The qRT-PCR was carried out using a SYBR Primix Ex Taq kit (TaKaRa, Dalian, China) following manufacturer's instructions and was analyzed in the ABI 7500 Real-Time System (Applied Biosystems, USA). The gene expression levels were calculated by ABI Prism 7500 Sequence Detection System Software (Applied Biosystems, USA). Each reaction contained 2 μ l cDNA template, 10 μ l 2 \times SYBR Green Master Mix, 0.4 μ l RT reaction mixture, 0.8 μ l forward and 0.8 μ l reverse primer (10 μ mol/ μ l) and water to a final volume of 20 μ l. The PCR amplification was carried out in a 96-well plate with the following cycling parameters: heating for 2 min at 95°C, 40 cycles of denaturation at 95°C for 10 s, annealing for 20 s at 60°C, and extension at 72°C for 35 s. Real-time quantitative PCR was performed in four replications for each sample and data were shown as mean values \pm SD (n = 4). Analysis of the relative gene expression data was conducted using the $2^{-\Delta\Delta C_t}$ method.

Result

Morphological analysis

T. erecta has a typical terminal capitulum consisting ray florets in the periphery and disk florets in the center (Fig 1). The ray florets have three whorl floral organs (sepal, petal and pistil), while the disk florets have four whorl floral organs (sepal, petal, stamen and pistil) (Fig 2). Based on the observation of the flower organs, we found that the petals of the ray and disc florets of MS plant developed into sepal-like structures, while the stamens developed into yellow filaments with no pollen formed (Fig 2). Scanning electron microscopy revealed that the deformed petal of MS plant was covered by unusual pappus hairs which were typically found in sepal, not in petal, and the distorted stamen was covered by trichomes that were only seen in stigma walls (Fig 3). From the observation of transverse semi-thin sections [11], we found that

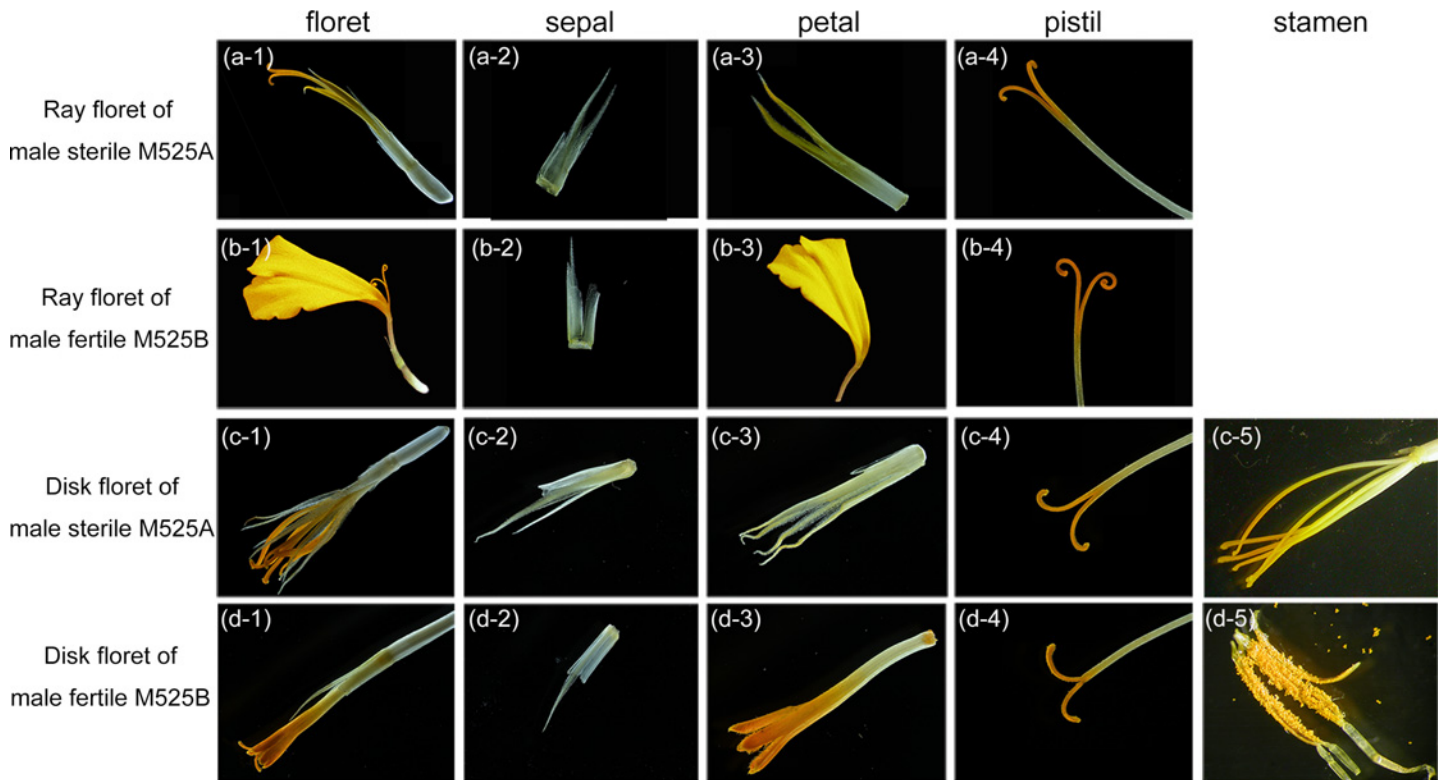


Fig 2. Floret morphology of the male sterile and male fertile two-type line M525AB of *T. erecta*. The ray florets of male sterile M525A (a-1) and male fertile M525B (b-1) had three whorls of floral organs, sepal (a-2, b-2), petal (a-3, b-3) and pistil (a-4, b-4), while the petal of ray floret in M525A developed into sepal-like structure. The disk florets of male sterile M525A (c-1) and male fertile M525B (d-1) had four whorls of floral organs, sepal (c-2, d-2), petal (c-3, d-3), stamen (c-5, d-5) and pistil (c-4, d-4). The petals of disc florets in M525A developed into sepal-like structures, while the stamens developed into yellow filaments.

doi:10.1371/journal.pone.0150892.g002

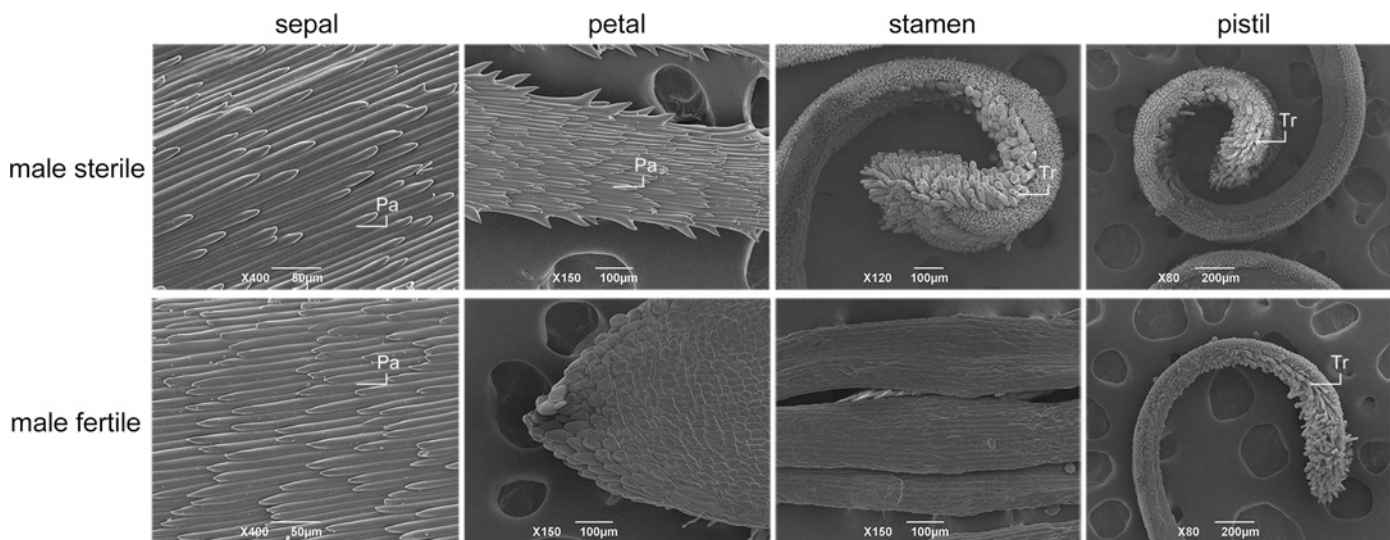


Fig 3. Scanning electron microscope observation of floret morphology of the male sterile and male fertile two-type line M525AB of *T. erecta*. The deformed petal of male sterile plant was covered by unusual pappus hairs which were typically found in sepal. The distorted stamen of male sterile plant was covered by trichomes that were only found in stigma walls. Pa: pappus hairs, Tr: trichomes.

doi:10.1371/journal.pone.0150892.g003

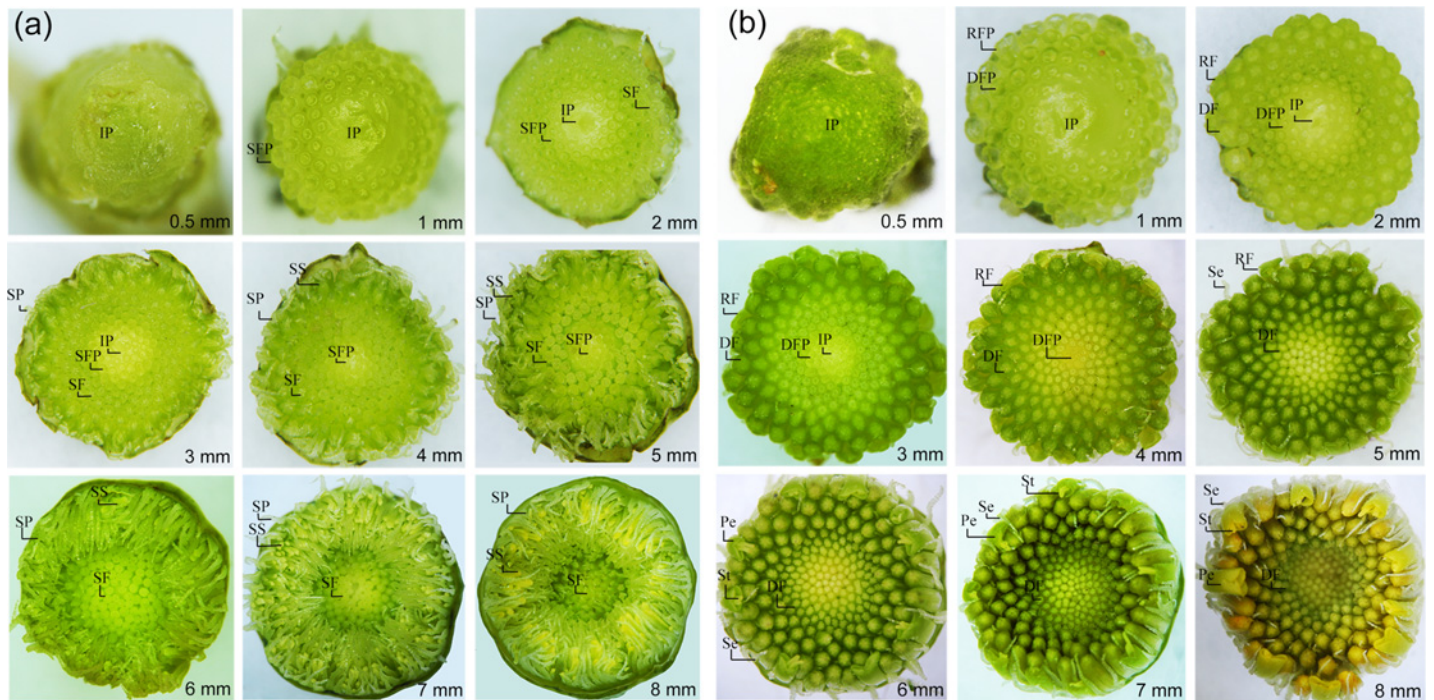


Fig 4. Flower buds development processes of male sterile and male fertile two-type line M525AB of *T. erecta*. (a) Developmental process of male sterile M525A's flower buds from 0.5 mm to 8 mm in diameter; (b) Developmental process of male fertile M525B's flower buds from 0.5 mm to 8 mm in diameter. IP: inflorescence primordium, SFP: sterile floret primordium, SF: sterile floret, SP: sepal and sepal-like petal of male sterile floret, SS: style-like stamen and stigma of male sterile floret, RFP: ray floret primordium, DFP: disc floret primordium, RF: ray floret, DF: disc floret, Se: sepal of fertile floret, Pe: petal of ray floret, St: stigma of ray floret.

doi:10.1371/journal.pone.0150892.g004

the development of the stamen primordia in MS plants failed to differentiate into archesporial cells, sporogenous cells, microspore mother cells, microspore tetrads and pollen grains and the stamens were partially converted to style-like structures. Thus, it is confirmed that male sterility in *T. erecta* was due to the inability to form normal archesporial cells and homeotic conversion of floral organs had occurred when the MS floret organs began to differentiate.

We also observed the developmental process of the flower bud under a stereo microscope. The results showed that only 3–5 rounds of florets in the peripheral of the inflorescence were in the stage of differentiation when the flower bud grew to 1 mm in diameter. When the flower bud grew to 4 mm in diameter, the florets in the center began to differentiate and the florets in the peripheral had completed differentiation with their height reaching 3.81 ± 0.03 mm at the outermost (Fig 4). He et al [11] reported that the floret organs completed differentiation process when the height of the floret reached about 4 mm. The homeotic conversion of floral organs took place when the MS floret organs began to differentiate. Based on our observation and former reports, we focused on the differentiation process of floret organs between MS and MF plant, and therefore chose flower buds 1 mm and 4 mm in diameter for transcriptome and digital gene expression (DGE) analysis.

Generating a reference transcriptome of flower development by Illumina sequencing

To generate a reference transcriptome, RNA was extracted from flower buds (1 mm and 4 mm in diameter) from ten homozygous MS plants and ten homozygous MF plants, and then pooled together for Illumina sequencing. A total of 90,547,072 raw tags were sequenced in the library

of *T. erecta*. After filtering out reads containing adapter, poly-N and other low quality reads from raw data, 87,473,431 clean tags remained in the library. The base average error rate was 0.03%, and the average Q20 and Q30 values were 97.08% and 90.85%, respectively. In addition, the average GC content was 41.82%. These data showed that the Illumina sequencing was of high quality. There were 128,937 transcripts of clean data assembled using Trinity software [23], and all further analyses were based on these transcripts. The average length of transcript was 1,188 bp, ranging from 201 bp to 13,680 bp. N50 and N90 (Put the splicing transcripts in the order of length. Those cumulative lengths more than 50% or 90% of the length of total splicing transcript are called N50 or N90) were 1,928 bp and 523 bp, respectively. There were 24,158 transcripts longer than 2 kbp (S2 Table). From these transcripts, 65,857 unigenes were identified with an average length of 777 bp, with the longest unigene 13,680 bp, and the shortest 201 bp (N50 was 1,379 bp, and N90 was 296 bp). A total of 5,734 unigenes were longer than 2 kbp (S2 Table).

Based on the annotation results shown in Table 1, there were 28,216 unigenes (42.84%) annotated in NR, 14,893 unigenes (22.61%) annotated in Nt, 8,714 unigenes (13.23%) annotated in KO, 21,085 unigenes (32.01%) annotated in SwissProt; 20,711 unigenes (31.44%) annotated in PFAM; 23,079 unigenes (35.04%) annotated in GO; 10,646 unigenes (16.16%) annotated in KOG. In summary, there were 3,481 unigenes (5.28%) annotated in all databases, and 34,176 unigenes (51.89%) annotated in at least one database.

A total of 10,646 unigenes were annotated in KOG, and these unigenes were categorized into 26 groups of KOG function clusters, among which the ‘general function prediction only’ cluster had the highest number of unigenes (1,904, 15.94%), and the ‘Posttranslational modification, protein turnover, chaperones’ cluster had the second largest number of unigenes (1,392, 11.65%), followed by the ‘signal transduction mechanisms’ cluster (1,011, 8.46%). By contrast, only four unigenes were classified into ‘cell motility’ (Fig 5).

Gene Ontology (GO) is an international standardized gene functional classification system that describes properties of genes and their products in any organism. A total of 23,079 unigenes annotated in GO could be categorized into three major categories (cellular component, molecular function and biological process) and 55 subcategories. In the biological process category, the ‘cellular process’ (14,063 unigenes) and the ‘metabolic process’ (13,241 unigenes) were the dominant subcategories. In respect of molecular functions, the major subcategories were ‘binding’ (13,532 unigenes) and ‘catalytic activity’ (11,492 unigenes). In the cellular component category, the ‘cell’ (8,748 unigenes) and ‘cell part’ (8,725 unigenes) were the largest subcategories (Fig 6).

Table 1. Number of unigenes annotated in databases.

Databases	Number of Unigenes	Percentage (%)
Annotated in Nr	28,216	42.84
Annotated in Nt	14,893	22.61
Annotated in KO	8,714	13.23
Annotated in SwissProt	21,085	32.01
Annotated in Pfam	20,711	31.44
Annotated in GO	23,079	35.04
Annotated in KOG	10,646	16.16
Annotated in all Databases	3,481	5.28
Annotated in at least one Database	34,176	51.89
Total Unigenes	65,857	100.00

doi:10.1371/journal.pone.0150892.t001

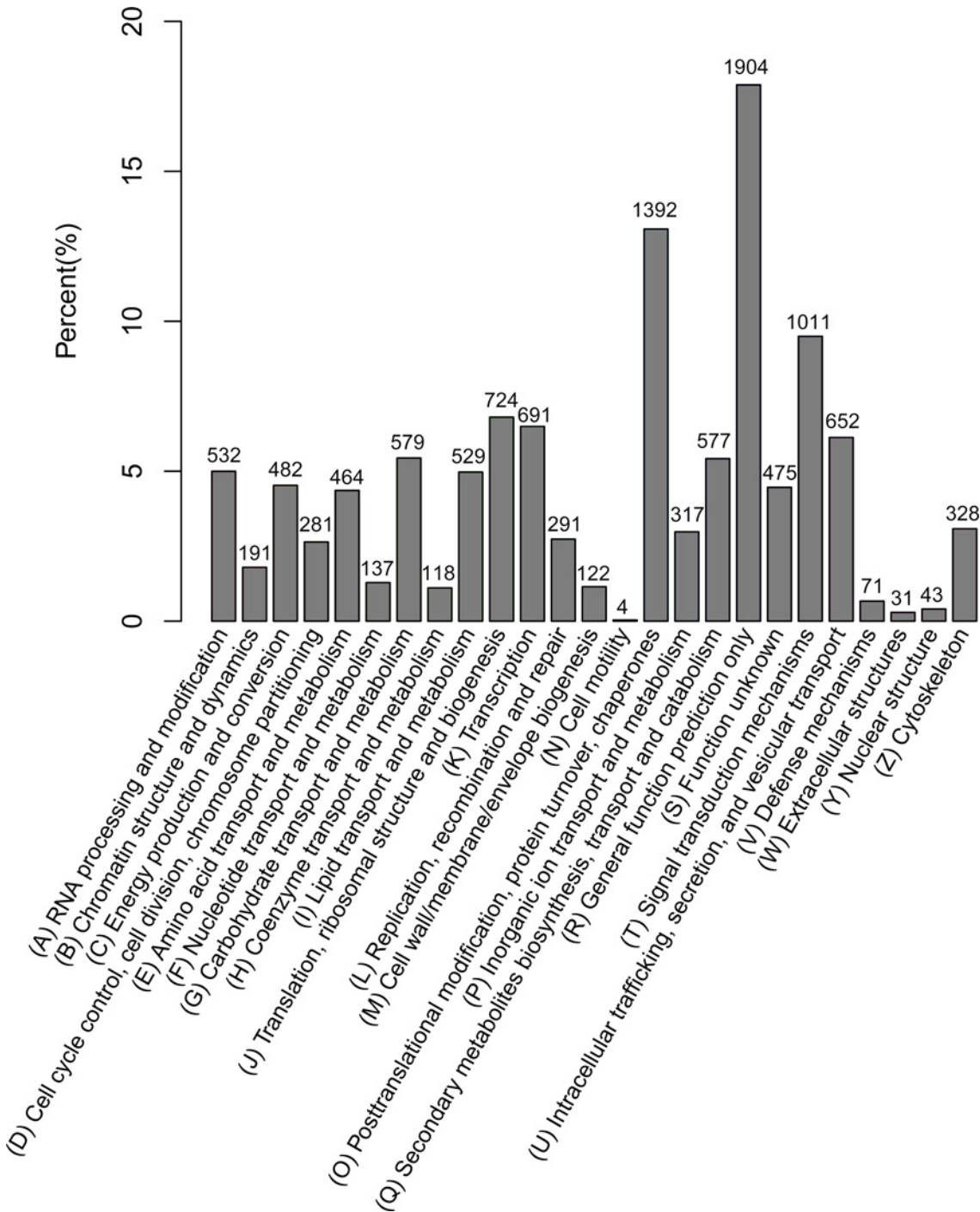


Fig 5. Functional classifications of the assembled unigenes according to the euKaryotic Ortholog Group categories. The x-axis indicated 26 groups of KOG. The y-axis indicated the percentage of the number of annotated genes under a group to the total number of annotated genes.

doi:10.1371/journal.pone.0150892.g005

KEGG pathway has been used to describe the cellular biological molecules that are involved in the metabolic pathways of network diagram, including metabolism pathways, genetic information processing pathways, environmental information processing pathways, cellular process pathways, organismal systems pathways and human diseases pathways. All the human diseases

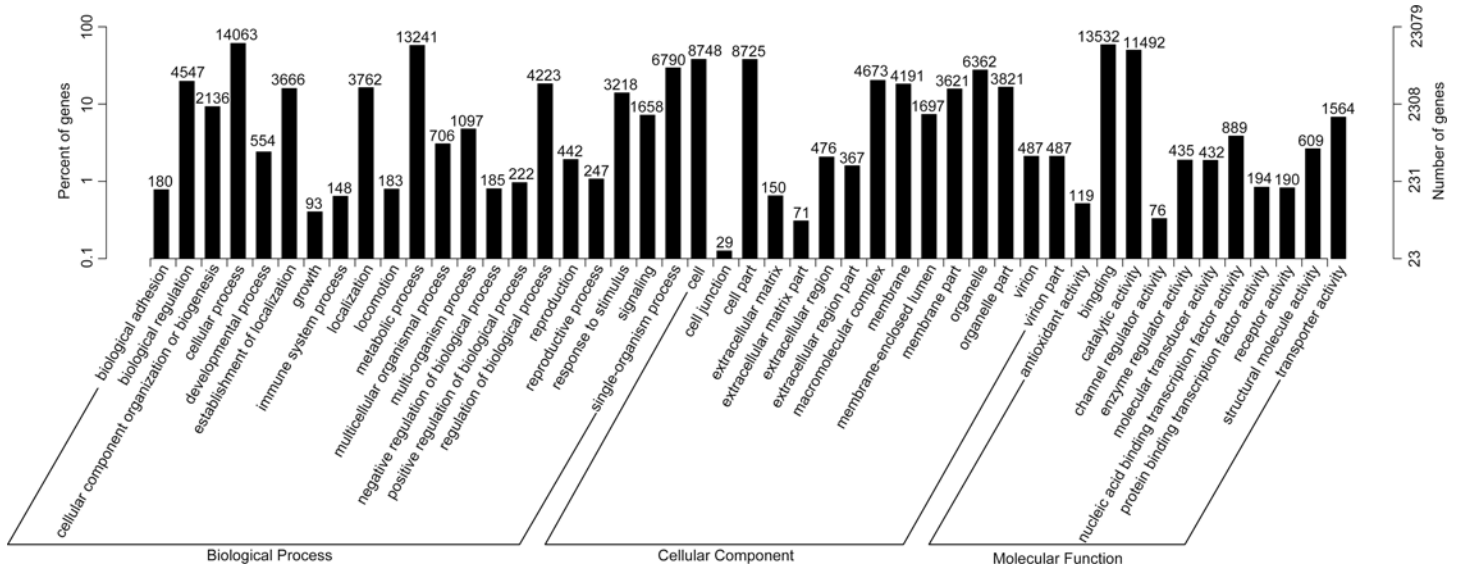


Fig 6. Gene Ontology classifications of the assembled unigenes. The results were categorized into three major categories: cellular component, molecular function, and biological process. The right y-axis indicated the number of genes in a category. The left y-axis indicated the percentage of a specific category of genes in that main category.

doi:10.1371/journal.pone.0150892.g006

pathways were removed in this study. By using KO annotations, we classified the genes into 32 groups based on their participation in KEGG metabolic pathways (Fig 7). In this study, the enriched pathways were ‘metabolism pathways’ (4,475 unigenes), followed by ‘genetic information processing pathways’ (1,939 unigenes).

We predicted the protein coding sequence (CDS) and the amino acid sequence of all unigenes using NCBI blast 2.2.28+ and Estscan (3.0.3) software to analyze unigene functions at the protein level. Firstly, the unigenes were searched in the Nr database and Swissprot database, and the corresponding ORF sequence of the unigenes were used to extract the predicted CDS sequence and translated into amino acid sequence with a standard genetic codon table (5' to 3'). The Nr database takes precedence over the Swissprot database. If the unigene did not hit any database, the software Estscan (3.0.3) was employed to predict its ORF which was then converted to CDS sequence and amino acid sequence. Altogether, a total of 29,054 unigenes (about 44.1%) were functionally annotated in the NR and Swissprot databases using NCBI blast 2.2.28+, and 17,554 not-hit unigenes (26.7%) were predicted by the Estscan (3.0.3) software. The length distributions of the predicted CDS sequences and amino acid sequences were displayed in Fig 8. In general, the length distribution of CDS prediction and translation were consistent with unigene assembly results.

Global analysis of differential gene expression during flower development

To obtain digital gene expression signatures during flower development of the MS and MF plant, we sequenced eleven libraries with three/two replications for flower buds 1 mm and 4 mm in diameter (designated as S1 and S2, F1 and F2). In total, raw reads generated from DGE libraries ranged from 16,249,267 to 21,996,609. After removal of adapter, poly-N and low quality reads, a total of 16,101,543 to 21,795,753 clean reads remained (S3 Table). These trimmed reads were mapped to the reference transcriptome database using RSEM software [25], and the results showed that the total mapped reads ranged from 15,269,622 (94.78%) to 20,699,094 (95.08%) (S3 Table).

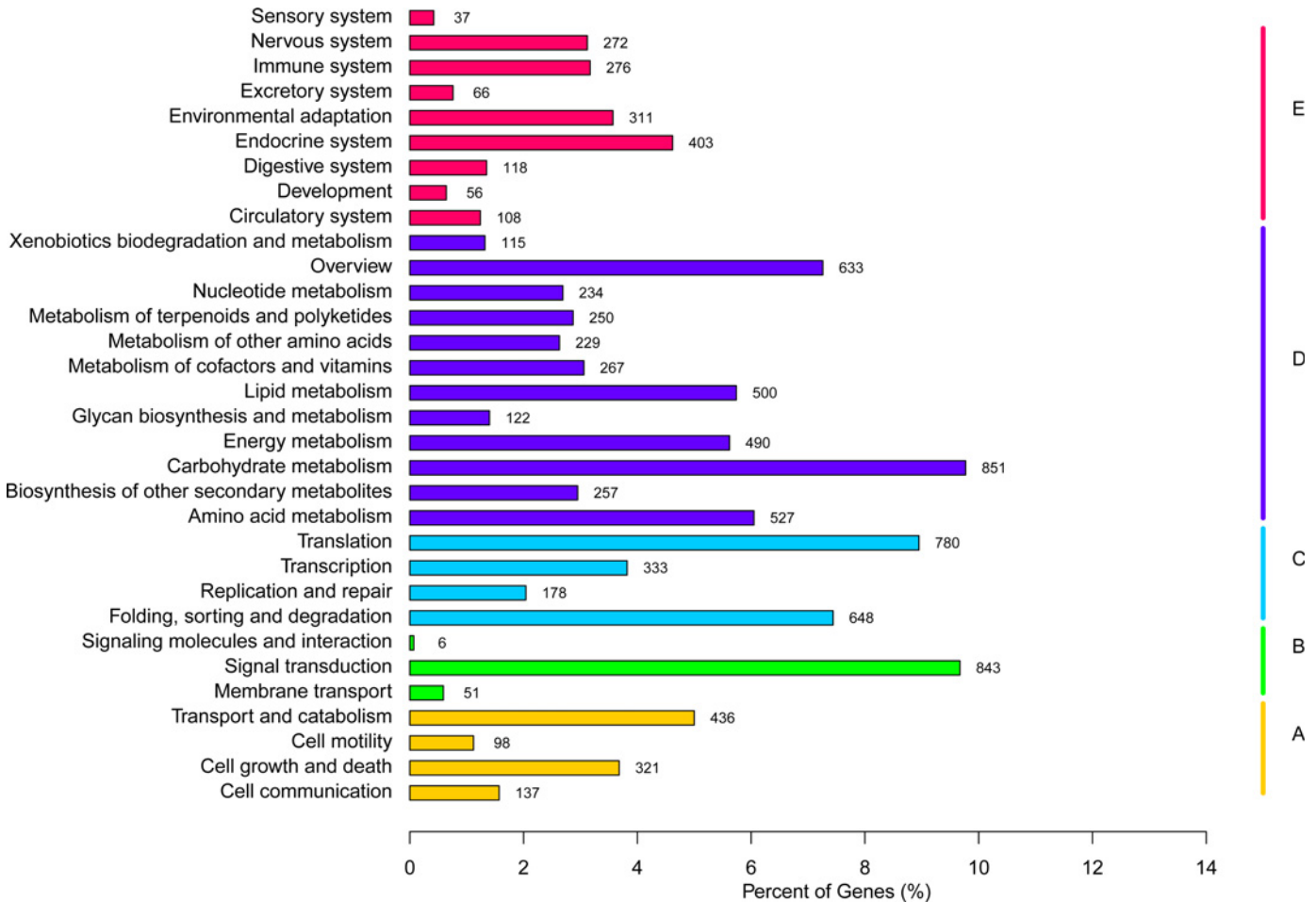


Fig 7. Functional classification of KEGG pathway of assembled unigenes. The KEGG pathways were summarized in five main categories: A, Cellular Processes; B, Environmental Information Processing; C, Genetic Information Processing; D, Metabolism; E, Organismal Systems. The y-axis indicated the name of the KEGG metabolic pathways. The x-axis indicated the percentage of the number of genes annotated under that pathway in the total number of annotated genes.

doi:10.1371/journal.pone.0150892.g007

Gene expression levels were quantified by RSEM [25] for each sample, and all read counts were normalized to FPKM value. To examine the reliability of data between biological replications, the Pearson’s correlation analysis of gene expression were carried out by SPSS software with transformation of $\log_{10}(\text{FPKM}+1)$. The Pearson’s correlation coefficients among replications of each sample were all higher than 0.95, indicating satisfactory repeatability (S1 Fig).

Genes having an adjusted P value < 0.05 found by DESeq were regarded as DEGs. By comparing with F1, 557 transcripts were found to be differentially expressed in the S1 library, which included 142 up-regulated genes and 415 down-regulated genes (Fig 9A). For S2, there were 785 differentially expressed transcripts when compared with F2 library, including 412 up-regulated genes and 373 down-regulated genes (Fig 9B). In addition, 125 transcripts showed significant differential expression levels in both developmental stages (Fig 9C).

To reveal significantly enriched GO terms in DEGs, GO enrichment analysis of functional significance on all DEGs was performed; besides, we also divided these terms into up-regulated and down-regulated groups. The GO term with P value < 0.05 was considered significantly enriched. For the DEGs between S1 and F1, there were two significantly enriched GO terms: oxidoreductase activity acting on paired donors with oxidation of a pair of donors resulting in

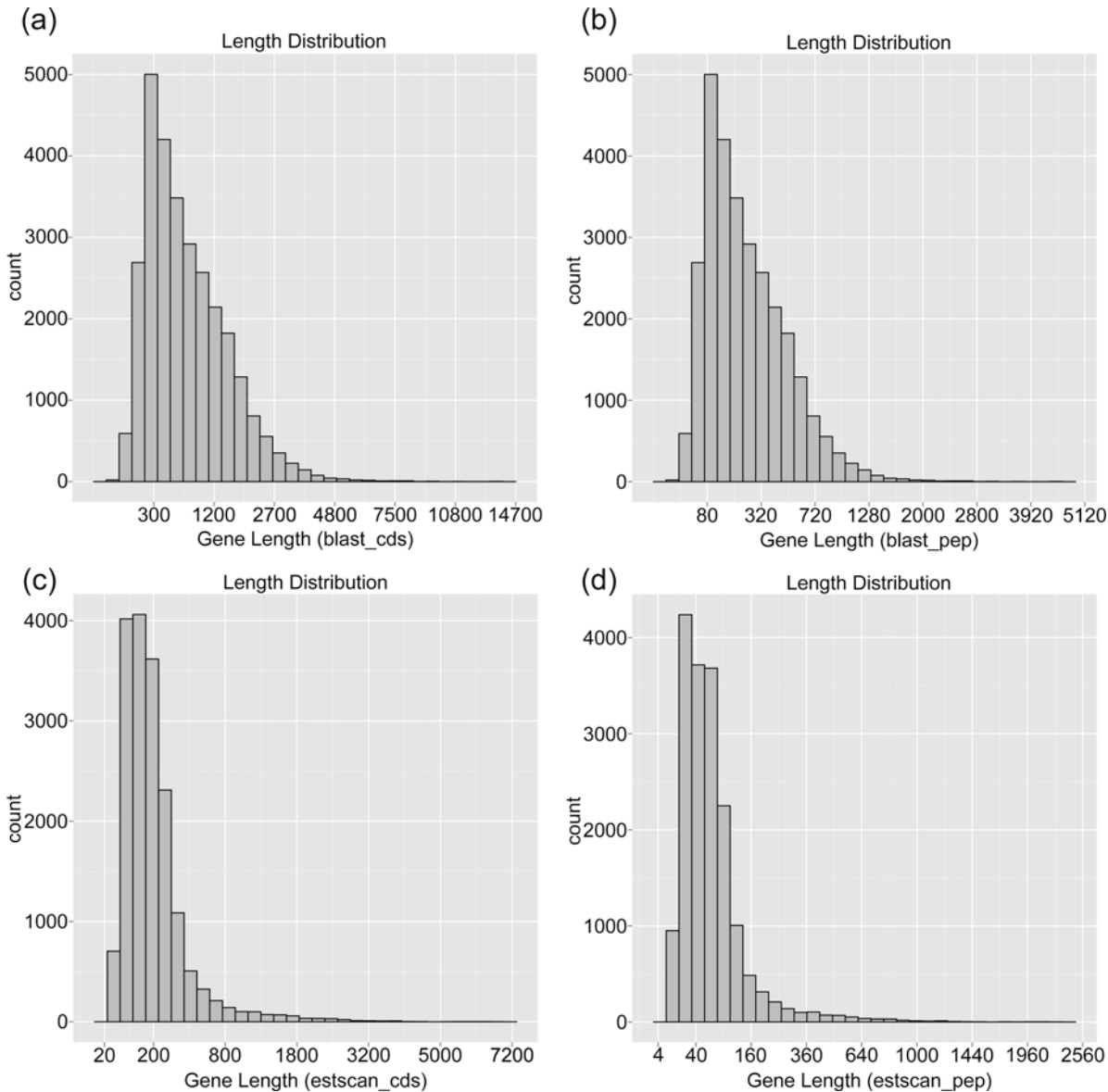


Fig 8. Length distribution of CDS prediction and translation. (a) The length distribution of the predicated CDS sequences using NCBI blast 2.2.28+; (b) The length distribution of the predicated amino acid sequences using NCBI blast 2.2.28+; (c) The length distribution of the predicated CDS sequences using Estscan (3.0.3) software; (d) The length distribution of the predicated amino acid sequences using Estscan (3.0.3) software.

doi:10.1371/journal.pone.0150892.g008

the reduction of molecular oxygen to two molecules of water (15 genes); and oxidoreductase activity acting on paired donors with incorporation or reduction of molecular oxygen (30 genes). Both GO terms participated in molecular function and most of the DEGs were down-regulated in S1 (Fig 10A). Other significantly down-regulated DEGs were presented in the “lipid metabolic process”, belonging to the biological process (S2 Fig).

Comparing the DEGs between S2 and F2, there were 18 significantly enriched GO terms, including 12 in biological process, 5 in molecular function, and 1 in cellular component. The significantly overrepresented GO terms were ‘carbohydrate metabolic process’ (70 genes), ‘transcription factor complex’ (48 genes), ‘cellular carbohydrate metabolic process’ (44 genes), ‘nucleic acid binding transcription factor activity’ (40 genes), ‘sequence-specific DNA binding

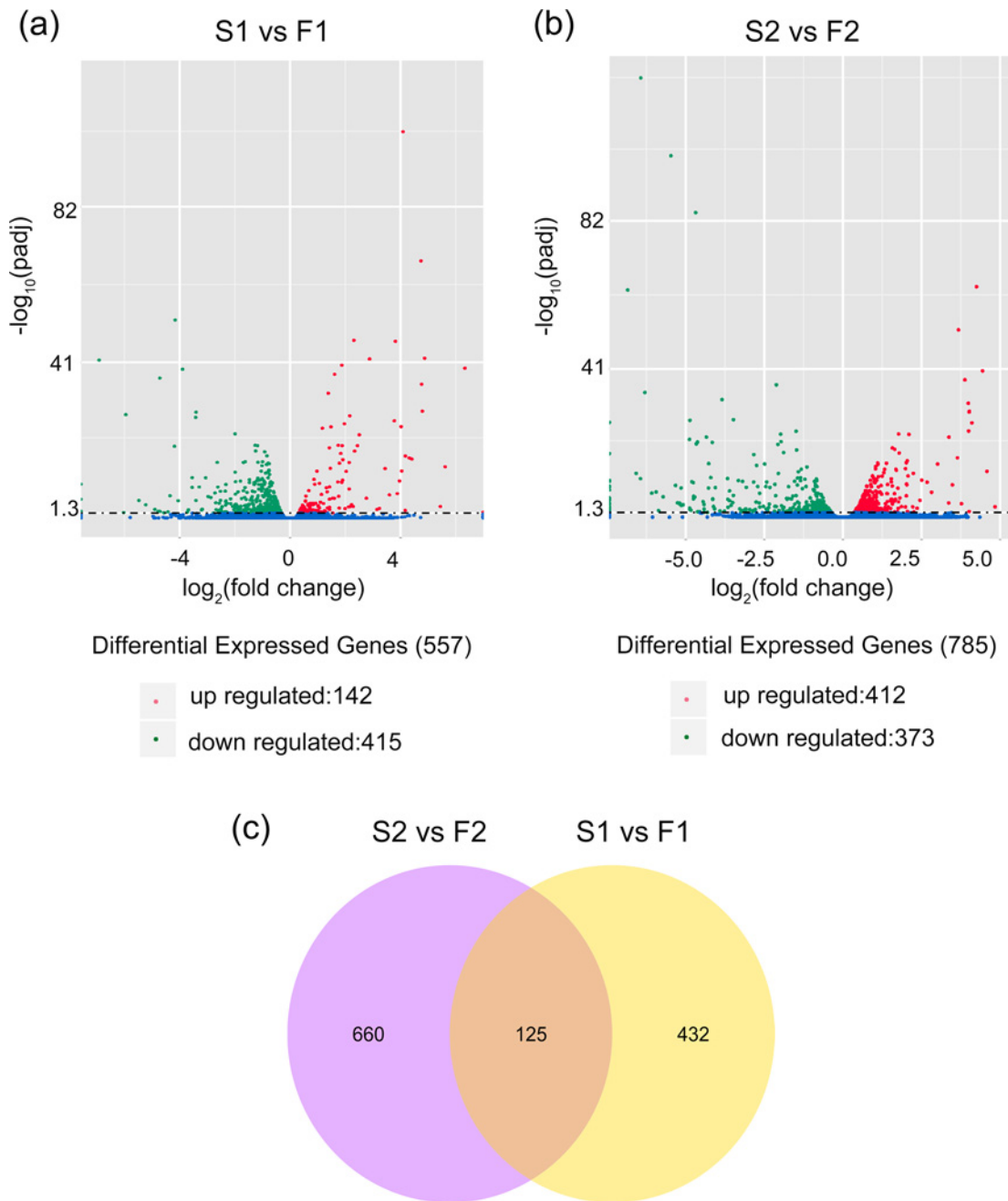


Fig 9. Differentially expressed genes (DEGs) of flower buds from male sterile and male fertile plants. (a) DEGs between S1 (1 mm flower buds of male sterile plants) and F1 (1 mm flower buds of male fertile plants); (b) DEGs between S2 (4 mm flower buds of male sterile plants) and F2 (4 mm flower buds of male fertile plants); (c) The Venn diagram showed specifically or commonly expressed DEGs in both development of flower buds. In the volcano figure, scattered dot represented each gene, blue dots indicated that the unigenes with no significant differential expression level, red dots indicated the significantly up-regulated unigenes while the green dots indicated the significantly down-regulated unigenes. In the Venn diagram, the number in the large circle represented total number of specifically expressed DEGs in 1 mm or 4 mm sized flower buds, while the number in the overlapping portion represented commonly expressed DEGs in both 1 mm and 4 mm sized flower buds.

doi:10.1371/journal.pone.0150892.g009

transcription factor activity' (40 genes), 'cellular polysaccharide metabolic process' (39 genes), and 'polysaccharide metabolic process' (39 genes) (Fig 10B). The down-regulated DEGs in S2 were involved in 10 significantly enriched GO terms. The significantly overrepresented GO

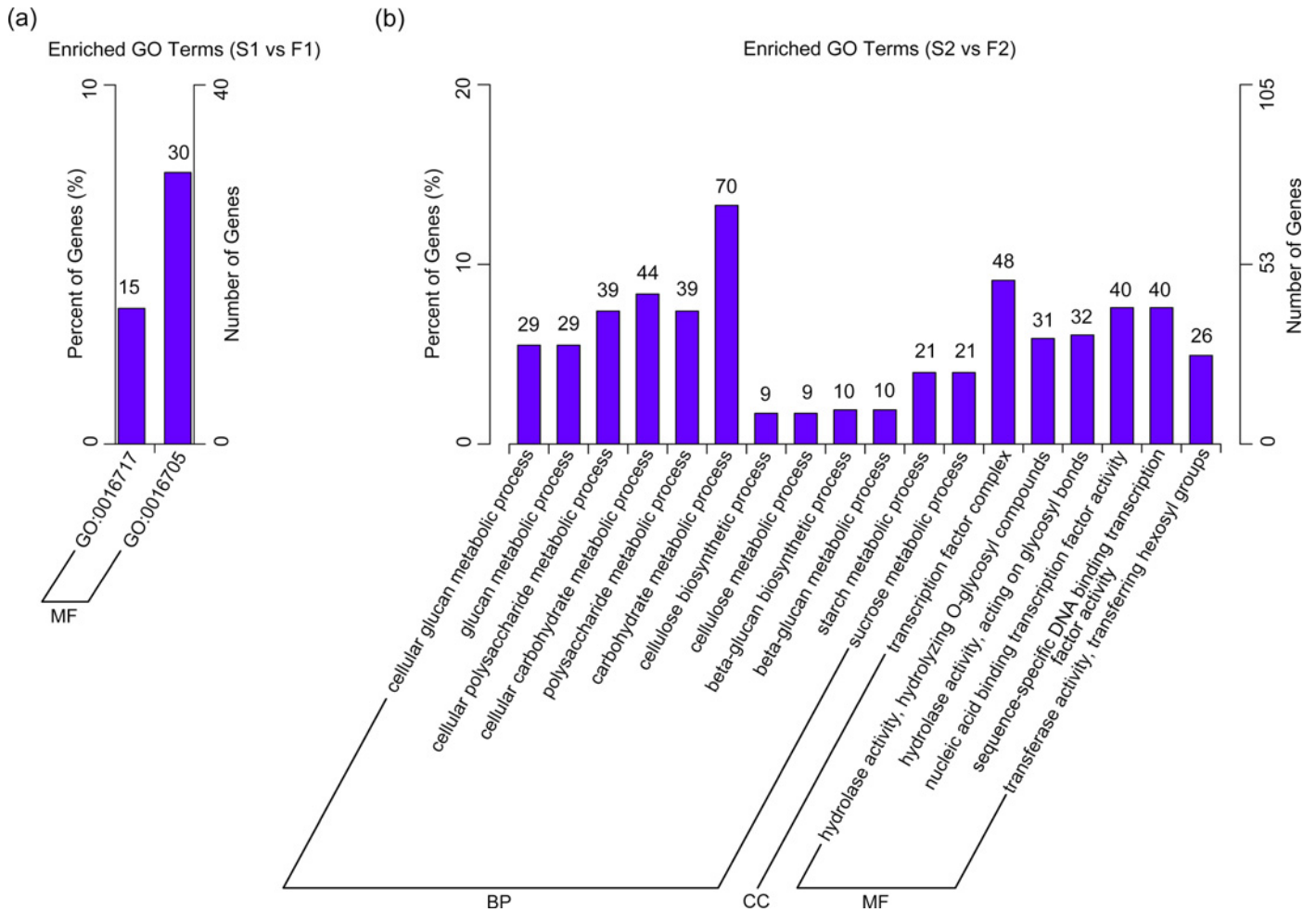


Fig 10. GO term enrichment analysis of differentially expressed genes of flower buds between male sterile and male fertile plants. (a) Enriched GO term between S1 (1 mm flower buds of male sterile plants) and F1 (1 mm flower buds of male fertile plants); (b) Enriched GO term between S2 (4 mm flower buds of male sterile plants) and F2 (4 mm flower buds of male fertile plants). The results were categorized into three major categories (BP: biological process, CC: cellular component, MF: molecular function). The left y-axis represented the percentage of DEGs annotated in this term. The digits above the GO terms represented the number of DEGs annotated in this term (including the sub-term).

doi:10.1371/journal.pone.0150892.g010

terms were “organelle lumen”, “intracellular organelle lumen” and “membrane-enclosed lumen”, all of which belonged to the cellular component category (S3 Fig). The up-regulated DEGs in S2 were showed in 24 significantly enriched GO terms, including 16 in biological process, 2 in cellular component, and 6 in molecular function. The major up-regulated DEGs were seen in “carbohydrate metabolic process”, “cellular carbohydrate metabolic process”, “cellular polysaccharide metabolic process” and “polysaccharide metabolic process” classifications (S4 Fig).

We also conducted KEGG pathway enrichment analysis of DGE to further understand the biological functions of DEGs. The KEGG pathway with corrected P value < 0.05 was considered significantly enriched. The top 20 enriched KEGG pathways corresponding to DEGs detected in both development stages of MS and MF plants were listed in S4 and S5 Tables, respectively. We also conducted the KEGG pathway enrichment analysis of the up-regulated and down-regulated DEGs groups, separately (S6–S9 Tables). For the DEGs between S1 and F1, there were six significantly enriched pathways, and the most significantly over-represented

enriched pathways were ‘biosynthesis of unsaturated fatty acids’ (rich factor = 0.3030, P value = 0, 20 genes) and ‘fatty acid metabolism’ (rich factor = 0.1429, P value = 1.94×10^{-9} , 20 genes) (S4 Table), both of which involved 20 down-regulated DEGs (S6 Table). Other significantly down-regulated DEGs were involved in ‘photosynthesis—antenna proteins’, ‘metabolism of xenobiotics by cytochrome P450’, ‘Drug metabolism—cytochrome P450’, and ‘flavone and flavonol biosynthesis’ pathways (S6 Table). Up-regulated DEGs were mainly found in the ‘Arginine and proline metabolism’ pathway (S7 Table).

Comparing the DEGs between S2 and F2, there were four significantly enriched pathways. The most highly enriched pathway was ‘phenylpropanoid biosynthesis’ (rich factor = 0.0915, P value = 0.0030, 13 genes) for containing most up-regulated DEGs. “Biosynthesis of secondary metabolites” involved the largest number of DEGs (rich factor = 0.0384, P value = 0.0277, 43 genes). Other two significantly enriched pathways were ‘flavonoid biosynthesis’ (rich factor = 0.1304, P value = 0.0304, six genes), and ‘phenylalanine metabolism’ (rich factor = 0.0860, P value = 0.0348, eight genes) (S5 Table), both of which contained both up-regulated and down-regulated DEGs (S8 and S9 Tables).

MADS-box Genes involved in flower development

It has been reported that spontaneous homeotic conversion of floral organs was the underlying cause of the male sterility in this marigold line [11]. So, we specially focused on the MADS-box genes for their regulatory function in floral organs development. The MIKCC-type MADS-box genes involved in plant growth and development, especially in specifying the floral organ identity, have been divided into 13 gene subfamilies, termed *AG*, *AGL6*, *AGL12*, *AGL15*, *AGL17*, *API-FUL*, *BS*, *FLC*, *PI-AP3*, *SEP*, *SVP*, *SOC1* and *TM8* [37–39]. In our study, 31 unigenes were annotated as the MADS-box transcription factors and displayed substantially different expression levels during the flower development (Fig 11). They could be further classified into 10 subfamilies which were *AG*, *AGL15*, *AGL17*, *API-FUL*, *FLC*, *PI-AP3*, *SEP*, *SOC1*, *SVP* and *TM8* (Fig 11).

We looked for the differential expressed genes between S1 and F1, and S2 and F2, respectively. Genes having an adjusted P value < 0.05 found by DESeq were assigned as DEGs. Only one *PI*-like gene (comp62794_c0) showed significantly different expression levels between S1 and F1, and the expression level in S1 was significantly lower than in F1 (adjusted P value = 5.18×10^{-7}). Between S2 and F2, there were 12 MADS-box unigenes showing significantly different expression levels (Table 2). Compared to the expression level in F2, there were 11 unigenes in S2 with significantly lower expression, including one *PI*-like gene (comp62794_c0), four *AP3*-like genes (comp37674_c0, comp37674_c1, comp67037_c0 and comp47648_c0), two *API*-like genes (comp38236_c0 and comp51042_c0), two *AGL15*-like genes (comp42748_c1 and comp53189_c0), one *SEP*-like gene (comp48314_c0), and one *SVP*-like gene (comp45522_c0). By contrast, there was only one *TM8*-like gene (comp46023_c0) expressed higher in S2.

Validation of Illumina sequencing results by qRT-PCR

To confirm the accuracy and reproducibility of the Illumina expression profiles, qRT-PCR analysis was performed to analyze the expression levels of seven MADS-box genes (Fig 12) and 19 randomly selected unigenes. The expression levels of each gene in S1, F1, S2, and F2 were measured through qRT-PCR and compared with its abundance from DGE sequencing data. The relative expression levels of the genes were calculated using the $2^{-\Delta\Delta Ct}$ method in qRT-PCR analysis. The DGE sequencing data were represented by the FPKM value of samples. Linear regression analysis showed significantly positive correlation ($R^2 = 0.885$) between DGE sequencing and qRT-PCR in the fold change of the gene expression ratios (Fig 13), suggesting

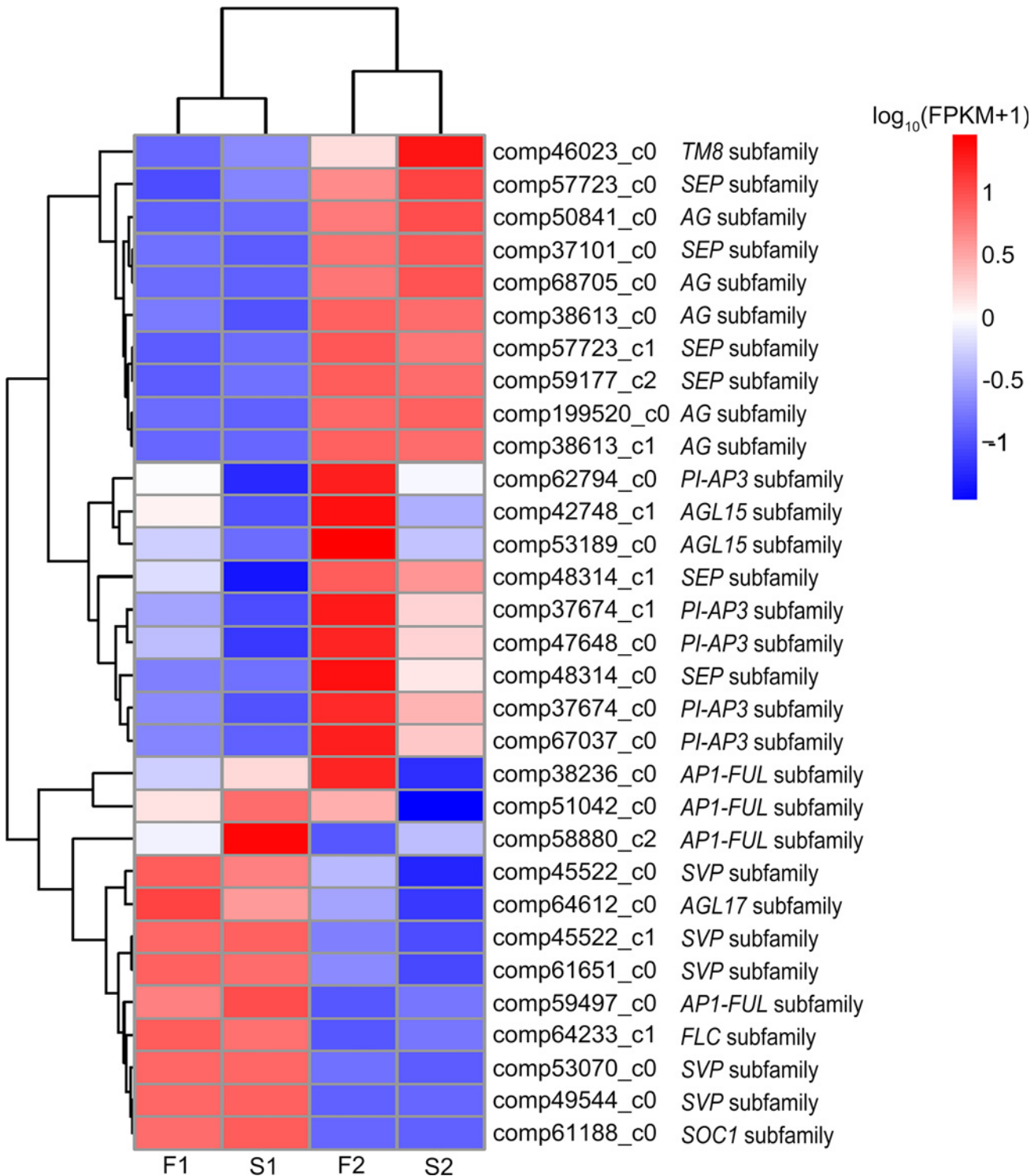


Fig 11. Heat map diagram of expression levels of DEGs annotated in the MADS-box transcription. Data for the relative expression levels of genes were obtained by DGE data after taking log₁₀ (FPKM+1). Color from red to blue, indicated that the log₁₀ (FPKM+1) values were from large to small, red color indicates high expression level and blue color indicates low expression level.

doi:10.1371/journal.pone.0150892.g011

Table 2. MADS-box unigenes showing significantly different expression between S2 and F2.

Gene ID	Homologous genes	log ₂ FoldChange	P value	Adjusted P value
comp62794_c0	<i>PI</i> -like gene	-2.0694	3.79E-24	7.42E-21
comp37674_c1	<i>AP3</i> -like gene	-1.3483	1.42E-14	1.26E-11
comp53189_c0	<i>AGL15</i> -like gene	-3.1943	1.85E-13	1.26E-10
comp47648_c0	<i>AP3</i> -like gene	-1.0424	1.42E-08	4.71E-06
comp48314_c0	<i>SEP</i> -like gene	-0.87983	7.32E-07	0.0001641
comp67037_c0	<i>AP3</i> -like gene	-0.99052	2.84E-06	0.0005529
comp51042_c0	<i>AP1</i> -like gene	-1.1712	6.33E-06	0.0011355
comp45522_c0	<i>SVP</i> -like gene	-3.1072	4.37E-05	0.0058644
comp38236_c0	<i>AP1</i> -like gene	-0.61246	5.62E-05	0.0072351
comp37674_c0	<i>AP3</i> -like gene	-0.96502	0.0001062	0.012092
comp46023_c0	<i>TM8</i> -like gene	1.3663	0.0002641	0.025272
comp42748_c1	<i>AGL15</i> -like gene	-2.1868	0.0004778	0.039343

doi:10.1371/journal.pone.0150892.t002

that the expression of the 26 unigenes revealed by qRT-PCR agreed well with the DGE analysis, thus confirmed the Illumina expression profiles analysis.

Discussion

So far, the lack of genome and transcriptome data has greatly restricted molecular studies in *T. erecta*. Here, we adopted the Illumina sequencing technology for de novo reference transcriptome assembly using flowering buds of *T. erecta*. A total of 87,473,431 clean reads were generated by Illumina HiSeq 2000, and 65,857 unigenes were assembled using the Trinity software, including many transcripts in the floral organ development. Among the Nr, Nt, Pfam, KOG, Swiss-Pro, KO and GO databases, 34,176 unigenes (51.89%) were annotated in at least one database and 3,481 unigenes (5.28%) were annotated in all databases, demonstrating that a large proportion of unigenes have clear descriptions of their functions. Through gene functional annotation, we could not only assess the functions of the unigenes, but get an insight into the putative conserved domains, gene ontology terms, and potential metabolic pathways [40]. This work is the first attempt to sequence and assemble a reference transcriptome in *T. erecta* using Illumina sequencing technology. Our results will provide a valuable resource for future genomic studies on *T. erecta* and other Asteraceae species, especially in flower organ development and/or differentiation. However, there were still nearly half of the unigenes cannot be annotated in any of the seven databases. Similar phenomena were also reported in other Asteraceae plants, such as *Carthamus tinctorius* [40], *Gerbera hybrida* [41], and *Chrysanthemum nankingense* [42]. The reason may lie in the uniqueness of unigenes in Asteraceae family and further studies are needed to understand the biological functions of those non-annotated unigenes.

DGE analysis is a powerful tool to identify and quantify gene expression on the whole genome level. When compared with traditional technologies, such as RDA (representational difference analysis), SSH (suppression subtractive hybridization), cDNA-AFLP (DNA amplified fragment length polymorphism) and RFDD-PCR (restriction fragment differential display PCR), DGE, a sequencing based method, could provide comprehensive sequencing data for studying differentially expressed genes [13]. Recently, transcriptome and DGE techniques have been successfully utilized to study the molecular mechanism of sterility and to identify the candidate regulators or genes responsible for anther and pollen development in many plant species [23–27]. In this study, 1 mm and 4 mm sized flower buds of MS and MF plants of *T. erecta*

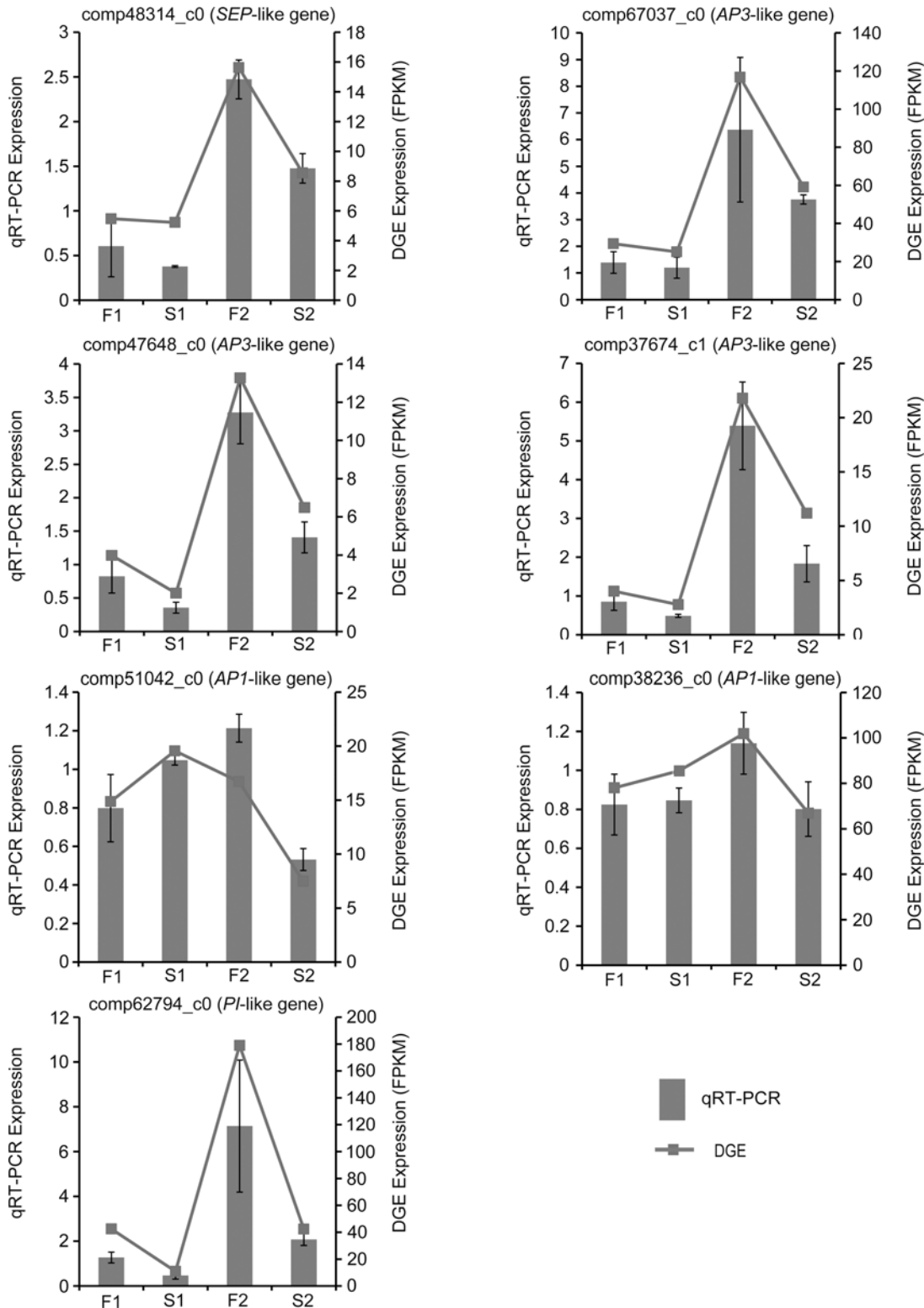


Fig 12. qRT-PCR verifications of seven MADS-box genes. The x axis represented four samples. S1: 1 mm flower buds of male sterile plants, F1: 1 mm flower buds of male fertile plants, S2: 4 mm flower buds of male sterile plants, F2: 4 mm flower buds of male fertile plants. The Left y axis represented the relative expression level by qRT-PCR. The right y axis is the FPKM value by DGE analysis.

doi:10.1371/journal.pone.0150892.g012

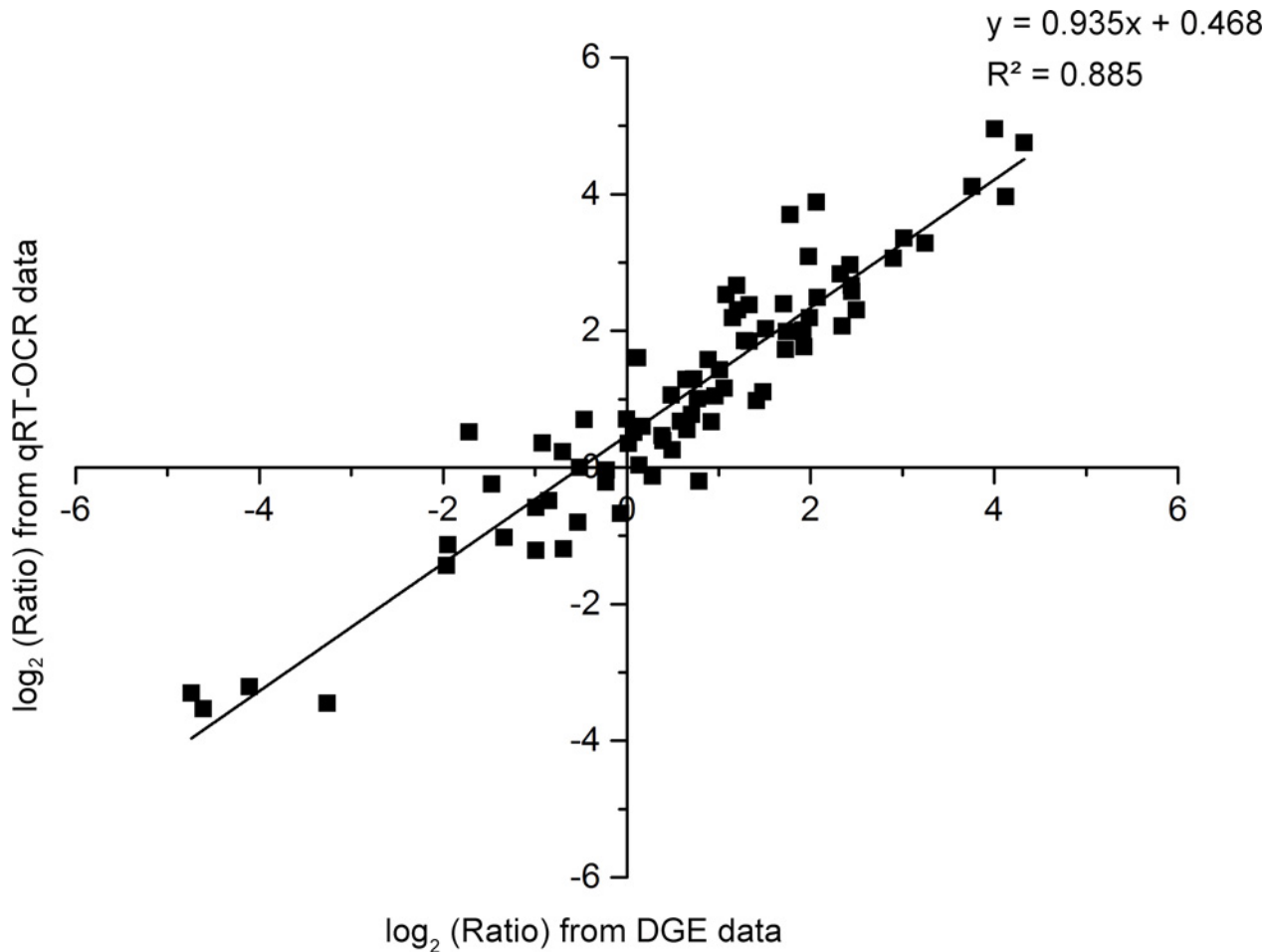


Fig 13. Linear regression analysis of the fold change of the gene expression ratios between DEG sequencing and qRT-PCR. 26 unigenes were selected for quantitative real-time PCR analysis to confirm the accuracy and reproducibility of the Illumina expression profiles using the same RNA samples that were used for DGE sequencing. The relative expression levels of the genes were calculated using the $2^{-\Delta\Delta Ct}$ method in qRT-PCR analysis. The DGE sequencing data were represented by the FPKM value of samples. Scatterplots were generated by the \log_2 expression ratios from DGE sequencing data (x-axis) and qRT-PCR data (y-axis).

doi:10.1371/journal.pone.0150892.g013

were designated for DGE analysis to profile the differences at the transcriptional level and identify candidate genes associated with male sterility. According to the DGE results, we detected 557 transcripts with significantly different expression levels between S1 and F1, and 785 transcripts between S2 and F2. Most of these differentially expressed genes were annotated in the public databases. These annotated genes might be candidates causing male sterility in *T. erecta* and could provide an invaluable resource to identify genes involved in flower development. To further understand the biological functions of DEGs, GO term and KEGG pathway enrichment analysis were employed to analyze the DEGs. These DGE analysis results will provide a better understanding in the molecular mechanism of male sterility in *T. erecta*.

The male sterility of *T. erecta* was not due to the failure of anther or pollen development, but as a result of the male organ abnormality caused by homeotic conversion of floral organs [11]. Most floral organ determined genes have been categorized into the family of MADS-box genes [12, 43–45] and have been further grouped into five different classes (A, B, C, D and E) based on their biological functions [38]. They were considered critical to define the differentiation of four whorl floral organs, and loss-of-function of any class of MADS-box genes may

result in homeotic conversion of floral organs. According to the ABCDE model of flower organ development, the class A genes (*API*, *CAL* and *AP2*) specify the sepal identity in the first whorl; the class A, B (*AP3* and *PI*) and E (*AGL3* and *SEP*) genes collectively control the petal identity in the second whorl; the class B, C (*AG*) and E genes all together control the stamen identity in the third whorl; the class C and E genes combined to determine the formation of the carpel in the fourth whorl; the class D (*SHPI* and *SHP2*, *AGL11* and *AGL13*) and E genes jointly determine the formation of the ovule [46–50].

In our study, only one *PI*-like unigene had significant differential expression levels at the beginning of floret differentiation between S1 and F1, and its expression level in S1 was significantly lower than in F1 (Fig 12). *PI*-like genes belonged to B class genes, and loss-of-function of B class genes produced homeotic phenotypes in which the second whorl organs developed into sepaloid structures, and the third whorl organs developed into carpeloid structures [51, 52]. This conclusion was confirmed by many other researches in the past decades. The co-suppression of *FBP1*, a *PI*-like gene in petunia, resulted in homeotic conversions of petals toward sepals and stamens toward carpels [53]. *MdPI*, identified in apple (*Malus domestica*), not only had a function of floral organ determination but played a role in apple parthenocarpy [54]. In grapevine (*Vitis vinifera*), the mutants showing abnormal petal / stamen structures had low expression level of *VvMADS9*, an orthologue of *PI* gene [55]. In California poppy (*Eschscholzia californica*), the truncation of highly conserved *PI* motif in *SEI-1* protein affected the formation of higher order complexes causing homeotic conversions [56]. Although the regulatory and protein-protein interactions of B-class factors have undergone changes during evolution, they still have conserved functions among flowering plants [12, 37, 57–60]. Thus, it seems likely that the *PI*-like gene might be the promising candidate gene conferring homeotic conversion in *T. erecta*.

Based on the DEGs results, 12 unigenes belonging to MADS-box family showed significantly different expression levels between S2 and F2 which contain florets at various differentiation stages, including the ones in the center that just began differentiation, and the ones on the peripheral that had already completed differentiation (Fig 12, Table 2). Compared to the expression level of F2, 11 unigenes expressed significantly lower in S2, including one *PI*-like gene, four *AP3*-like genes, two *API*-like genes, two *AGL15*-like genes, one *SEP*-like gene, and one *SVP*-like gene. By contrast, there was one *TM8*-like gene expressing higher in S2. The *SVP*-like, *AGL15*-like and *TM8*-like genes were reported to be involved in floral transition and to determine flowering time [61–63]. The *PI*-like, *AP3*-like, *API*-like, and *SEP*-like genes were floral organ identity genes, belonging to the B-, B-, A-, and E-class floral homeotic genes, respectively. Based on the ABCDE model, *AP3* and *PI* proteins are functional partners interacting with each other to form obligate heterodimers for DNA binding in vitro and to regulate gene expression by binding to the *CAR*G motif of their promoters. A complex comprising of *AP3/PI/SEP3/AP1* was postulated to specify petals formation and a complex comprising of *AP3/PI/SEP3/AG* specify stamen development [64]. The decreased expression levels of the A-, B-, and E-class genes might influence the formation and function of the heterodimers or high order complex in MS plants. We hypothesized that male sterility of *T. erecta* might be related to the suppressed expression of *PI*-like gene at the beginning of floret differentiation, which could affect the formation of *PI/AP3* heterodimer and furtherly influence the quaternary complexes of *AP3/PI/SEP3/AP1* and *AP3/PI/SEP3/AG*, leading to the absence of normal petal and stamen organs in MS *T. erecta*.

Supporting Information

S1 Fig. Pearson's correlation analysis of gene expression between samples. F1-1, F1-2, F1-3 and F2-1, F2-2, F2-3 were different replications of F1 (1 mm flower buds of male fertile plants)

and F2 (4 mm flower buds of male fertile plants), respectively. S1-1, S1-2, and S2-1, S2-2, S2-3 were different replications of S1 (1 mm flower buds of male sterile plants) and S2 (4 mm flower buds of male sterile plants), respectively. The number represented the Pearson's correlation analysis of gene expression between samples, the value ranges from 0 to 1. A high value between the biological samples indicated that the samples have good repeatability.

(TIF)

S2 Fig. GO term enrichment analysis of down-regulated DEGs of 1 mm flower buds between male sterile and male fertile plants. BP: biological process, MF: molecular function. The x-axis represents the categories of GO terms, the left y-axis represents the percentage of DEGs annotated in this term, and the digits above the GO terms represent the number of DEGs annotated in this term.

(TIF)

S3 Fig. GO term enrichment analysis of down-regulated DEGs of 4 mm flower buds between male sterile and male fertile plants. CC: cellular component, MF: molecular function. The x-axis represents the categories of GO terms, the left y-axis represents the percentage of DEGs annotated in this term, and the digits above the GO terms represent the number of DEGs annotated in this term.

(TIF)

S4 Fig. GO term enrichment analysis of up-regulated DEGs of 4 mm flower buds between male sterile and male fertile plants. BP: biological process, CC: cellular component, MF: molecular function. The x-axis represents the categories of GO terms, the left y-axis represents the percentage of DEGs annotated in this term, and the digits above the GO terms represent the number of DEGs annotated in this term.

(TIF)

S1 Table. Primers of the selected unigenes for qRT-PCR.

(DOCX)

S2 Table. Length distribution of unigenes and transcripts.

(DOCX)

S3 Table. Summary of the sequencing data quality of the eleven digital gene expression profiles.

(DOCX)

S4 Table. The top 20 enriched KEGG pathways of differentially expressed genes of 1 mm flower buds between male sterile and male fertile plants.

(DOCX)

S5 Table. The top 20 enriched KEGG pathways of differentially expressed genes of 4 mm flower buds between male sterile and male fertile plants.

(DOCX)

S6 Table. The top 20 enriched KEGG pathways of down-regulated DEGs of 1 mm flower buds between male sterile and male fertile plants.

(DOCX)

S7 Table. The top 16 enriched KEGG pathways of up-regulated DEGs of 1 mm flower buds between male sterile and male fertile plants.

(DOCX)

S8 Table. The top 20 enriched KEGG pathways of down-regulated DEGs of 4 mm flower buds between male sterile and male fertile plants.

(DOCX)

S9 Table. The top 20 enriched KEGG pathways of up-regulated DEGs of 4 mm flower buds between male sterile and male fertile plants.

(DOCX)

Acknowledgments

We thank all past and present colleagues in our lab for their constructive suggestions and technical support.

Author Contributions

Conceived and designed the experiments: MZB YHH YA. Performed the experiments: YA QHZ. Analyzed the data: YA QHZ WNW ZC. Contributed reagents/materials/analysis tools: YA QHZ CLZ. Wrote the paper: YA YHH MZB. Plant cultivation: YA CLZ. Revised the paper: YA MZB YHH WNW.

References

1. Liu Z, Cai X, Seiler GJ, Jan CC (2014) Interspecific amphiploid-derived alloplasmic male sterility with defective anthers, narrow disc florets and small ray flowers in sunflower. *Plant Breeding* 133: 742–747. doi: [10.1111/pbr.12216](https://doi.org/10.1111/pbr.12216)
2. Ai Y, Zhang Q, Pan C, Zhang H, Ma S, He Y, et al. (2015) A study of heterosis, combining ability and heritability between two male sterile lines and ten inbred lines of *Tagetes patula*. *Euphytica* 203: 349–366. doi: [10.1007/s10681-014-1278-y](https://doi.org/10.1007/s10681-014-1278-y)
3. Vasudevan P, Kashyap S, Sharma S (1997) *Tagetes*: a multipurpose plant. *Bioresource Technology* 62: 29–33. doi: [10.1016/S0960-8524\(97\)00101-6](https://doi.org/10.1016/S0960-8524(97)00101-6)
4. Siriamompuna S, Kaisoona O, Meesoc N (2012) Changes in colour, antioxidant activities and carotenoids (lycopene, β -carotene, lutein) of marigold flower (*Tagetes erecta* L.) resulting from different drying processes. *Journal of Functional Foods* 4: 757–766. doi: [10.1016/j.jff.2012.05.002](https://doi.org/10.1016/j.jff.2012.05.002)
5. Bhatt BJ (2013) Comparative analysis of larvicidal activity of essential oils of *Cymbopogon flexuosus* (Lemon grass) and *Tagetes erecta* (Marigold) against *Aedes aegypti* larvae. *Euro J Exp Bio* 3: 422–427.
6. Towner JW (1961) The inheritance of Femina, a male-sterile character in *Tagetes erecta*. *Proc Amer Soc Hort Sci California: AAS—Pacific Davis* 2.
7. Singh B, Swarup V (1971) Heterosis and combining ability in African marigold. *Indian J Genet PI Br* 31: 407–415.
8. Sreekala C, Raghava SP (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. doi: [10.1007/s00122-002-1143-6](https://doi.org/10.1007/s00122-002-1143-6) PMID: [12596009](https://pubmed.ncbi.nlm.nih.gov/12596009/)
9. Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. *Bot Rev* 38: 425–454.
10. Budar F, Pelletier G (2001) Male sterility in plants: occurrence, determinism, significance and use. *Life Sci* 324: 543–550. doi: [10.1016/S0764-4469\(01\)01324-5](https://doi.org/10.1016/S0764-4469(01)01324-5)
11. 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 Breeding* 26: 19–29. doi: [10.1007/s11032-009-9372-x](https://doi.org/10.1007/s11032-009-9372-x)
12. Theißen G (2001) Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol* 4: 75–85. doi: [10.1016/S1369-5266\(00\)00139-4](https://doi.org/10.1016/S1369-5266(00)00139-4) PMID: [11163172](https://pubmed.ncbi.nlm.nih.gov/11163172/)
13. Mardis ER (2008) The impact of next generation sequencing technology on the genetics. *Trends Genet* 24: 133–141. doi: [10.1016/j.tig.2007.12.007](https://doi.org/10.1016/j.tig.2007.12.007) PMID: [18262675](https://pubmed.ncbi.nlm.nih.gov/18262675/)
14. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344–1349. doi: [10.1126/science.1158441](https://doi.org/10.1126/science.1158441) PMID: [18451266](https://pubmed.ncbi.nlm.nih.gov/18451266/)

15. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10: 57–63. doi: [10.1038/nrg2484](https://doi.org/10.1038/nrg2484) PMID: [19015660](https://pubmed.ncbi.nlm.nih.gov/19015660/)
16. AC't Hoen P, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, et al. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36: e141. doi: [10.1093/nar/gkn705](https://doi.org/10.1093/nar/gkn705) PMID: [18927111](https://pubmed.ncbi.nlm.nih.gov/18927111/)
17. Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, et al. (2009) Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* 19: 1825–1835. doi: [10.1101/gr.094482.109](https://doi.org/10.1101/gr.094482.109) PMID: [19541910](https://pubmed.ncbi.nlm.nih.gov/19541910/)
18. Morrissy AS, Griffith M, Marra MA (2011) Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res* 21: 1203–1212. doi: [10.1101/gr.113431.110](https://doi.org/10.1101/gr.113431.110) PMID: [21719572](https://pubmed.ncbi.nlm.nih.gov/21719572/)
19. Hong LZ, Li J, Schmidt-Kuntzel A, Warren WC, Barsh GS (2011) Digital gene expression for non-model organisms. *Genome Res* 21: 1905–1915. doi: [10.1101/gr.122135.111](https://doi.org/10.1101/gr.122135.111) PMID: [21844123](https://pubmed.ncbi.nlm.nih.gov/21844123/)
20. Tang Q, Ma XJ, Mo CM, Wilson IW, Song C, Zhao H, et al. (2011) An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics* 12: 343. doi: [10.1186/1471-2164-12-343](https://doi.org/10.1186/1471-2164-12-343) PMID: [21729270](https://pubmed.ncbi.nlm.nih.gov/21729270/)
21. Chen P, Ran S, Li R, Huang Z, Qian J, Yu M, et al. (2014) Transcriptome de novo assembly and differentially expressed genes related to cytoplasmic male sterility in kenaf (*Hibiscus cannabinus* L.). *Mol Breeding* 34: 1879–1891. doi: [10.1007/s11032-014-0146-8](https://doi.org/10.1007/s11032-014-0146-8)
22. Guo Q, Ma X, Wei S, Qiu D, Wilson IW, Wu P, et al. (2014) *De novo* transcriptome sequencing and digital gene expression analysis predict biosynthetic pathway of rhynchophylline and isorhynchophylline from *Uncaria rhynchophylla*, a non-model plant with potent anti-alzheimer's properties. *BMC Genomics* 15: 676. doi: [10.1186/1471-2164-15-676](https://doi.org/10.1186/1471-2164-15-676) PMID: [25112168](https://pubmed.ncbi.nlm.nih.gov/25112168/)
23. Zheng BB, Wu XM, Ge XX, Deng XX, Grosser JW, Guo WW (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. doi: [10.1371/journal.pone.0043758](https://doi.org/10.1371/journal.pone.0043758) PMID: [22952758](https://pubmed.ncbi.nlm.nih.gov/22952758/)
24. Liu C, Ma N, Wang PY, Fu N, 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. doi: [10.1371/journal.pone.0065209](https://doi.org/10.1371/journal.pone.0065209) PMID: [23750245](https://pubmed.ncbi.nlm.nih.gov/23750245/)
25. Qu C, Fu F, Liu M, Zhao H, Liu C, Li J, et al. (2015) Comparative transcriptome analysis of recessive male sterility (RGMS) in sterile and fertile *Brassica napus* lines. *PLoS One* 10: e0144118. doi: [10.1371/journal.pone.0144118](https://doi.org/10.1371/journal.pone.0144118) PMID: [26656530](https://pubmed.ncbi.nlm.nih.gov/26656530/)
26. 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 Genomics* 14: 930–938.
27. Fang W, Zhao F, Sun Y, Xie D, Sun L, Xu Z, et al. (2015) Transcriptomic profiling reveals complex molecular regulation in cotton genic male sterile mutant Yu98-8A. *PLoS One* 10: e0133425. doi: [10.1371/journal.pone.0133425](https://doi.org/10.1371/journal.pone.0133425) PMID: [26382878](https://pubmed.ncbi.nlm.nih.gov/26382878/)
28. Ai Y, He Y, Hu Y, Zhang Q, Pan C, Bao M (2014) Characterization of a novel male sterile mutant of *Tagetes patula* induced by heat shock. *Euphytica* 200: 159–173. doi: [10.1007/s10681-014-1116-2](https://doi.org/10.1007/s10681-014-1116-2)
29. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
30. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402. doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389) PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
31. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323. doi: [10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323) PMID: [21816040](https://pubmed.ncbi.nlm.nih.gov/21816040/)
32. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28: 511–515. doi: [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621) PMID: [20436464](https://pubmed.ncbi.nlm.nih.gov/20436464/)
33. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society* 57:289–300. doi: [10.2307/2346101](https://doi.org/10.2307/2346101)
34. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, 2010, 11:1–12. doi: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106)
35. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11: R14. doi: <http://genomebiology.com/2010/11/2/R14> PMID: [20132535](https://pubmed.ncbi.nlm.nih.gov/20132535/)

36. Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21: 3787–3793. doi: [10.1093/bioinformatics/bti430](https://doi.org/10.1093/bioinformatics/bti430) PMID: [15817693](https://pubmed.ncbi.nlm.nih.gov/15817693/)
37. Alvarez-Buylla ER, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, et al. (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci USA* 97: 5328–5333. doi: [10.1073/pnas.97.10.5328](https://doi.org/10.1073/pnas.97.10.5328) PMID: [10805792](https://pubmed.ncbi.nlm.nih.gov/10805792/)
38. Becker A, Theißen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29: 464–489. doi: [10.1016/S1055-7903\(03\)00207-0](https://doi.org/10.1016/S1055-7903(03)00207-0) PMID: [14615187](https://pubmed.ncbi.nlm.nih.gov/14615187/)
39. Diaz-Riquelme J, Lijavetzky D, Martinez-Zapater JM, Carmona MJ (2009) Genome-wide analysis of MIKC^C-type MADS box genes in grapevine. *Plant Physiol* 149: 354–369. doi: [10.1104/pp.108.131052](https://doi.org/10.1104/pp.108.131052) PMID: [18997115](https://pubmed.ncbi.nlm.nih.gov/18997115/)
40. Huang L, Yang X, Sun P, Tong W, Hu S (2012) The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7: e38653. doi: [10.1371/journal.pone.0038653](https://doi.org/10.1371/journal.pone.0038653) PMID: [22723874](https://pubmed.ncbi.nlm.nih.gov/22723874/)
41. Laitinen RA, Immanen J, Auvinen P, Rudd S, Alatalo E, Paulin L, et al. (2005). Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae). *Genome Res* 15: 475–486. doi: [10.1101/gr.3043705](https://doi.org/10.1101/gr.3043705) PMID: [15781570](https://pubmed.ncbi.nlm.nih.gov/15781570/)
42. Wang H, Jiang J, Chen S, Qi X, Peng H, Li P, et al. (2013) Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. *PLoS One* 8: e62293. doi: [10.1371/journal.pone.0062293](https://doi.org/10.1371/journal.pone.0062293) PMID: [23626799](https://pubmed.ncbi.nlm.nih.gov/23626799/)
43. Pařenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15: 1538–1551. doi: [10.1105/tpc.011544](https://doi.org/10.1105/tpc.011544) PMID: [12837945](https://pubmed.ncbi.nlm.nih.gov/12837945/)
44. Messenguy F, Dubois E (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene* 316: 1–21. doi: [10.1016/S0378-1119\(03\)00747-9](https://doi.org/10.1016/S0378-1119(03)00747-9) PMID: [14563547](https://pubmed.ncbi.nlm.nih.gov/14563547/)
45. De Folter S, Angenent GC (2006) *trans* meets *cis* in MADS science. *Trends Plant Sci* 11: 224–231. doi: [10.1016/j.tplants.2006.03.008](https://doi.org/10.1016/j.tplants.2006.03.008) PMID: [16616581](https://pubmed.ncbi.nlm.nih.gov/16616581/)
46. Davies B, Schwarz-Sommer Z (1994) Control of floral organ identity by homeotic MADS-box transcription factors. *Results Probl. Cell Differ* 20: 235–258. doi: [10.1007/978-3-540-48037-2_11](https://doi.org/10.1007/978-3-540-48037-2_11)
47. Ma H (1994). The unfolding drama of flower development: recent results from genetic and molecular analyses. *Genes Dev* 8: 745–756. PMID: [7926764](https://pubmed.ncbi.nlm.nih.gov/7926764/)
48. Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. *Cell* 78: 203–209. doi: [10.1016/0092-8674\(94\)90291-7](https://doi.org/10.1016/0092-8674(94)90291-7) PMID: [7913881](https://pubmed.ncbi.nlm.nih.gov/7913881/)
49. Gramzow L, Ritz MS, Theißen G (2010) On the origin of MADS domain transcription factors. *Trends Genet* 26: 149–153. doi: [10.1016/j.tig.2010.01.004](https://doi.org/10.1016/j.tig.2010.01.004) PMID: [20219261](https://pubmed.ncbi.nlm.nih.gov/20219261/)
50. Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* 23: 865–872. doi: [10.1105/tpc.110.081737](https://doi.org/10.1105/tpc.110.081737) PMID: [21378131](https://pubmed.ncbi.nlm.nih.gov/21378131/)
51. Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönning WE, et al. (1992) *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *EMBO J* 11: 4693–4704. PMID: [1361166](https://pubmed.ncbi.nlm.nih.gov/1361166/)
52. Goto K, Meyerowitz EM (1994) Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev* 8: 1548–1560. doi: [10.1101/gad.8.13.1548](https://doi.org/10.1101/gad.8.13.1548) PMID: [7958839](https://pubmed.ncbi.nlm.nih.gov/7958839/)
53. Angenent GC, Franken J, Busscher M, Colombo L, van Tunen AJ (1993) Petal and stamen formation in petunia is regulated by the homeotic gene *fbp1*. *Plant J* 4: 101–112. doi: [10.1046/j.1365-313X.1993.04010101.x](https://doi.org/10.1046/j.1365-313X.1993.04010101.x) PMID: [8106081](https://pubmed.ncbi.nlm.nih.gov/8106081/)
54. Yao JL, Dong YH, Morris BAM (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. *Proc Natl Acad Sci USA* 98: 1306–1311. doi: [10.1073/pnas.98.3.1306](https://doi.org/10.1073/pnas.98.3.1306) PMID: [11158635](https://pubmed.ncbi.nlm.nih.gov/11158635/)
55. Sreekantan L, Torregrosa L, Fernandez L, Thomas MR (2006) *Vvmads9*, a class B MADS-box gene involved in grapevine flowering, shows different expression patterns in mutants with abnormal petal and stamen structures. *Funct Plant Biol* 33: 877–886. <http://dx.doi.org/10.1071/FP06016>
56. Lange M, Orashakova S, Lange S, Melzer R, Theißen G, Smyth DR, et al. (2013) The seirena B class floral homeotic mutant of California Poppy (*Eschscholzia californica*) reveals a function of the enigmatic PI motif in the formation of specific multimeric MADS domain protein complexes. *Plant Cell* 25: 438–453. doi: [10.1105/tpc.112.105809](https://doi.org/10.1105/tpc.112.105809) PMID: [23444328](https://pubmed.ncbi.nlm.nih.gov/23444328/)

57. Winter KU, Weiser C, Kaufmann K, Bohne A, Kirchner C, Kanno A, et al. (2002) Evolution of class B floral homeotic proteins: Obligate heterodimerization originated from homodimerization. *Mol Biol Evol* 19: 587–596. PMID: [11961093](#)
58. Whipple CJ, Ciceri P, Padilla CM, Ambrose BA, Bandong SL, Schmidt RJ (2004) Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* 131: 6083–6091. doi: [10.1242/dev.01523](#) PMID: [15537689](#)
59. Bartlett ME, Specht CD (2010) Evidence for the involvement of *GLOBOSA*-like gene duplications and expression divergence in the evolution of floral morphology in the Zingiberales. *New Phytologist* 187: 521–541. doi: [10.1111/j.1469-8137.2010.03279.x](#) PMID: [20456055](#)
60. Smaczniak C, Immink RGH, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADS domain factors: insights from recent studies. *Development* 139: 3081–3098. doi: [10.1242/dev.074674](#) PMID: [22872082](#)
61. Mandel MA, Yanofsky MF (1998) The *Arabidopsis* *AGL9* MADS box gene is expressed in young flower primordia. *Sex Plant Reprod* 11: 22–28. doi: [10.1007/s004970050116](#)
62. Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development*, 127: 725–734.
63. Adamczyk BJ, Lehti-Shiu MD, Fernandez DE (2007) The MADS domain factors *AGL15* and *AGL18* act redundantly as repressors of the floral transition in *Arabidopsis*. *Plant J* 50: 1007–1019. doi: [10.1111/j.1365-3113X.2007.03105.x](#) PMID: [17521410](#)
64. Theißen G, Saedler H (2001) Plant biology-floral quartets. *Nature* 409: 469–471. doi: [10.1038/35054172](#) PMID: [11206529](#)