Pak. J. Bot., 48(6): 2423-2431, 2016.

ISOLATION AND ANALYSIS OF DIFFERENTIALLY EXPRESSED GENES BETWEEN MALE FERTILE AND MALE STERILE FLOWER BUDS OF MARIGOLD (*TAGETES ERECTA* L.)

ZHIQIANG HOU[1,2], DAOCHENG TANG[2] AND NAN TANG[1, 2*]

[1]*State Key Laboratory of Plateau Ecology and Agriculture, College of Agriculture and Animal Husbandry, Qinghai University, 810016 Xining, China*
[2]*Key Laboratory of Qinghai Province for Landscape Plants Research, Plateau Flower Research Center, Qinghai University, 810016 Xining, China*
**Corresponding author's email: natasha_tn@hotmail.com; telephone: (86)09715362782*

Abstract

Male sterility is an important approach in utilization of heterosis in marigold (*Tagetes erecta* L.). Study on the mechanism of male sterility is very important, especially in mining of fertility-related genes. Three suppression subtractive hybridization (SSH) cDNA libraries were constructed between male fertile and male sterile flower buds of marigold. Out of 1920 clones, five hundred and six positive clones were verified by dot-blot hybridization. Two hundred and eighty-six non-redundant ESTs were obtained of which, one hundred and ninety-two ESTs corresponding to proteins with known functions. Through GO function annotation, fifteen candidate genes that may have a function in male sterility were identified. These genes involved in hormone pathways and cell cycles as well as encoded transcription factors and protein kinases. Further more, five of them were verified by quantitative real-time PCR, they were *CDKB2;1* functioned in cell division, *AMS* involved in anther wall tapetum development, *LAP3* played a role in pollen exine formation, *ACOS5* and *CYP703A2* involved in sporopollenin biosynthetic process. This is the first study that constructing cDNA libraries containing differentially expressed gene pools associate with male fertility using SSH strategy, and provides a first step to understand the mechanism of male reproductive development in marigold.

Key words: *Tagetes erecta* L., Male sterility, Suppression subtractive hybridization, Differentially expressed genes, Quantitative real-time PCR.

Introduction

Marigold (*Tagetes erecta* L.), an annual herb belonging to the composite family, is widely planted all over the world for its multi-purposes in ornamental, pigment and biological control (Vasudevan *et al*., 1997; Ballcoelho *et al*., 2003; Bosma *et al*., 2003). Numerous F1 hybrid cultivars have been bred since 1960, including some hybrids with excellent ornamental characters such as short plant height, big flower size, double petals, and so on. These hybrids benefited from the application of male sterile lines in hybrid seed production. To be specific, one male sterile line is half of a 'male sterile two-type line' of which contains 50% male fertile plants and 50% male sterile plants which have completely abortive stamens but normal and fertile pistils. This character makes it easy to pollenate male sterile plants with pollens from restorer lines, and successfully produce elite F1 hybrids. In addition, the male sterility trait can be maintained through crossing between male sterile and male fertile plants of 'male sterile two-type line' (Sreekala *et al*., 2003).

Genetic analysis in marigold supported the model of recessive genic male sterility (RGMS), i.e., a pair of recessive nuclear genes (msms) was responsible for male sterility in marigold (He *et al*., 2010). Histological and cytological observation showed that male fertile and male sterile flowers of marigold 'male sterile two-type line' varied a lot in morphology and cytology (He *et al*., 2010). However, little is known about molecular and cellular processes of stamen development including genes involved in male fertility and causes for male sterility. As male sterility is an important tool for breeding in general and for marigold breeding-in particular, it is important to understand which gene or genes underlying male sterility. In this way, more directed selection for male sterility can be developed, and homologous genes in other crops can be tested and employed as well to increase breeding efficiency. Therefore, as a first step towards revealing the molecular mechanism of marigold male sterility, we aimed at mining fertility-related genes of marigold.

Suppression subtractive hybridization (SSH) has been widely used in the isolation of differentially expressed genes from two comparative samples (Sahebi *et al*., 2015). Firstly, by subtractive hybridization, the common transcripts are hybridized between tester and driver (reference) cDNA, and the remaining un-hybridized (differentially expressed) transcripts in tester cDNA are normalized. Secondly, by suppression PCR, differentially expressed cDNAs in tester cDNA are selectively amplified. In this way, SSH libraries representing specific (differentially expressed) transcripts in tester cDNAs are constructed (Diatchenko *et al*., 1996). SSH is useful in exploring transcriptional responses between male sterile and male fertile lines for instance in *Zinnia elegans* (Pang *et al*., 2012), *Brassica napus* (Wu *et al*., 2007) and *Triticum aestivum* (Chang *et al*., 2006; Yang *et al*., 2011).

In this study, three lines with different fertility genotypes have been used for SSH, thus constructed 3 libraries containing differentially expressed (up-regulated) cDNAs in male fertile and male sterile flower buds of marigold. Sanger sequencing was performed to characterize these cDNAs and generated expressed sequence tags (ESTs). Then, homology search and gene ontology (GO) annotation were performed based on public databases to annotate to the retrieved genes. The male fertility-related candidate genes were further analyzed by quantitative real-time PCR (qRT-PCR). In this way, this study explored the relationships between male fertility/sterility and mRNA expression patterns, to obtain more insight in male reproductive development of marigold.

Materials and Methods

Plant materials: This study used a recessive genic male sterile two-type line '2-2' (provided by Plateau Flower Research Center of Qinghai University, Xining, China). The line '2-2' include 50% male sterile population (genotype, msms) and 50% male fertile population (Msms), and is propagated by crossing 'msms×Msms'. Besides, the homozygous male fertile population (MsMs) was generated by two generations of self-crossing, 'Msms×Msms' and 'MsMs×MsMs'. The florets in male sterile capitulum have no anther and pollen structures in stamens tissue, and show homotic transformation of petals to sepals (sepalloid-petals) and stamens to pistils (pistilloid-stamens) (Fig. 1a). The male fertile plants have normal flower organs (Fig. 1b). Three types of flower buds (MsMs, Msms and msms) with diameter=5mm were collected in the field, put into liquid nitrogen immediately, and then kept in -80º until RNA isolation.

RNA isolation and detection: Total RNA were extracted from 600 mg tissues of each type flower buds using the Trizol protocol (Tiangen, China). The purity and integrity of total RNA was judged by A260/A280 and 1% agarose gel electrophoresis. DNase I (Thermo scientific, USA) was used to remove genomic DNA. Then mRNA was picked by oligo(dT) cellulose (Sangon, China) from total RNA, and was also detected by electrophoresis.

Subtracted cDNA libraries construction: Three SSH libraries were constructed to enrich differentially expressed genes in each of three mRNA populations (Msms, MsMs, msms) using 'Msms/msms', 'MsMs/msms' and 'msms/Msms' as 'tester/driver' combinations. Accordingly, we called these three SSH libraries FSL-I (male fertile subtractive library I), FSL-II (male fertile subtractive library II) and SSL (male sterile subtractive library). SSH procedure was conducted using the PCR-Select^{TM} cDNA Subtraction Kit (Clontech, USA) according to the manufacturer's instructions. The first step, mRNA was reverse transcripted and finally double-strand cDNA was synthesized. Secondly, cDNA was digested with *Rsa* I to generate shorter, blunt-ended cDNA fragments, and then two tester populations were created with different adaptors (tester with adaptor 1 and tester with adaptor 2R). For the third, two testers were separately hybridized with driver cDNA (in excess) to equalize and enrich differentially expressed sequences, and the two hybridized products were added to excess driver cDNA for further hybridization. In this way, templates for PCR amplification were generated, and then for the final step, exponentially amplifying the differentially expressed sequences with nest primers. In addition, to ensure the efficiency of SSH, TubF/TubR (5'-TGATTCGGTTCTGGATGTT-3'/5'-CACGAGATGTGAGTGGAGC-3', designed from sequence of *Zinnia elegans* house-keeping gene β-tubulin, NCBI No. D63138.1) was used to analyze the ligation efficiency and subtraction efficiency.

Screening to clones of SSH libraries: The nested PCR amplified products were inserted into pGM-T vectors (Tiangen), and the recombinant plasmids were cloned in *Escherichia coli* DH5α cells under conditions of ampicillin and X-gal/IPTG. White colonies were picked, and colony PCR were then performed to screen single band clones using nested primers pair NP1/NP2R. In order to further selected positive clones, all white colonies were put into dot-blot hybridization using DIG High Prime Labeling and Detection Starter Kit I (Roche, Switzerland). DIG probe preparing, membrane dotting, hybridization and signal detection were conducted according to the manufacturer's instructions. To be more specific, the second round nested PCR products of each library were purified and then labeled by DIG-High Prime to produce male fertile and male sterile DNA probe for hybridization. One micro-liter colony PCR products was dotted on membrane (Roche, USA), and then baked at +80º to fix the DNA. All clone were performed Dot-blot with both fertile and sterile probes separately, and the hybridization signal was compared by naked eye observation. In FSLs, clones which had significant stronger hybridization signal in male fertile probe than male sterile probe were identified as positive clones, and it was opposite in SSL.

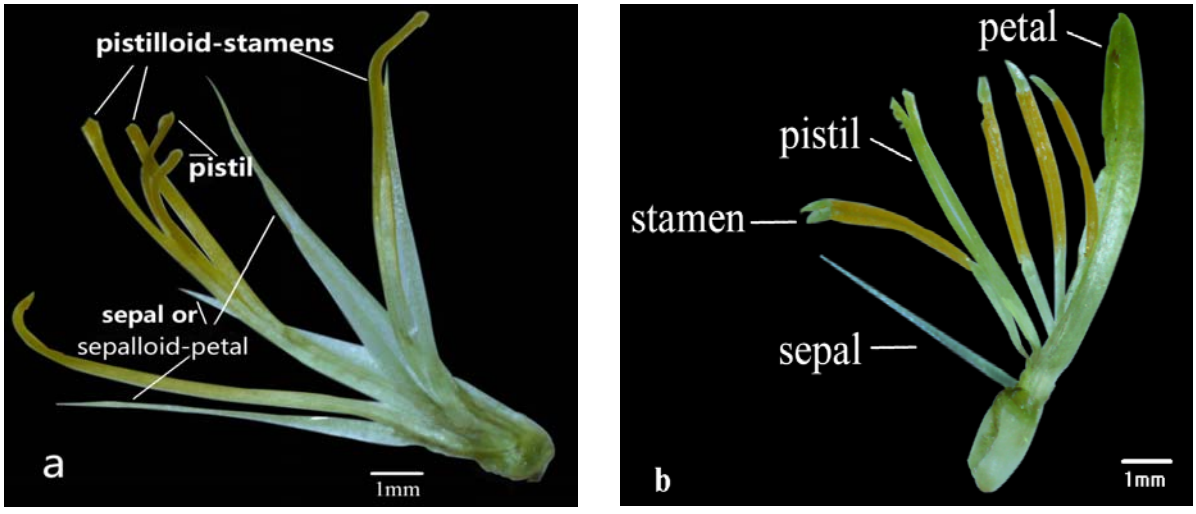

Fig. 1. Morphology of male sterile floret (a) and male fertile tubiform floret (b). Florets with 10 mm length were dissected, the anatomy observation was carried out by microscope OLYMPUS SZX7, and the two images were modified by Photo-Shop software. Some sepals, petals or stamens were lost during florets dissection.

Sequencing and ESTs analysis: Positive and single clones were sequenced by M13F primer using ABI 3730 Genetic Analyzer (Sangon, China). Low quality sequences, vector and adapter sequences were cleaned manually, and similar sequences (similarity>95%) were detected and removed, thus obtaining a collection of expressed sequence tags (ESTs) for each library. All ESTs were annotated to the non-redundant protein sequences database (NR) by Basic Local Alignment Search Tool (BLAST) in Blast2go software, with BLASTX *Expect*-value (*E*-value) threshold <1 and Gene Ontology (GO) annotation threshold *E*<0.001. Ribosomal sequences were further removed. Sequence length, *E*-value, protein description, and GO terms were examined. Blast 2 go is an automated tool for sequence BLAST and assignment of GO terms using novel sequence database (Conesa & Götz, 2008). Gene Ontology is a hierarchic system for gene annotation starting with three main categories, biological process, molecular function, and cellular component.

Quantitative real-time PCR (qRT-PCR): qRT-PCR was performed in order to confirm the differential expression pattern of candidate genes. For each genotype (Msms, MsMs, msms), flower buds (diameter=5mm) were sampled separately from three individual plants (three biological replications), and total RNA were extracted using the Trizol protocol (Tiangen). First strand cDNA synthesis and qRT-PCR were performed using the Fast Quant RT Kit and Super Real Pre Mix Plus Reagent SYBR Green (Tiangen). Gene-specific primers and β-tubulin primers (inner-reference) were designed by Premier Primer 5 software (Table 3). The PCR mixture were 1μL cDNA template (1st cDNA reaction), 10μL 2 × Super Real PreMix Plus, 0.5μL of each forward and reverse primer (10μmol/L), and water to a final volume of 20μL. The thermal cycling conditions were programmed with an initial denaturing step at 95° for 15min, followed by 40 cycles of 95° for 10s and 60° for 30s. The relative expression level of target genes in different cDNA templates was calculated by $2^{-\Delta\Delta ct}$ method , defining the relative higher expressed level as 1(Livak & Schmittgen, 2001).

Results

Detection of total RNA and mRNA: For each library, the detection of total RNA integrity with agarose gel electrophoresis showed 2 bright bands, corresponding 28S and 18S ribosomal RNA with a ratio of intensities of 2:1 (Fig. 2). The values of A260/A280 of two total RNA samples were both 2.000. Two hundred micrograms of total RNA was used to pick mRNA, and the detection of mRNA quality with agarose gel electrophoresis appeared smears between 0.3kb to over 2kb (Fig. 2). This indicated that the quality of total RNA and mRNA were satisfactorily high.

Quality control and screening of three SSH cDNA libraries: To ensure the quality of three cDNA libraries, the efficiency of the SSH was evaluated by PCR with TubF/TubR specific for a house-keeping gene β-tubulin. In FSL-I, the subtracted tester showed the β-tubulin band until 35 cycles, while the unsubtracted tester showed the same band as early as 20 cycles (Fig. 3). In FSL-II and SSL, the un-subtracted tester showed 5 and 10 cycles earlier than the subtracted tester (Fig. 3). This demonstrated common transcripts including the house keeping gene were effectively subtracted, so the SSH procedure was effective in three libraries. By picking white clones, nested-primer PCR and dot-blot hybridization, positive and single clones representing differentially expressed sequences were screened (Fig. 4). As a result, a total of 506 positive single clones were characterized from 1920 white colonies in 3 libraries, with ratio in SSL (70/576) was obviously smaller than in FSL-I (203/768) and FSL-II (233/576) (Table 1).

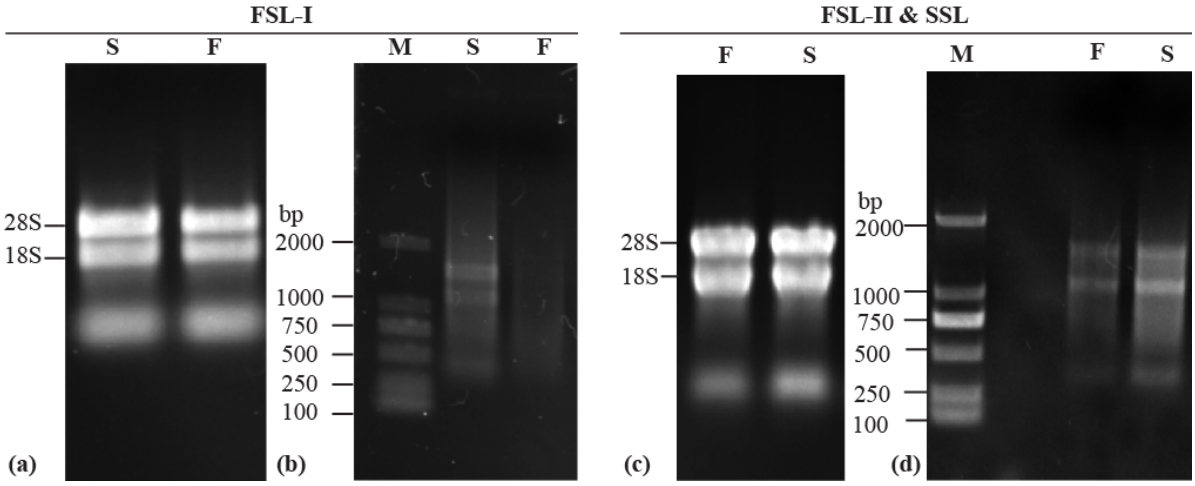

Fig. 2. Detection of total RNA and mRNA by gel electrophoresis. (a) and (b), total RNA and mRNA samples for FSL-I library construction. (c) and (d), total RNA and mRNA for FSL-II & SSL libraries construction (forward and reverse subtracted). F, fertile plant. S, sterile plant. M, DL 2000 marker.

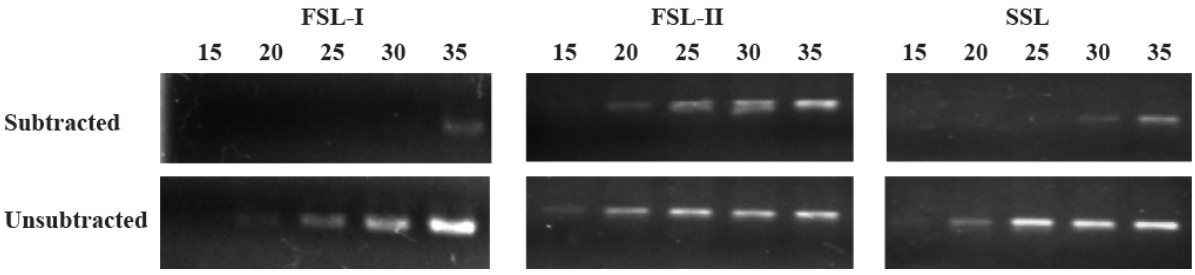

Fig. 3. Detection of subtraction efficiency for 3 SSH libraries through identifying amplified products of β-tubulin at 15, 20, 25, 30 and 35 PCR cycles.

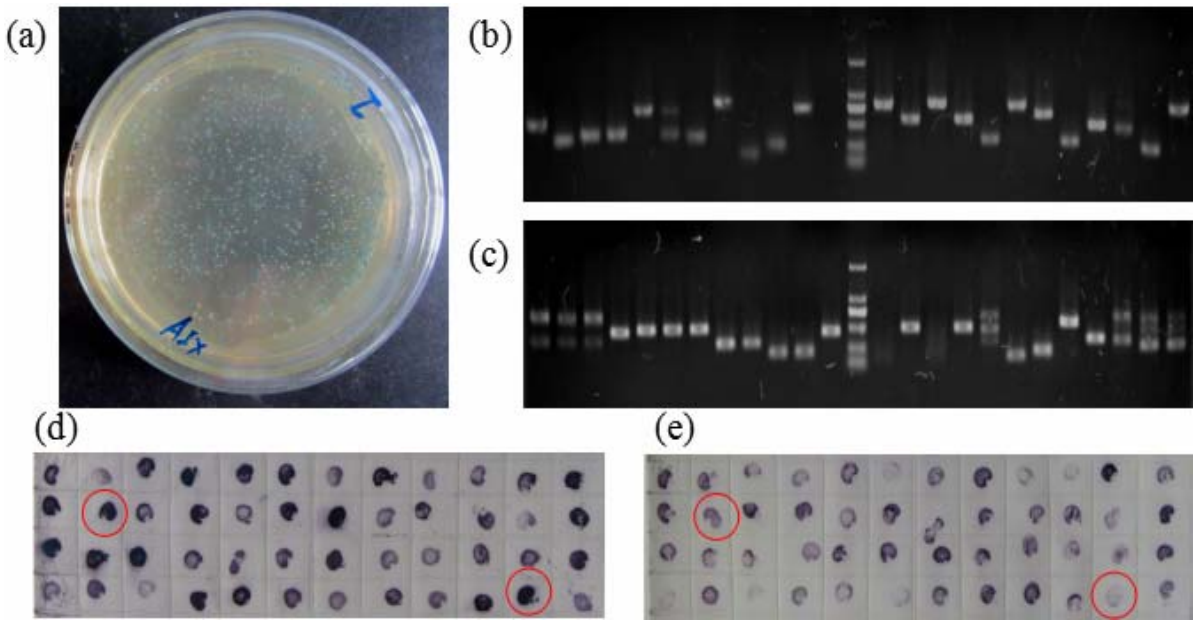

Fig. 4. Example of screening SSH library clones. (a), petridish include blue and white clonies. (b)&(c), PCR identification of white clonies TeF1301~TeF1348 (from FSL-II library) using nested primers pair NP1/NP2R. (d)&(e), Dot-blot signals of clonies TeF1301~TeF1348 using fertile (d) & sterile (e) probes. Red circle labeled dots were examples of up-regulated clonies in FSL-II library.

Table 1. Statistics of SSH libraries.

	FSL-I	FSL-II	SSL	Total
White bacilli	768	576	576	1920
Dot-blot positive clones (ESTs)	203	233	70	506
Non-repetitive ESTs (Nr-ESTs)	154	78	54	286
Average length of Nr-ESTs (bp)	418	413	308	396
ESTs with/without BLAST hits	128/26	53/25	48/6	229/57
ESTs with known/unknown protein functions	112/16	39/14	41/7	192/37
Assigned GO terms	134	183	184	501

ESTs BLAST and GO annotation: ESTs of each library were generated by Sanger sequencing. By sequences similarity checking in blast2go software (sequence similarity<5%), number of non-repetitive ESTs were 154 (FSL-I), 78 (FSL-II) and 54 (SSL), and no similar sequences were found among these libraries (Table 1). The average length of these 286 sequences was 396 bp. Results of BLASTX showed that a total of 229 ESTs with BLAST hits were obtained of which, one hundred and ninety-two ESTs corresponding to proteins with known functions. A relative higher ratio of ESTs with known protein functions (112/192) was found in FSL-I library, while only 41 were found in SSL library (Table 1). GO annotation was operated to those ESTs with known protein functions and, the results showed a total of 317 GO terms were assigned to FSL-I and FSL-II library, while 184 GO terms were assigned to SSL library. Usually, an EST was assigned to several GO terms at the same or different hierarchic levels, in this way, all ESTs were gradually divided into different subcategories and, this process started from the 192 annotated ESTs being repeatedly assigned to level 2 GO terms in each of three main categories, biological process, cellular component and molecular function (Fig. 5).

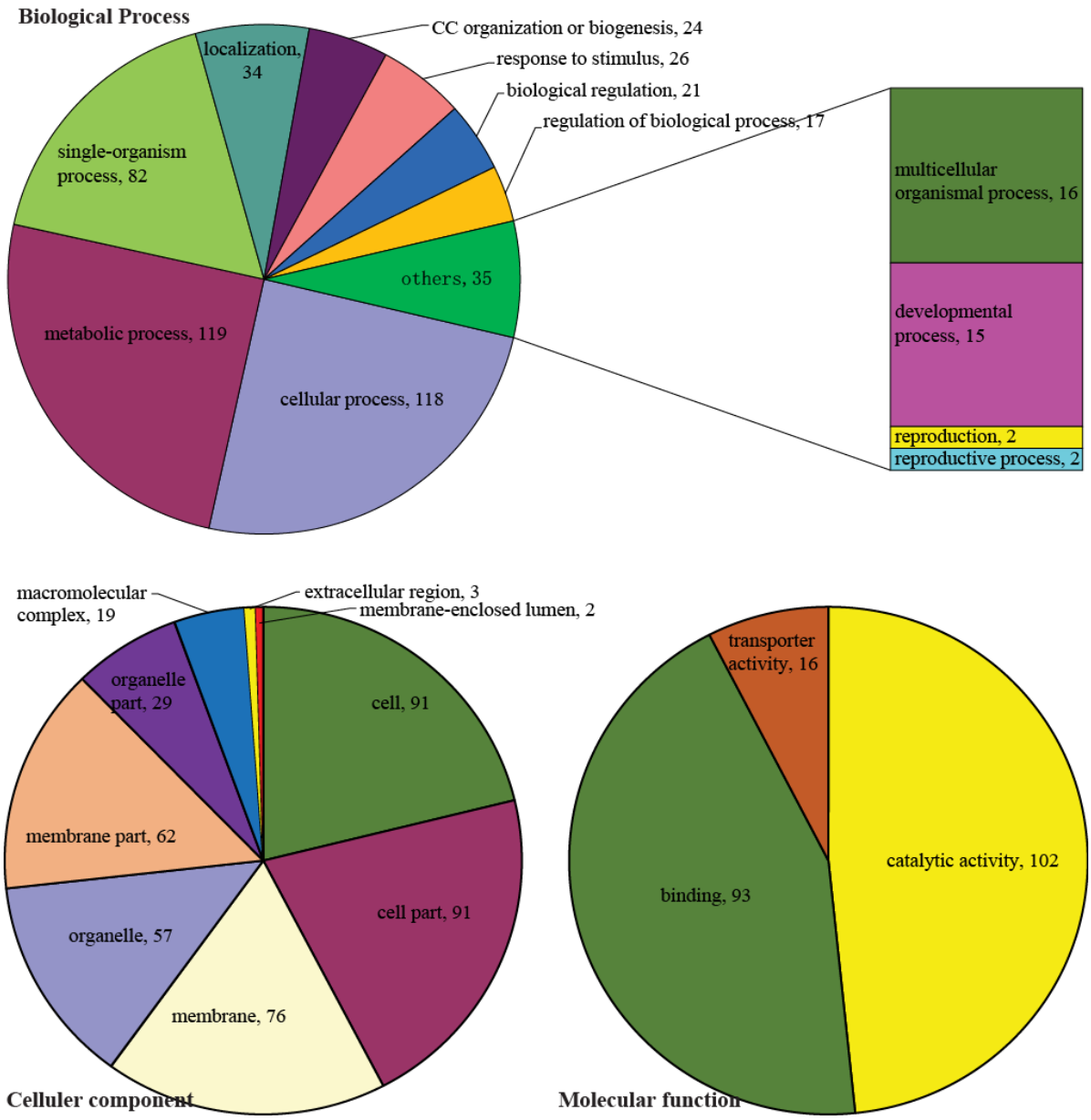

Fig. 5. Pie charts of GO annotation terms distribution at hierarchic level 2 in three main categories, Biological Process, Cellular Component and Molecular Function.

Biological process refers to a biological objective to which the gene or gene product contributes (Ashburner *et al*., 2000). In this category, 'metabolic process' (level 2) was the most represented term including 119 ESTs from three libraries (Fig. 5). This term were then divided into many lower level GO terms (not shown in Fig. 5), e.g., 'phosphorus metabolic process' (27 ESTs),'carbohydrate metabolic process' (22ESTs), 'protein metabolic process' (47 ESTs) and 'cellular nitrogen compound metabolic process' (36 ESTs). Further more, GO terms 'nucleobase-containing small molecule metabolic process', 'carbohydrate derivative metabolic process' and 'glycosyl compound metabolic process' were specifically assigned to 6 ESTs in SSL library, on the other hand, 'lipid metabolic process' (4 ESTs) and 'sulfur compound metabolic process' (2 ESTs) were specific to FSL-II library. This result implied some metabolic responses related to male fertility including carbohydrate, lipid and sulfur compounds. The second largest term 'cellular process' included 118 ESTs (Fig. 5). They were mainly assigned to 'regulation of cellular process' (15 ESTs), 'cellular component organization' (24 ESTs), 'celluar metabolic process' (87 ESTs) at level 3. The third term 'developmental process' had 15 ESTs, and was of particular interest to us. This group contained many GO terms related to male fertility, e.g., 'pollen development', 'regulation of meristem growth', 'anther wall tapetum development' and 'anther dehiscence'. In addition, other GO terms like 'response to stimulus' and 'regulation of biological process' were assigned to 26 and 17 ESTs in three libraries (Fig. 5).

Cellular component refers to the place in the cell where a gene product is active (Ashburner *et al*., 2000). This category contained significant representation of 'cell' (91 ESTs), 'organelle' (57 ESTs), 'membrane' (76 ESTs), 'macromolecular complex' (19 ESTs) (Fig. 5). Some were specific to SSL library, e.g., level 5 terms 'mitochondrial envelope' (3 ESTs) and 'chloroplast part' (3 ESTs). This indicated some mitochondrial and chloroplast genes were disturbed in male sterile flower buds. Finally, molecular function is defined as the biochemical activity of a gene product (Ashburner *et al*., 2000). In this category, it showed the most represented GO terms were of binding function such as 'ion binding', 'lipid binding', 'histone binding', 'enzyme binding', 'nucleic acid binding' in addition to all types of activities such as 'hydrolase activity', 'ligase activity', 'oxidoreductase activity', 'transferase activity' and 'kinase activity' (Fig. 5). Some terms e.g. 'kinase activity', 'methyltransferase activity' and 'phosphotransferase activity' were specific to two FSL libraries.

Based on above GO functional annotations, fifteen candidate genes were identified as homologues involved in flower developent or male reproductive process in marigold (Table 2). These homologous genes were: (1) protein kinase genes and hormone-related genes *PK5, MPK1, CDKB2;1, CDKB2;2, SERK2* and *LOX3*; (2) transcription factors and flower transition genes *AMS*, *HUA2* and *SHD*; (3) cell cycle genes *GSL10* , *MGH3* and *MSH2*; (4) pollen wall formation related genes *CYP703A2*, *ACOS5* and *LAP3*.

Quantitative real-time PCR (qRT-PCR): By qRT-PCR, relative expression level of these 15 male fertility related candidate genes were compared between 3 male fertile plants and 3 male sterile plant. For candidate genes selected from FSL-I library, *CDKB2;1* (Clone No. TeF832) was the only one that was significantly differentially expressed (up-regulated) in male fertile flower buds (FBU) than in male sterile buds (SBU) (mean value of relative expression fold >3), others were down-regulated or similar expressed (mean value of relative expression fold <3) (Table 3). This result was not consistent with the data of Dot-blot hybridization and, the causes needed further analyze. In library FSL-II, four candidate genes (Clone No. TeF1116, TeF1141, TeF1215, TeF1314) were significantly up-regulated in FBU than in SBU, while other two (Clone No. TeF1167, TeF1180) were not significantly up-regulated. In SSL library, two candidate genes (Clone No. TeS562, TeS590) were not significantly up-regulated in SBU than in FBU. In addition, expression of these genes were not stable according to the big standard deviation value (SD) in 3 individuals (Table 3). Consequently, five candidate genes (*CDKB2;1*, *LAP3*, *ACOS5*, *AMS* and *CYP703A2*) were confirmed as up-regulated expression in male fertile buds, other 10 genes needed further study on the temporal expression profiles as well as the sequence characters and biological functions.

Table 2 List of candidate genes related to male fertility.

Clone No.	Size (bp)	*E*-value	Description	TAIR locus	Gene name	References
TeF535	289	1.40E-08	mitogen-activated protein kinase 1	AT1G10210	*MPK1*	Hamel *et al*., 2006
TeF543	674	5.60E-28	glucan synthase-like 10	AT3G07160	*GSL10*	Armin *et al*., 2008
TeF735	791	9.30E-70	D6 protein kinase like 2	AT5G47750	*PK5*	Barbosa *et al*., 2014
TeF752	178	3.20E-16	cyclin-dependent kinase B2;2	AT1G20930	*CDKB2;2*	Andersen *et al*., 2006
TeF783	363	4.40E-09	lipoxygenase 3	AT1G17420	*LOX3*	Caldelari *et al*., 2010
TeF807	927	4.90E-04	flower transition repressor	AT5G23150	*HUA2*	Doyle *et al*., 2005
TeF832	333	5.30E-36	cyclin-dependent kinase B2;1	AT1G76540	*CDKB2;1*	Andersen *et al*., 2006
TeF1116	455	1.20E-42	calcium-dependent phosphotriesterase	AT3G59530	*LAP3*	Dobrista *et al*., 2009
TeF1141	373	6.70E-22	acyl-CoA synthetase 5	AT1G62940	*ACOS5*	Morant *et al*., 2007
TeF1167	584	4.90E-60	male-gamete-specific histone H3	AT1G19890	*MGH3*	Mathieu *et al*., 2007
TeF1180	346	8.00E-39	floral meristem maintenance	AT4G24190	*SHD*	Ishiguro *et al*., 2002
TeF1215	550	9.50E-27	bHLH DNA-binding transcription factor	AT2G16910	*AMS*	Xu *et al*., 2014
TeF1314	590	6.10E-56	cytochrome P450 703A2	AT1G01280	*CYP703A2*	de Azevedo *et al*., 2009
TeS562	193	6.00E-04	MUTS homolog 2	AT3G18524	*MSH2*	Emannuel *et al*., 2006
TeS590	198	1.30E-06	somatic embryogenesis receptor-like kinase 2	AT1G34210	*SERK2*	Albrecht *et al*., 2005 & 2008

Table 3. Primer sequences and results of quantitative real-time PCR (qRT-PCR).

Clone No.	Forward primer [5'→3']	Reverse primer [5'→3']	Product size (bp)	Relative expression fold (M±SD)	
				FBU	SBU
TeF535	TAGATGAAGATTGGGGTGAG	CATTACATCCATGCTTATTACAG	170	0.60±0.25	1±0.36
TeF543	GGTTCCCATTTGTTTCCACT	CGTTAGCTTTGTTTCCAGCA	102	0.62±0.21	1±0.46
TeF735	TGGAGTTTTGCCCTGGAG	GAGCATTATGTGACCGTCTTCT	193	0.17±0.26	1±0.92
TeF752	AGCTCGAGCGAGTCCAAG	GGGATCTTAAACCGCACAAT	80	1±0.11	0.92±0.23
TeF783	TCTTGCACCAAGTTCAGAGG	ACTACAATTACATCCCTCAATATGCT	128	0.92±0.11	1±1.33
TeF807	TTACCCTCCTGGTCTTCGAC	TGTTGTCAATTTGGGTCCAG	129	0.85±0.13	1±0.39
TeF832	GTCGTCAAGCTGATGGATGT	CTTGATGACTCTGGGTGGAA	150	1±0.17	0.33±0.14
TeF1116	CACAAACTGCAGGCTAATGAA	GCTTTAGCCCTGCAACAATC	150	1±1.08	1.3E-06±7.4E-07
TeF1141	TATTTCATTGTGCAAGGGTA	GGTGAGGGTGCTTCATTT	122	1±0.59	7.8E-05±6.2E-05
TeF1167	GCTCTGCTGTTGCTGCTC	AATCATCCCAAACCCTAA	178	1±0.02	0.92±0.82
TeF1180	TAACCAAATGGTGGAGGGA	AGGTTACCACCACACAAGGG	96	1±1.55	0.55±0.20
TeF1215	CATAACCCAGACCCTAAT	TAACCTCACAAATCCTCC	135	1±0.92	0.003±0.004
TeF1314	CATCCCTCACGAGTCCTT	CGGGCACTTCCTCTTTCC	229	1±5.73	0.008±0.008
TeS562	TCTAACTGAGCTACTAGAAACTGCATC	CGTCCAAGTTCATCTAATGCC	81	0.61±0.16	1±1.2
TeS590	TTGGGTCTTTTACTGCTCTTC	TTTCCACTATTCACTTCTGC	79	0.45±0.11	1±1.09
β-tubulin	GGGTGGAATGTCACAAACGG	GAGCACCAAGGAAGTGGATGAA	112		

Discussion

This study constructed three qualified SSH cDNA libraries enriching differentially expressed genes between male fertile and male sterile flower buds. More differentially expressed genes were isolated by constructing 2 male fertile libraries FSL-I and FSL-II, but it was unexpected that no common gene (sequence similarity >95%) was found between them. It was speculated that this enormous variation may be caused by using male fertile lines with different genotype (Msms/MsMs) in FSL-I/FSL-II libraries. Further more, it would be better to use both heterozygous and homozygous lines when constructing male fertility-related SSH libraries in marigold as well as in other species. Secondly, these libraries were part of temporary transcription profiles, so it was easier to isolate a large number of downstream genes, e.g., metabolism process genes, whereas it was more difficult to isolate the target gene (Ms/ms) or upstream genes. By GO annotation, fifteen candidate genes were identified out of all ESTs, and they are supposed to be more close to upstream genes related to male fertility than other ESTs. Finally, some candidate genes were failed to be confirmed by qRT-PCR, this conflicting results may be due to the complexity of gene expression in temporal and spatial scale. In spite of this, the value of these genes should not be overlooked.

Male fertile plants produce and spread viable pollen, while male sterile plants exhibit abnormal anther or pollen development (Jack, 2004; McCormick, 2004; Ma, 2005). Many genes specifying or regulating stamen development have been identified, including *APETALA3/ PISTILLATA (AP3/PI), UNUSUAL FLORAL ORGANS (UFO), SPOROCYTELESS/ NOZZLE (SPL/NZZ)* (Coen & Meyerowitz, 1991; Jack *et al*., 1992; Lohmann & Weigel, 2002; Eunyoung *et al.*, 2008; Liu *et al.*, 2009). By SSH strategy, some homologous genes in other crops have been characterized in marigold, these genes function in many anther developmental stages, and may have important roles in flower development and male fertility of marigold. For instance, *HUA2* has a probable role in floral homeotic AGAMOUS pathway and negatively regulates timing of transition from vegetative to reproductive stage in *Arabidopsis* (Doyle *et al.*, 2005), *SERK2* involves in brassinosteroid mediated signaling pathway and is very important in controlling anther cell differentiation in *Arabidopsis* (Zhao, 2009; Albrecht *et al.*, 2005 & 2008), *LOX3* functions in jasmonic acid biosynthesis and shows male sterile in *Arabidopsis lox3lox4* double mutants (Caldelari *et al.*, 2010), *GSL10* has a probable role in callose deposition in cell wall and the microspores can not able to enter into mitosis in *Arabidopsis gsl10* mutants (Armin *et al.*, 2008), *AMS* plays the central role in coordinating sporopollenin biosynthesis and pollen wall patterning in *Arabidopsis* (Xu *et al.*, 2014). Therefore, it was speculated that genes encode transcription factors and protein kinases together with genes involved in hormone pathways and cell cycles probably play important roles in maintaining of male fertility in marigold. Moreover, it was reported that dominant genic male sterility *Brassica napus* was probably caused by the abnormality in the DNA damage repair system during meiosis (Wu & Yang, 2008), and similarly in this study, TeS562 was identified as homologous gene *MSH2* which functions in DNA repair system in *Arabidopsis* (Emmanuel *et al.*, 2006). So, TeS562 was selected as a candidate gene that probably related to marigold male sterility.

The male sterility trait in marigold was caused by spontaneous floral organ homeotic transformation of petals into sepals (sepallody) and stamens into pistil-like structures (pistillody) (He *et al.*, 2010). We also observed this phenomenon in line '2-2' (Fig. 1a). Further more, the male fertile line and the male sterile line of '2-2' share the same cytoplasm and exhibit similar plant phenotype except floral morphology. These demonstrated that the male sterility trait were most probably happened at the stamen identity stage, and were very likely caused by the loss-of-function of floral homeotic B function genes. According to the ABC model of flower organ identity, the B function genes specify petal and stamen development, and sepallody or pistillody is attributed to mutations of B function genes *AP3*/*PI* in *Arabidopsis* (Bowman *et al.*, 1989; Goto & Meyerowitz, 1994; Krizek & Fletcher, 2005; Murai, 2013). In this study, none floral homeotic B function homologue was identified among 286 ESTs and, this may be related to relative small volume of cDNA libraries.

It is well known that the floral homeotic genes were regulated by other genes in floral pathway. For instance, it was reported another protein kinase gene *ASK1 (ARABIDOPSIS SERINE/THREONINE KINASE 1)* regulates early *AP3* and *PI* expression in *Arabidopsis* and, this support the importance of 'floral building genes' for floral organ identity (Zhao *et al.*, 2001). Besides, wheat line Norin 26 with cytoplasm of *Aegilops crassa* L. showed pistil-like stamens, and the nuclear-cytoplasm interaction altered the expression pattern of class-B MADS-box gene (Saraike *et al.*, 2007; Zhu *et al.*, 2008; Yamamoto *et al.*, 2013). In this study, it was found the sequence TeF735 was high similar with gene *WPPK1* (*wheat pistillody-related protein kinase 1*) which was reported strongly expressed in pistils and pistil-like stamens in the wheat pistillody line than the normal line (Saraike *et al.*, 2007). This result suggested the importance of TeF735, and also supported other genes may be involved in floral identity by regulating class B function genes in male sterile marigold.

This study improved our understanding to male sterility in marigold. Some limitations of this study are: (1) relative small volume of these libraries especially in male sterile subtractive library; (2) some candidate genes were failed to be confirmed; (3) no floral homeotic genes was identified. In spite of this, SSH is still an effective strategy in mining male fertility-related genes, and these problems could be solved if we construct larger libraries and perform more accurate qRT-PCR analysis. Further more, some methods should be performed to enlarge the value of these cDNA libraries, e.g., cloning full length cDNA by RACE (rapid-amplification of cDNA ends), and then performing bio-information analysis of a candidate gene.

Conclusion

With purpose of mining male fertility-related genes of marigold, this study isolated 286 differentially expressed ESTs from male sterile and male fertile flower buds (5mm diameter) by constructing and screening suppression subtractive hybridization (SSH) cDNA libraries. Using homology search and GO annotation, fifteen candidate genes were identified. They were: (1) protein kinase genes and hormone-related genes *PK5, MPK1, CDKB2;1, CDKB2;2, SERK2* and *LOX3*; (2) transcription factors and flower transition genes *AMS*, *HUA2* and *SHD*; (3) cell cycle genes *GSL10* , *MGH3* and *MSH2*; (4) pollen wall formation related genes *CYP703A2*, *ACOS5* and *LAP3*. Among these candidate genes, five (*CDKB2;1, LAP3, ACOS5, AMS* and *CYP703A2*) were confirmed as significantly differentially expressed (up-regulated) genes in male fertile buds than in male sterile buds (relative expression level folds >3) by qRT-PCR analysis, other 10 genes needed further confirmation.

Acknowledgments

We would like to thank the support by the CAS 'Light of West China' Program for supporting the research 'Development of Molecular Markers for Fertility Identification of *Tagetes erecta* L. Male Sterile AB Line During Seedling Stage', the support by Department of Science and Technology of Qinghai Province for the Development of Key Laboratory of Qinghai Province for Landscape Plants Research. We want to thank Tao He, Hou-hua Li and Yu-hu Shen for their helpful comments on an earlier version of this paper.

References

Albrecht, C., E. Russinova, B. Kemmerling, M. Kwaaitaal and S.C. de Vries. 2008. *Arabidopsis* somatic embryogenesis receptor kinase proteins serve brassinosteroid-dependent and -independent signaling pathways. *Plant Physiol.,* 148: 611-619.

Albrecht, C., E. Russinova, V. Hecht, E. Baaijens and S.C. de Vries. 2005. The *Arabidopsis thaliana* somatic embryogenesis receptor-like Kinases 1 and 2 control male sporogenesis. *Plant Cell,* 17: 3337-3349.

Andersen, S.U., S. Buechel, Z. Zhao, K. Ljung, O. Novák, W. Busch, C. Schuster and J.U. Lohmann. 2008. Requirement of B2-type cyclin-dependent kinases for meristem integrity in *Arabidopsis thaliana*. *Plant Cell,* 20: 88-100.

Armin, T., B. Lynette, N. Christina, T. David and S. Paul. 2008. Dual function of *Arabidopsis* glucan synthase-like genes *GSL8* and *GSL10* in male gametophyte development and plant growth. *Plant J.,* 54: 911-923.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. *Nature Genet.,* 25(1): 25-29.

Ballcoelho, B., A.J. Bruin, R.C. Roy and E. Riga. 2003. Forage pearl millet and marigold as rotation crops for biological control of root-lesion nematodes in potato. *Agron. J.,* 95: 282-292.

Barbosa, I.C.R., M. Zourelidou, B. Willige, B. Weller and C. Schwechheimer. 2014. D6 protein kinase activates auxin transport-dependent growth and pin-formed phosphorylation at the plasma membrane. *Dev. Cell*, 29: 674-685.

Bosma, T.L., J.M. Dole and N.O. Maness. 2003. Optimizing marigold (*Tagetes erecta* L.) petal and pigment yield. *Crop Sci.,* 43: 2118-2124.

Bowman, J.L., D.R. Smyth and E.M. Meyerowitz. 1989. Genes directing flower development in *Arabidopsis*. *Plant Cell,* 1(1): 37-52.

Caldelari, D., G. Wang, E.E. Farmer and X. Dong. 2011. *Arabidopsislox3 lox4* double mutants are male sterile and defective in global proliferative arrest. *Plant Mol. Biol.,* 75: 25-33.

Chang, Q.S., R.H. Zhou, X.Y. Kong, Z.L. Yu and J.Z. Jia. 2006. Identification of differentially expressed genes during anther abortion of Taigu genic male sterile wheat by combining suppression subtractive hybridization and cDNA array. *J. Integr. Plant Biol*., 48: 1348-1354.

Coen, E.S. and E.M. Meyerowitz. 1991. The war of the whorls: Genetic interactions controlling flower development. *Nature,* 353: 31-37.

Conesa, A. and S. Götz. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.*, 2008: 619-832.

de Azevedo, S.C., S.S. Kim, S. Koch, L. Kienow, K. Schneider, S.M. McKim, G.W. Haughn, E. Kombrink and C.J. Douglas. 2009. A novel fatty Acyl-CoA Synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. *Plant Cell,* 21: 507-525.

Diatchenko, L., Y.F.C. Lau, A.P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E.D. Sverdlov and P.D. Siebert. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA,* 93: 6025-6030.

Dobritsa, A.A., S.I. Nishikawa, D. Preuss, E. Urbanczyk-Wochniak, L.W. Sumner, A. Hammond, A.L. Carlson and R.J. Swanson. 2009. *LAP3*, a novel plant protein required for pollen development, is essential for proper exine formation. *Sex. Plant Reprod.,* 22: 167-177.

Doyle, M.R., C.M. Bizzell, M.R. Keller, S.D. Michaels, J. Song, Y.S. Noh and R.M. Amasino. 2005. *HUA2* is required for the expression of floral repressors in *Arabidopsis thaliana*. *Plant J.*, 41: 376-385.

Emmanuel, E., E. Yehuda, C. Melamed-Bessudo, N. Avivi-Ragolsky and A.A. Levy. 2006. The role of *ATMSH2* in homologous recombination in *Arabidopsis thaliana*. *EMBO Rep.*, 7: 100-105.

Eunyoung, C., K.G. T. Queenie, T.A. Hill and V.F. Irish. 2008. An *Arabidopsis* F-box protein acts as a transcriptional co-factor to regulate floral development. *Development,* 135: 1235-1245.

Goto, K. and E.M. Meyerowitz. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Dev.*, 8: 1548-1560.

Hamel, L.P., M.C. Nicole, S. Sritubtim, M.J. Morency, M. Ellis, J. Ehlting, N. Beaudoin, B. Barbazuk, D. Klessig, J. Lee, G. Martin, J. Mundy, Y. Ohashi, D. Scheel, J. Sheen, T. Xing, S. Zhang, A. Seguin and B.E. Ellis. 2006. Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.,* 11: 192-198.

He, Y.H., G.G. Ning, Y.L. Sun, Y. Hu, X.Y. Zhao and M.Z. Bao. 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.

Ishiguro, S., Y. Watanabe, N. Ito, H. Nonaka, N. Takeda, T. Sakai, H. Kanaya and K. Okada. 2002. *Shepherd* is the *Arabidopsis* GRP94 responsible for the formation of functional *CLAVATA* proteins. *Embo J.,* 21: 898-908.

Jack, T. 2004. Molecular and genetic mechanisms of floral control. *Plant Cell,* 16: S1-S17.

Jack, T., L.L. Brockman and E.M. Meyerowitz. 1992. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell,* 68: 683-697.

Krizek, B. A. and J.C. Fletcher. 2005. Molecular mechanisms of flower development: an armchair guide. *Nat. Rev. Genet.,* 6: 688-698.

Liu, X., J. Huang, S. Parameswaran, T. Ito, B. Seubert, M. Auer, A. Rymaszewsky, G. Jia, H.A. Owen and D. Zhao. 2009. The *SPOROCYTELESS/NOZZLE* gene is involved in controlling stamen identity in *Arabidopsis*. *Plant Physiol.,* 151: 1401-1411.

Livak, K.J. and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle C^T}$ method. *Methods,* 25: 402-408.

Lohmann, J.U. and D. Weigel. 2002. Building beauty: The genetic control of floral patterning. *Dev. Cell,* 2: 135-142.

Ma Hong. 2005. Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu. Rev. Plant Biol.,* 56: 393-434.

Mathieu, I., H. Yuki, G. Mathieu, H. Tetsuya and B. Frédéric. 2007. Distinct dynamics of *HISTONE3*variants between the two fertilization products in plants. *Curr. Biol.,* 17: 1032-1037.

McCormick, S. 2004. Control of male gametophyte development. *Plant Cell,* 16: S1265-S1275.

Morant,M., K. Jørgensen, H. Schaller, F. Pinot, B.M. Møller, D. Werck-Reichhart and S. Bak. 2007. *CYP703* is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell*, 19: 1473-87.

Murai, K. 2013. Homeotic genes and the ABCDE model for floral organ formation in wheat. *Plants,* 2: 379-395.

Pang, R.H., G.Y. Ma, X.Y. Lou, M.Z. Bao and Y.M. Ye. 2012. Construction and analysis of suppression subtractive cDNA library from fertile disk floret bud of *Zinnia elegans*. *J. Integr. Agr.,* 11: 567-575.

Sahebi, M., M.M. Hanafi, P. Azizi, A. Hakim, S. Ashkani and R. Abiri. 2015. Suppression subtractive hybridization versus next-generation sequencing in plant genetic engineering: challenges and perspectives. *Mol. Biotechno.,* 57: 880-903.

Saraike, T., N. Shitsukawa, Y. Yamamoto, H. Hagita, Y. Iwasaki, S. Takumi and K. Murai. 2007. Identification of a protein kinase gene associated with pistillody, homeotic transformation of stamens into pistil-like structures, in alloplasmic wheat. *Planta,* 227: 211-221.

Sreekala, C. and S. Raghava. 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.

Vasudevan, P., S. Kashyap and S. Sharma. 1997. *Tagetes*: a multipurpose plant. *Bioresource Technol.*, 7: 29-35.

Wu, J., J. Shen, X. Mao, K. Liu, L. Wei, P. Liu and G. Yang. 2007. Isolation and analysis of differentially expressed genes in dominant genic male sterility (DGMS) *Brassica napus* L. using subtractive PCR and cDNA microarray. *Plant Sci.* 172: 204-211.

Wu, J.Y. and G.S. Yang. 2008. Meiotic abnormality in dominant genic male sterile *Brassica napus*. *Mol. Biol.,* 42(4): 572-578.

Xu, J., Z. Ding, G. Vizcay-Barrena, J. Shi, W. Liang, Z. Yuan, D. Werck-Reichhart, L. Schreiber, Z.A. Wilson and D. Zhang. 2014. Aborted microspores acts as a master regulator of pollen wall formation in *Arabidopsis*. *Plant Cell*, 26: 1544-1556.

Yamamoto, M., N. Shitsukawa, M. Yamada, K. Kato, S. Takumi, K. Kawaura, Y. Ogihara and K. Murai. 2013. Identification of a novel homolog for a calmodulin-binding protein that is upregulated in alloplasmic wheat showing pistillody. *Planta,* 237: 1001-1013.

Yang, Z., Z. Peng, H. Yang, J. Yang, S. Wei and P. Cai. 2011. Suppression subtractive hybridization identified differentially expressed genes in pistil mutations in wheat. *Plant Mol. Biol. Rep.,* 29: 431-439.

Zhao, D. 2009. Control of anther cell differentiation: a teamwork of receptor-like kinases. *Sex. Plant Reprod.,* 22: 221-228.

Zhao, D., Q. Yu, M. Chen and H. Ma. 2001. The *ASK1* gene regulates B function gene expression in cooperation with *UFO* and *LEAFY* in *Arabidopsis*. *Development,* 128: 2735-2746.

Zhu, Y., T. Saraike, Y. Yamamoto, H. Hagita, S. Takumi and K. Murai. 2008. *Orf260cra*, a novel mitochondrial gene, is associated with the homeotic transformation of stamens into pistil-like structures (pistillody) in alloplasmic wheat. *Plant Cell Physiol.,* 49: 1723-1733.

(Received for publication 16 December 2015)