BMC Genomics

RESEARCH ARTICLE

Open Access

CrossMark

Transcriptome profiling of differentially expressed genes in floral buds and flowers of male sterile and fertile lines in watermelon

Sun-Ju Rhee[1†], Minseok Seo[2,3†], Yoon-Jeong Jang[1], Seoae Cho[3] and Gung Pyo Lee[1*]

Abstract

Background: Male sterility is an important mechanism for the production of hybrid seeds in watermelon. Although fruit development has been studied extensively in watermelon, there are no reports on gene expression in floral organs. In this study, RNA-sequencing (RNA-seq) was performed in two near-isogenic watermelon lines (genic male sterile [GMS] line, DAH3615-MS and male fertile line, DAH3615) to identify the differentially expressed genes (DEGs) related to male sterility.

Results: DEG analysis showed that 1259 genes were significantly associated with male sterility at a FDR *P*-value of < 0.01. Most of these genes were only expressed in the male fertile line. In addition, 11 functional clusters were identified using DAVID functional classification analysis. Of detected genes in RNA-seq analysis, 19 were successfully validated by qRT-PCR.

Conclusions: In this study, we carried out a comprehensive floral transcriptome sequence comparison of a male fertile line and its near-isogenic male sterile line in watermelon. This analysis revealed essential genes responsible for stamen development, including pollen development and pollen tube elongation, and allowed their functional classification. These results provided new information on global mechanisms related to male sterility in watermelon.

Keywords: Watermelon, RNA-seq, Male sterile, Male fertile

Background

Male sterility allows the large-scale hybrid seed production, since it evades the need of emasculation. There are three reported types of male sterility based on the inheritance or origin: genic male sterility (GMS), cytoplasmic male sterility (CMS), and cytoplasmic-genic male sterility (CGMS) [1]. GMS is caused by defective nuclear genes, ordinarily inherited by a single recessive gene; CMS is maternally inherited and caused by the extranuclear genome of mitochondria or chloroplasts; and CGMS is derived by both cytoplasmic and nuclear genes. Male sterility is manifested in different ways based on the nature of plant species: (1) absence of stamens or production of abnormal stamens in bisexual plants or absence of male flowers in dioecious plants; (2) failure to develop normal anther sporogenous tissue; (3) abnormal microsporogenesis in deformed or viable pollen; (4) abnormal pollen maturation or inability to germinate on compatible stigmata; (5) defective anther dehiscence, but viable pollen formation; or (6) barriers other than incompatibility that prevent pollen from reaching the ovule [2].

Although male sterility is regarded as a gift of nature by reducing the labor to produce high-yield hybrids, the mechanism is still unclear in watermelon. To date, RNA-sequencing (RNA-seq) techniques helps to understand the global gene expression patterns and infer the candidate genes relating the male sterility mechanisms. RNA-seq analysis of male sterile floral organs have been reported in

* Correspondence: gplee@cau.ac.kr
†Equal contributors
[1]Department of Integrative Plant Science, Chung-Ang University, Ansung 456-756, Republic of Korea
Full list of author information is available at the end of the article

Rhee *et al. BMC Genomics* (2015) 16:914

Page 2 of 14

many plants such as tomato [3], cotton [4, 5], chili pepper [6], Brassica napus [7, 8], soybean [9] and so on.

Watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) belongs to the family of Cucurbitaceae, which is one of the most economically important crops in the world. It has 11 chromosomes and a haploid genome of approximately 425 Mb ($2n = 2\times = 22$). Watermelon genome sequence was released in 2013 [10], followed by cucumber [11] and melon [12], which makes it possible to apply next-generation sequencing (NGS). To date, some watermelon male sterile mutants, such as glabrous male-sterile (*gms*) [13–15] and male sterile dwarf (*ms-dw*) [16] *ms-1* [17], *ms-2* [18], and *ms-3* [19], have been reported. Watermelon has two flowering patterns; one is monoecious, having male and female flowers on the same plant and the other is andromonoecious, having both male flowers and hermaphrodite flowers on the same plant [20]. Male flower consists of sepal, petal, and stamen and produces pollen grains in the anther.

Anther development is a complex process that has sequential phases to produce male gametes. The anther is composed of several cell types required for pollen development including the epidermis, endothecium, and tapetum that surround the microsporocyte [21, 22]. Anther development is composed of two phases [23]. Phase 1 is defined by cell specification, tissue differentiation, and anther morphology establishment. At the early stages of anther development, archesporial cells are divided into two cell types: the primary parietal cells and the sporogenous cells. The primary parietal cells differentiate to epidermis, endothecium, middle layer, and tapetum (from outer to inner). The sporogenous cells give rise to pollen mother cell and then undergo meiosis to generate the haploid tetrads of microspores. Phase 2 is defined by pollen development, anther enlargement, tapetum/callose tissue degeneration, anther dehiscence, and pollen grain release [23, 24]. Male sterile mutants show disruption of genes related to anther and pollen development [3, 25].

Previous studies identified several genes that are related to male sterility. *SPOROCYTELESS (SPL)/NOZZLE (NZZ)* encodes a protein related to MADS box transcription factors and is required for cell differentiation and division of anther cell walls in *Arabidopsis* [26–29]. *MALE STERILITY 1* (*MS1*) [30–33] and MYB family members [34–37] are important for the tapetum and pollen development *Arabidopsis*. A basic helix-loop-helix (bHLH) transcription factor, *ABORTED MICRO-SPORES* (*AMS*), and *DYSFUNCTIONAL TAPETUM 1* (*DYT1*) are required for tapetum biosynthesis and microspore development in rice [38] and *Arabidopsis* [39]. A deficiency of *AtMYB103* gives rise to complete male sterility, because callose degeneration, tapetal development, and exine formation during anther

development are interrupted [36]. These key genes related to anther development from archesporial tissue development to pollen grain release are responsible for male sterility.

Transcriptome profiling showed similar patterns of functional categories, such as cell wall modification and metabolism, cell cycle regulation, hormonal regulation, energy metabolism, pollen development and cytoskeleton dynamics, between male sterile and fertile lines [4, 40, 41]. In Cucurbitaceae, transcriptome profiling data of floral development are limited to floral sex determination [42, 43]. These studies suggested that ethylene is the most important hormone for regulating female sex determination in cucumber and that auxin, abscisic acid, and brassinosteroids may have an effect on sex determination through their interaction with ethylene. While many transcriptome analyses have been performed in various plant species, only a few have been conducted on fruit development, fruit color, and incompatible interaction with pathogen in watermelon [44–50].

In this study, we focused on the detecting differentially expressed genes (DEGs) related to male sterility. In order to this, RNA-seq was performed with floral buds and mature flowers from two different watermelon lines [genic male sterile (GMS) line, DAH3615-MS, and male fertile line, DAH3615], because two different stages of organs would be reduced to make a false decision for understanding the mechanisms. In addition, significantly detected DEGs across the RNA-seq analysis were technically validated by real-time reverse transcription polymerase chain reaction (qRT-PCR).

Results
Pre-processed RNA-seq data
Raw RNA-seq data were pre-processed using Trimmomatic [51] for generating clean reads. As shown in Table 1, an average of 12,645,328 (94.09 %) surviving reads was retrieved from each data set. All clean reads were mapped on the reference genome of watermelon (cv. 97103) [10]. The overall mapping rate ranged from 58.3 to 62.2 % and the concordant alignment rate from 51 to 54.7 %. A total of 23,440 gene abundances were measured using HTSeq [52] and 19,732 genes remained after filtering out non-expressed genes across all samples.

Identification of DEGs
We compared mRNA expression in floral buds and flowers of DAH3615 and DAH3615-MS samples (Fig. 1) using two different statistical tests [analysis of deviance (ANODEV) and Fisher's exact test], in order to detect sterility-related genes (Fig. 2a and Additional file 1). A total of 649 DEGs were identified from both tests in both organs (floral bud and flower). The highest total

Rhee *et al. BMC Genomics* (2015) 16:914

Page 3 of 14

Table 1 Pre-processed RNA-sequencing data using Trimmomatic

Sample	# of surviving reads	Surviving reads	Overall mapping rate	Alignment rate
Mf_bud	11,834,220	93.55 %	62.2 %	54.7 %
Mf_flower	13,845,808	93.90 %	60.3 %	52.9 %
Ms_bud	12,411,872	94.64 %	61.1 %	53.7 %
Ms_flower	12,489,412	94.25 %	58.3 %	51.0 %

Each sample represents a floral organ in a breeding line; *Mf* male fertile line (DAH3615), *Ms* male sterile line (DAH3615-MS), *bud* floral bud, and *flower* mature flower, as shown in Fig. 1. Including all statistics was calculated based on the paired-end reads

number of DEGs [1259; false discovery rate (FDR) adjusted *P*-value <0.01] was revealed by ANODEV. Out of 1259 DEGs, 352 were exclusively detected by ANODEV. Fisher's exact test revealed 42 and 92 DEGs in the floral bud and flower, respectively, with a lower FDR adjusted *P*-value than that in ANODEV. Overall, ANODEV showed a higher statistical power than Fisher's exact test, and we decided to focus on the DEGs detected from the former test. Hierarchical clustering analysis was performed with $k = 2$ and results are shown in Fig. 2b. DAH3615 and DAH3615-MS were clearly distinguishable, since strong differences were observed between the lines and not between the organs. The relationship of all the significant DEGs was visualized in a heatmap as shown in Fig. 2c. Two regions (Region 1 and 2) were highly different between DAH3615 and DAH3615-MS. In these regions, DEGs showed patterns that were only

identified in DAH3615, regardless the organ. In order to further investigate these results, line plots of the top 10 genes were drawn (Fig. 2d) and confirmed that these genes had similar patterns that were only expressed in DAH3615.

Male sterility related DEGs in relation to biasness of gene expression and chromosomal location

DEG analysis showed that most of the detected genes were exclusively expressed in DAH3615; hence, the trend of gene expression was expected to be skewed towards male fertility. Detected DEGs were characterized using biasness information, calculated as the differences between DAH3615 and DAH3615-MS in floral buds and flowers using the log expression ratios (trimmed mean of M values). As shown in Fig. 3a, approximately 50 % of the genes in DAH3615 and DAH3615-MS were biased
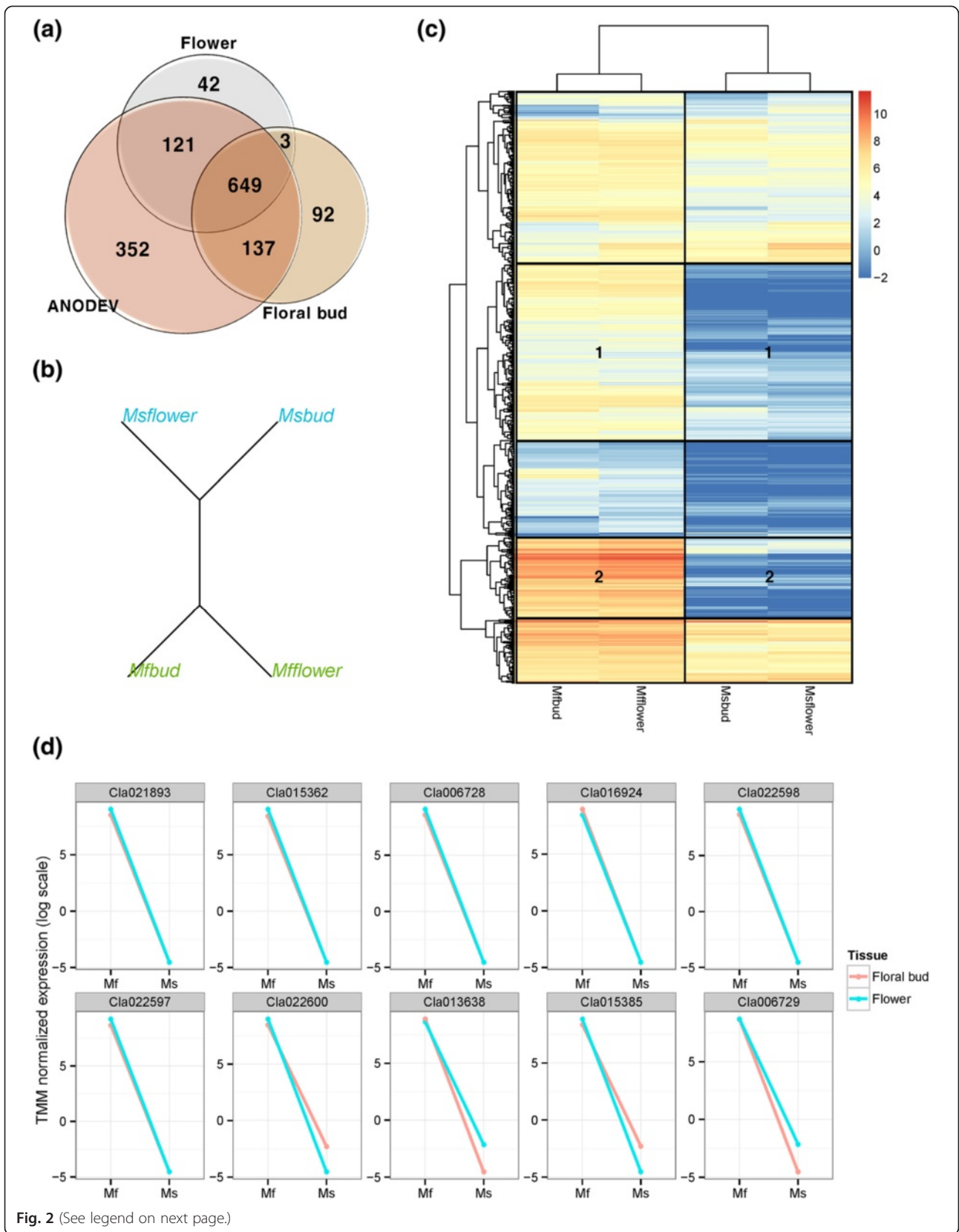

Fig. 1 Morphological differences in floral organs of a male sterile line (DAH3615-MS) and a male fertile line (DAH3615). The flower (**a**) and floral bud (**c**) of the male sterile line, DAH3615-MS, are smaller than those (**b**, **d**) of the male fertile line, DAH3615. The male sterile flower (**a**) shows distinctive defective stamen and absence of pollen grains. The images were taken at 63 d after sowing

Rhee *et al. BMC Genomics* (2015) 16:914

Page 4 of 14

Fig. 2 (See legend on next page.)

Rhee *et al. BMC Genomics* (2015) 16:914

Page 5 of 14

in both organs, when considering gene expressions of the whole genes. However, analysis of the significantly detected DEGs showed that 87.53 % of the genes were DAH3615-biased and only 12.47 % of the genes were DAH3615-MS-biased in both organs. Moreover, the DAH3615-biased genes in the floral bud were also biased in the flower. These results indicate that the gene expression pattern of the detected genes was maintained from the floral bud to the flower.

We also investigated the chromosomal location of the detected sterility-related genes using the watermelon reference genome that is composed of 12 chromosomes
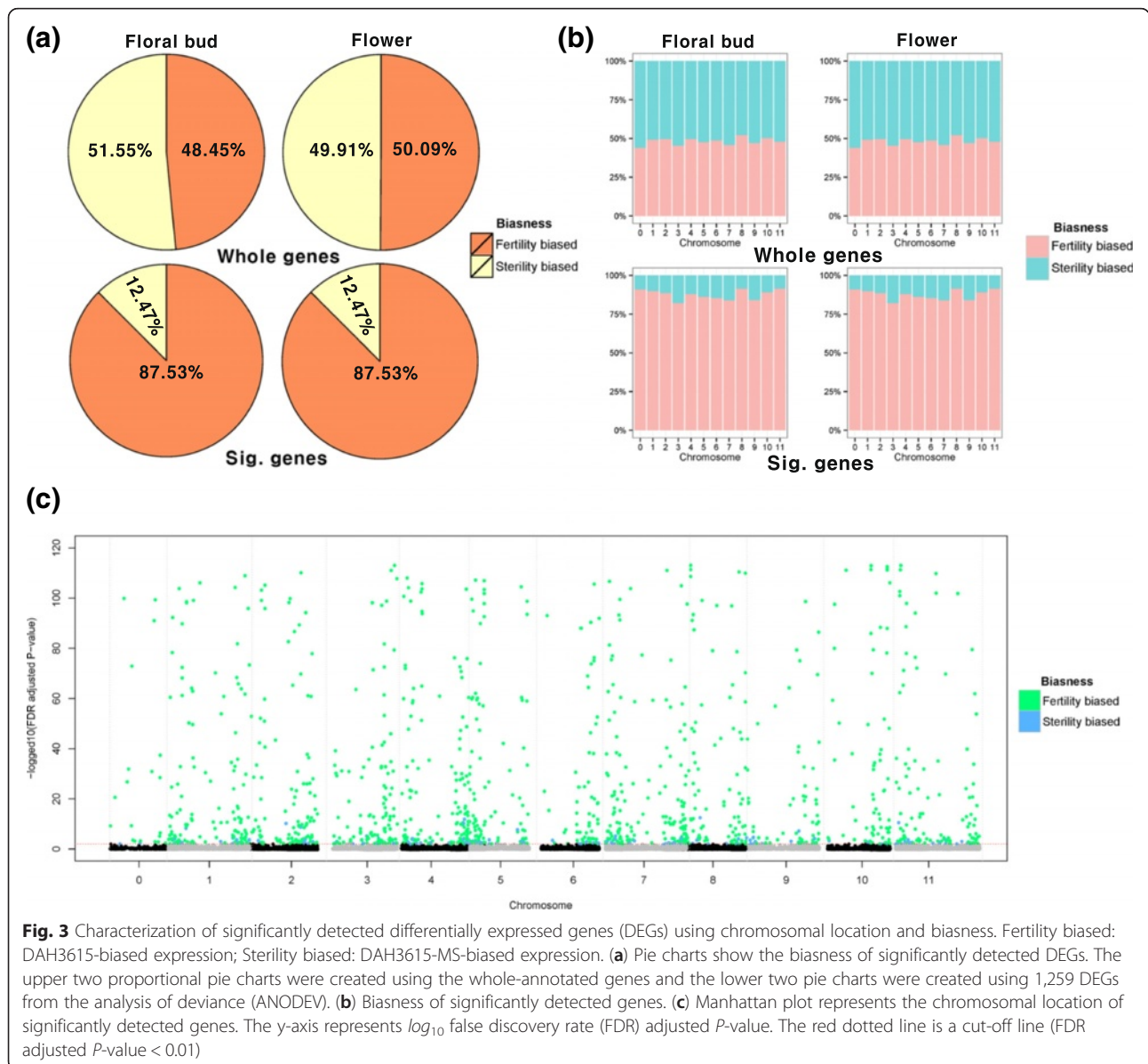

Fig. 3 Characterization of significantly detected differentially expressed genes (DEGs) using chromosomal location and biasness. Fertility biased: DAH3615-biased expression; Sterility biased: DAH3615-MS-biased expression. (**a**) Pie charts show the biasness of significantly detected DEGs. The upper two proportional pie charts were created using the whole-annotated genes and the lower two pie charts were created using 1,259 DEGs from the analysis of deviance (ANODEV). (**b**) Biasness of significantly detected genes. (**c**) Manhattan plot represents the chromosomal location of significantly detected genes. The y-axis represents log_{10} false discovery rate (FDR) adjusted *P*-value. The red dotted line is a cut-off line (FDR adjusted *P*-value < 0.01)

Rhee *et al. BMC Genomics* (2015) 16:914

Page 6 of 14

including chr 0 (undetermined chromosome). The biasness of the significantly detected genes was investigated as shown in Fig. 3b. We observed that the sterility-related genes were not clustered in specific chromosomes, but they were uniformly distributed along the genome. To further investigate these results, the significantly detected genes were visualized in a Manhattan plot as shown in Fig. 3c. The Manhattan plot is usually used in genome-wide association studies (GWAS) to investigate the chromosomal location of single nucleotide polymorphisms (SNPs). It can be applied in RNA-seq analysis because the location of each gene is also known, based on the reference genome. Therefore, the representative location value of each gene can be determined by a Manhattan plot from the annotated mRNA central position on the reference genome. The plot showed a strong signal originated by the DAH3615-biased genes, and that the telomere of each chromosome was enriched to sterility related genes.

Gene-set enrichment analysis for functional annotation of significantly detected sterility-related genes

In order to investigate the biological function of the 1259 detected genes (1105 and 154 genes in DAH3615 and DAH3615-MS biased expression, respectively), gene-set enrichment analysis was performed by DAVID tool using the default stringency level (medium), which is considered [53]. Since only a few studies were performed in watermelon, we used *Arabidopsis* gene annotation as a background database, which is relatively well composed. As a result, DAVID analysis with 1105 DAH3615 biased genes showed that many significantly enriched biological terms including gene ontology (GO) and INTERPRO terms as shown in Table 2. On the other hand, no significant terms were detected in DAVID analysis with the detected 154 DAH3615-MS biased genes. In the Table 2, 11 clusters with their enrichment scores (ES) were reported from the functional annotation clustering analysis. Cluster 1 included copper binding sites, cupredoxin, and multicopper oxidases (type 1, 2, and 3). Cluster 2 and 4 included several lipid-related terms, while cluster 3 several kinesin-related terms and activity-related terms in microtubule. Cluster 5 included signal-related biological terms, while cluster 6 included the highest number of biological terms such as kinase-related terms, ATP-related terms, and nucleotide-binding-related terms. Cluster 7 included diverse development-related terms, while cluster 8 included membrane- and topology-related terms. Cluster 9 and 10 included several binding related terms. Finally, Cluster 11 included DNA- and RNA-related terms. Functional classification analysis using significantly detected

Table 2 DAVID functional classification analysis

Cluster	Terms	ES
Cluster 1	Cupredoxin, copper ion binding, Multicopper oxidase, type 3, Multicopper oxidase, type 2, Multicopper oxidase, type 1	2.25
Cluster 2	Proteoglycan, lipid moiety-binding region: GPI-anchor amidated serine, lipoprotein, propeptide: Removed in mature form, GPI-anchor, cell membrane, glycoprotein	1.35
Cluster 3	Kinesin, motor region, conserved site, KISc, Kinesin, motor region, microtubule associated complex, microtubule motor activity, motor protein, microtubule-based movement, microtubule, motor activity, microtubule-based process, microtubule, microtubule cytoskeleton, cytoskeletal part, cytoskeleton, intracellular non-membrane-bounded organelle, non-membrane-bounded organelle	1.13
Cluster 4	Secondary metabolic process, cellular amino acid derivative metabolic process, phenylpropanoid metabolic process	0.95
Cluster 5	Signal peptide, signal, glycoprotein, lycosylation site:N-linked (GlcNAc…), extracellular region, Secreted	0.94
Cluster 6	protein serine/threonine kinase activity, Serine/threonine protein kinase, active site, serine/threonine-protein kinase, protein amino acid phosphorylation, phosphate metabolic process, phosphorus metabolic process, protein kinase activity, receptor kinase, phosphorylation, Protein kinase, core, protein tyrosine kinase activity, Protein kinase, ATP binding site, nucleotide-binding, atp-binding, ATP binding, adenyl ribonucleotide binding, ribonucleotide binding, purine ribonucleotide binding, Serine/threonine protein kinase-related, adenyl nucleotide binding, purine nucleoside binding, nucleoside binding, purine nucleotide binding, nucleotide binding	0.83
Cluster 7	Developmental growth involved in morphogenesis, unidimensional cell growth, developmental growth, cell morphogenesis, cell growth, regulation of cell size, cellular component morphogenesis, regulation of cellular component size, growth	0.7
Cluster 8	Intrinsic to membrane, topological domain: Cytoplasmic, topological domain: Extracellular, membrane, transmembrane region, transmembrane, integral to membrane	0.36
Cluster 9	Cation binding, ion binding, metal ion binding, transition metal ion binding	0.33
Cluster 10	Transition metal ion binding, oxidation reduction, oxidoreductase, metal-binding	0.09
Cluster 11	Intracellular signaling cascade, transcription regulation, nucleus, Transcription, response to endogenous stimulus, transcription, regulation of transcription, DNA-dependent, regulation of RNA metabolic process, response to organic substance, response to hormone stimulus, dna-binding, transcription factor activity, transcription regulator activity, regulation of transcription, DNA binding	0.08

DAVID functional annotation analysis was performed using 1,105 DAH3615-biased genes derived from the RNA analysis. Twelve clusters were identified including their enrichment score (ES)

Rhee *et al. BMC Genomics* (2015) 16:914

Page 7 of 14

DEGs successfully revealed the associated function of the fertility-related genes.

Technical validation of detected fertility-related genes using qRT-PCR

As the RNA-seq experiment was performed without biological replicates, we have conducted verification experiment with several biological replicated samples in order to confirm the determination of DEGs. Although RNA-seq is more accurate than previous known chip-based transcriptome measuring platforms, such as microarray, biological replication was still important to accurately estimate the expression mean and variance in each group. Many RNA-seq based transcriptome studies have been carried out without biological replications, because RNA-seq is often used for pre-screening, in order to reduce the candidate genes for qRT-PCR. The qRT-PCR methods has been traditionally used for measuring gene expression level, but it can measure gene expression relatively more accurately than high-throughput platforms such as RNA-seq and microarray. In our study, the first RNA-seq analysis was performed, in order to select candidate genes for qRT-PCR. A total of 21 genes were randomly selected from the significantly detected DEGs in the RNA-seq analysis (FDR adjusted P-value < 0.01). We performed qRT-PCR using these 21 genes with 3 biological replicates from each group, namely male fertile floral bud, male fertile flower, male sterile floral bud, and male sterile flower ($n = 12$). Of 21 genes, 19 were successfully measured by qRT-PCR, while 2 genes (calmodulin- and fasciclin-like arabinogalactan protein 17) failed to reach the threshold (Additional file 2). Most of tested genes showed apparent expression in the floral buds and flowers of DAH3615, but no or slight expression in those of DAH3615-MS. Of these genes, S-adenosylmethionine synthase and elongation factor 1-alpha were reduced by approximately 10-fold and 5-fold in the floral buds of DAH3615-MS and those of DAH3615, respectively, but not detectable in DAH3615-MS flowers. DEGs of DAH3615 and DAH3615-MS were consistent with RNA-seq data. In order to visually compare RNA-seq and qRT-PCR results, the quantile normalization method was used for adjusting different scales of gene expression. Based on the normalized gene expression data, relative heatmaps were generated as shown in Fig. 4a. In both RNA-seq and qRT-PCR, clear separation was depended on the breeding line, and the patterns of hierarchical clustering had high similarity. In addition, differences between the biological replications were very small in qRT-PCR as shown in Fig. 4a and b. Student's *t*-test was used to separate the differences between DAH3615 and DAH3615-MS lines, and the results are presented in Table 3. Based on these results, we confirmed that the 19 genes were fertility-related genes and

that RNA-seq was highly consistent. In addition, the male fertility biased gene expression pattern was also technically validated by qRT-PCR.

Discussion

GMS is caused by abnormal gene expression in several developmental stages of stamen development, male gametophyte development, pollen formation, pollen nutrient support, and pollen tube elongation. Previous studies used transcriptional profiling or proteomic analysis to detect DEGs from male sterile and male fertile lines [5, 6, 8, 54]. In this study, we performed RNA-seq and compared transcripts of floral buds and flowers between a male sterile line, DAH3615-MS, and its near-isogenic fertile line, DAH3615, in watermelon. Overall, 1259 genes were detected as DEG between DAH3615 and DAH3615-MS. Of detected genes, 1105 genes were identified as only expressed in DAH3615 and they were annotated into 11 clusters' biological terms by DAVID analysis (Table 2). These clusters included copper-ion binding sites (cluster 1), lipid metabolism terms (cluster 2), cytoskeleton terms (cluster 3), cellular metabolism terms (cluster 4), signaling terms (cluster 5), kinase-related terms (cluster 6), developmental growth terms (cluster 7), membrane location terms (cluster 8), ion-binding terms (cluster 9), oxidation reduction terms (cluster 10), and DNA- and RNA-related terms (cluster 11). For validating RNA-seq data obtained by qRT-PCR, we selected 21 genes from each cluster. Of those, 19 showed consistent results by both platforms (Fig. 3). In this study, DEGs were involved in stamen development, and pollen development, formation, and germination.

A previous study showed that significant genes related to programed cell death (PCD) are expressed spatially and temporally for tapetum degeneration and anther dehiscence [55]. In this study, DEGs for PCD were subtilisin-like protease, disease resistance-responsive (Dirigent-like) family protein, and cysteine proteinase found in cluster 7. Subtilisin-like proteinase is expressed at the tetrad stage in lily[56] and up-regulated in the floral bud of wild-type cabbage [57], while 5B-protein, an anther-specific protease, is associated with the degradation of regulatory proteins in tomato anthers [58]. Subtilisin-like proteinase and dirigent-like protein are involved in lignin biosynthesis and participate in anther dehiscence [59].

Genes related to cell wall loosening were prominently expressed in DAH3615 and included in cluster 7 such as polygalacturonase (PG), expansins, pectate lyase, cellulase (glycosyl hydrolase family 5), endoglucanase, and glucan endo-1 3-beta-glucosidase 5. PG is involved in anther dehiscence [60, 61], pollen mother cell wall degradation [62], pollen intine and exine formation [63], and pollen tube growth [64], while pectate lyase in
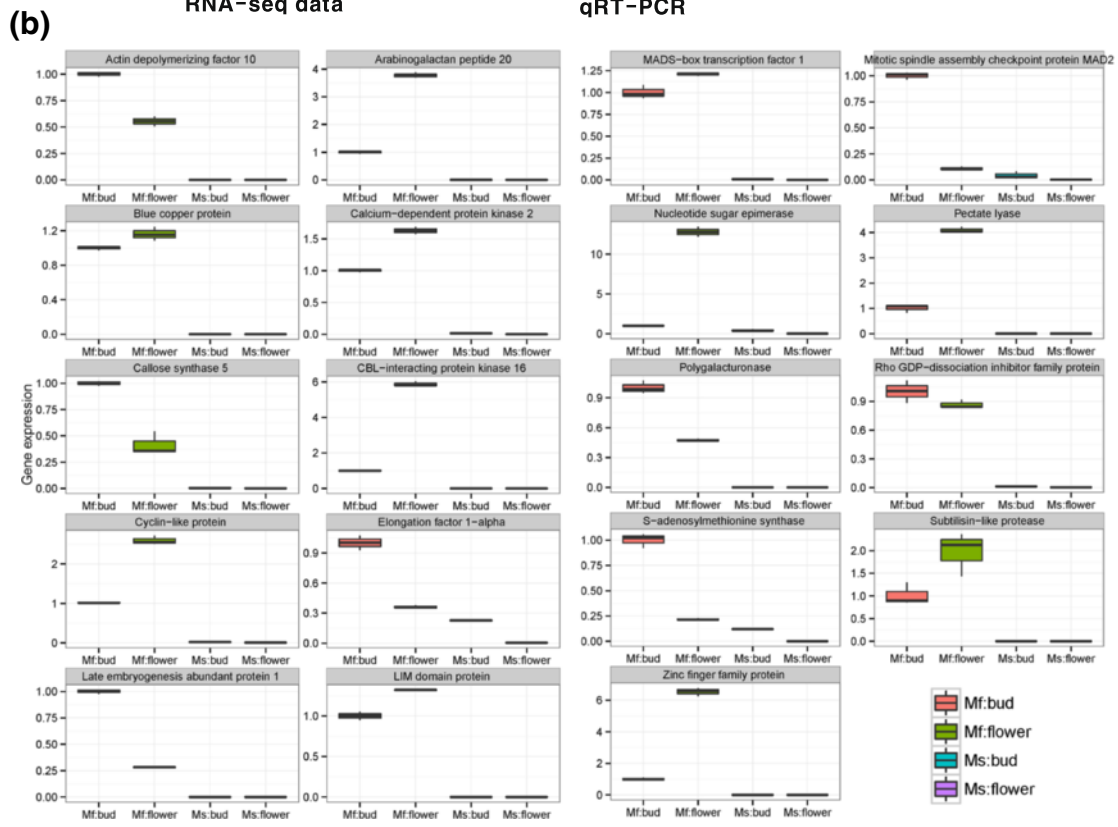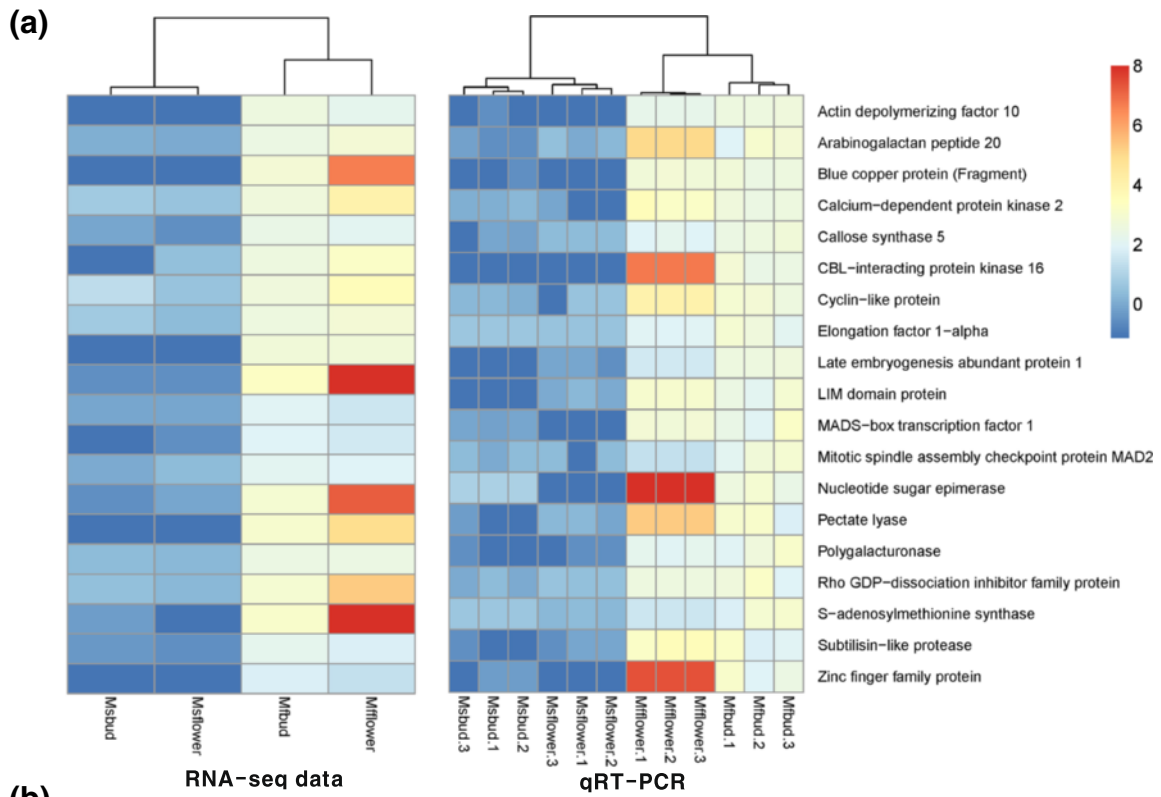
Rhee *et al. BMC Genomics* (2015) 16:914

Page 8 of 14

Fig. 4 (See legend on next page.)

Rhee *et al. BMC Genomics* (2015) 16:914

Page 9 of 14

(See figure on previous page.)
Fig. 4 Technical validation of 19 randomly selected differentially expressed genes (DEGs) detected by RNA-sequencing (RNA-seq) using qRT-PCR. (**a**) Relative heatmaps of RNA-seq and qRT-PCR. Gene expression from the two platforms was normalized by the quantile normalization method. (**b**) Box-plots of 19 DEGs detected by qRT-PCR. The y-axis represents the gene expression level, which is the $-2^{\Delta\Delta Ct}$ value of qRT-PCR compared to the control gene

pollen wall development [65, 66] and pollen tube growth [67, 68]. In this study, male-sterile flowers did not produce any pollen, since many genes related to pollen formation, development, and pollen tube growth were distinguishably expressed in the floral buds and flowers of DAH3615, but not in those of DAH3615-MS.

Callose, beta-1, 3-glucan, is a polysaccharide that separates developing pollen grains, preventing their walls from fusion. The callose wall is temporarily synthesized by callose synthase 5 (Cal5) that requires microspore callose deposition to prevent cell fusion [69, 70] and for this reason Cal5 knockout mutant shows reduced fertility [71]. Callose wall should be timely broken down to release the microspores into the locules. Thus, the mistiming of callose degradation leads to male-sterility. Beta-1, 3-glucanase (callase), a PCD-related gene expressed in the reproductive organs, is secreted from tapetum cells and causes callose degradation [24]. In this study, arabinogalactan proteins (AGPs), such as GPI-anchored proteins, COBRA-like proteins encoding GPI-

anchored proteins, AG peptide 20 and 23, and fasciclin-like AGP 17, were annotated in the DEG data. AGPs participate in cell expansion, division, seed germination, and pollen tube growth and guidance [72].

Calcium-gradient mediated pollen tube growth is one of the best-characterized metabolism systems [73]. In this study, many genes of calcium-gradient mediated pollen tube growth were expressed in DAH3615 such as calcium-dependent protein kinase 25, Rho GDP-dissociation inhibitor family protein, CALCINEURIN B-LIKE (CBL)-interacting protein kinase 16, calmodulin, and calmodulin-like (CaM) proteins. Calcium-binding proteins act as calcium sensors and relay calcium signals. Major calcium sensors have been reported such as calcium-dependent protein kinase, CaM, CBL proteins, and CML protein [73]. In *Arabidopsis*, CIPK19 is expressed specifically in pollen grains and controls pollen polarity [74]. Rho GDP-dissociation inhibitor protein isolated from *Nicotiana tabacum* (Nt-

Table 3 qRT-PCR test results of 19 genes detected by RNA-sequencing

Gene description	Uniport_ID	Floral Bud	Flower
Actin depolymerizing factor 10	B9I489_POPTR	1.35E–04***	2.44E–03**
Arabinogalactan peptide 20	AGP20_ARATH	8.00E–04***	2.10E–04***
Blue copper protein (Fragment)	O82576_MAIZE	1.87E–04***	1.56E–03**
Calcium–dependent protein kinase 2	Q3YAT0_PETIN	1.10E–04***	4.03E–04***
Callose synthase 5	C8C9X3_ARATH	8.66E–05***	2.26E–02*
CBL–interacting protein kinase 16	C4P7W5_VITVI	2.92E–04***	1.92E–04***
Cyclin–like protein	Q6ZIX9_ORYSJ	2.59E–05***	6.36E–04***
Elongation factor 1–alpha	B9SPV9_RICCO	2.75E–03**	7.44E–04***
Late embryogenesis abundant protein 1	LEA1_CICAR	1.41E–04***	2.35E–04***
LIM domain protein	Q306K1_BRANA	8.75E–04***	5.13E–06***
MADS–box transcription factor 1	D1MFS6_HEVBR	1.99E–03**	6.63E–05***
Mitotic spindle assembly checkpoint protein MAD2	D2V3A0_NAEGR	5.34E–06***	8.07E–03**
Nucleotide sugar epimerase	B1WNM2_CYAA5	1.02E–02*	8.12E–04***
Pectate lyase	B6TSP4_MAIZE	7.71E–03**	2.84E–04***
Polygalacturonase	E3VSV7_CUCPE	1.42E–03**	1.96E–04***
Rho GDP–dissociation inhibitor family protein	D7L724_ARALL	4.51E–03**	9.35E–04***
S–adenosylmethionine synthase	B9S0G1_RICCO	2.14E–03**	6.32E–04***
Subtilisin–like protease	Q6WNU4_SOYBN	1.86E–02*	1.93E–02*
Zinc finger family protein	D7MD22_ARALL	3.19E–03**	5.41E–04***

Student's *t*–test was used for separating the differences between a male sterile line (DAH3615–MS) and a male fertile line (DAH3615) in floral buds and flowers.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

Rhee *et al. BMC Genomics* (2015) 16:914

Page 10 of 14

RhoGDI2) is specifically expressed in elongating tobacco pollen tubes and inhibits the formation of transversal actin bands [75].

Defective cytoskeleton dynamics is also one of the factors causing male sterility [76]. In this study, genes related to cytoskeleton were included in cluster 3. Male-sterile lines of wheat and Chinese cabbage have much less actin in the anther than the fertile wild-type plants. Therefore, it has been suggested that low actin levels in the anther are associated with male sterility [77]. LIM domain-containing proteins, which are classified in the novel family of actin bundling proteins, act as transcriptional activators of lignin biosynthesis. Several LIM domain-containing proteins are specifically expressed in the pollen [78]. A previous study showed that the PLIM2c promoter initiated its expression in the pollen during stamen filament elongation and the PLIM2c protein was expressed in the developing pollen grains of *Arabidopsis*. In addition, the PLIM2c-knockout mutant had sterile pollen [79]. Actin depolymerizing factor (ADF), an actin binding protein, affects cytoskeleton architecture dynamics that play a role in the regulation of F-actin filament assembly, which is involved in the polarized tip growth by regulating the actin cytoskeleton in *Arabidopsis* [80, 81].

The final developmental stage of pollen maturation includes a programmed desiccation process for enhancing pollen-geminating efficiency. Therefore, pollen grains have desiccation tolerance during pollen maturation [82]. Dehydrins or late embryogenesis abundant (LEA) proteins protect pollen from desiccation [83]. In this study, the related genes (stress-induced genes) were included in cluster 9.

MADS box transcription factor 1 and zinc-finger family proteins were included in cluster 11. MADS-box transcription factor controls the specification of stamen primordia [84], while several zinc-finger family proteins play important roles in flower development. The latter proteins are especially activated during anther development in petunia [85].

Development of the male reproductive organ includes sporophytic cell division through mitosis and male gametogenesis through meiosis. Therefore, mitotic and meiotic-specific gene destruction during cell division process can lead to male sterility [86]. In this study, cyclin-like protein, kinesin-like protein, and mitotic spindle assembly checkpoint protein MAD2 were differentially expressed in DAH3615 than in DAH3615-MS.

Interestingly three DEGs in our result have scant information in male sterility related previous reports.

Elongation factor-1 alpha (EF-1α) is a well-known housekeeping gene, which is expressed in whole organs including reproductive organs. In vivo assay which visualization of EF-1α using EF-1α-Gus expressed transgenic tobacco plant showed Gus expression in meristems of shoot, root and floral organs [87]. However, our RNA-seq results and qRT-PCR validation in this study were deduced to absence in male sterile flower buds and flowers. It has a previously report that proteome analysis using *Zea mays* ms8 mutant anther have shown that EF-1α as a differentially expressed protein [88]. EF-1α is a multi-functional protein which play a role such as elongation of translation, cytoskeleton regulation, and signal transduction. Additionally, it is participated in interaction of calmodulin [87], substrate of calcium dependent protein kinases [89] and Rho associated kinase [90], and putative regulator of DNA replication/ repair protein [91]. Most of these EF-1α functions are deduced in our DEG results. Therefore, EF-1α is crucially required for the successful development of male reproductive organ of watermelon.

Blue copper protein was the most significant expressed gene that included in cluster 1, which was the most enriched cluster. Plantacyanins, a sub-subfamily of phytocyanins that belongs to type 1 'blue copper protein', are secreted from pistil and guide pollen and pollen tube to interact with the pistil. Although there is limited information whether blue copper protein is associated with male sterility, transcriptome analysis showed that blue copper protein is down regulated in CMS cotton line [4]. So, blue copper protein may be involved in determination of the male sterility in watermelon, but it would be required for progress studies.

In cucumber, cDNA subtractive hybridization using gynoecious and hermaphrodite floral buds and flowers has shown that nucleotide sugar epimerase may be involve in stamen development [92]. Nucleotide sugar epimerase showed low expression in DAH3615-MS. Even though the gene has little information related to floral organ development or male sterility, the result of our RNA-seq data provides possibility of the gene would be a putative novel gene which implicates for inducing male sterility by affect cell expansion, especially of stamen primordia.

Conclusions

In conclusion, we successfully detected 1259 DEGs (FDR adjusted *P*-value < 0.01) by comparing floral buds and flowers of a male-sterile line (DAH3615-MS) and a male-fertile line (DAH3615) using two-way ANODEV. To our knowledge, this is the first report on transcriptional profiling in watermelon for identifying DEGs related to male sterility and investigating expression patterns in male-sterile and fertile reproductive organs. We also anticipated some functions, such as stamen development, pollen formation, and pollen tube elongation, of essential genes in the fertile male flower that referred in previous reports. Our results provide new information

Rhee *et al. BMC Genomics* (2015) 16:914

Page 11 of 14

on global mechanisms related to male sterility in watermelon.

Methods

Plant materials

Floral buds and flowers were harvested from progenies of an 1:1 segregating watermelon population between a GMS line (DAH3615-MS, *msms*) derived from the *ms-1* Chinese male sterile line [17] and its fertile near-isogenic line (DAH3615, *Msms*), which was selected from a backcross program. The plants were grown in a greenhouse during the spring of 2014. Floral buds of DAH3615-MS and DAH3615 were collected from the 22nd–23rd node (3–4 mm and 7–8 mm in length, respectively) 1 d before flowering, while mature flowers were collected the first day of flowering from the 16–17th node. The samples were directly frozen in liquid nitrogen and then stored at –80 °C until RNA extraction.

RNA-seq experiment

To generate RNA-seq reads, we followed Illumina sequencing protocol. Total RNA was isolated from floral buds and flowers of DAH3615-MS and DAH3615 using Tri-Reagent® (MRC, OH, USA) added 1 % mercaptoethanol according to the manufacturer's manual. Quantification was performed using Nanodrop spectrophotometer (Thermo Scientific, USA), while quality assessment was performed with RNA 6000 Nano assay kit (Agilent, USA) using Bioanalyser 2100 (Agilent, USA). NGS sequencing libraries were generated from 1 µg of total RNA using Truseq RNA Sample Prep Kit (Illumina, USA) according to the manufacturer's protocol. The resulting libraries were then paired-end sequenced (2x101bp) with the Illumina HiSeq™ 2000 system. Finally, FastQC V0.10.1 with ASCII Q-score offset 33 was used to check sequencing quality. More detailed QC results can be shown in Additional file 3.

Reference genome and RNA-seq data pre-processing

We used watermelon reference genome (cv. 97103) version 1 from the Curcurbit Genomics Database. Prior to aligning, we employed Trimmomatic [51] to generate clean reads, removing Illumina adapter sequences. Clean reads were mapped on watermelon genome using Bowtie2 [93] included in Tophat2 [94]. Finally, we used HTSeq [52] along with watermelon gene annotation file from the Curcurbit Genomics Database for measuring mRNA expression. In this study, we focused only on the quantification of known mRNA sequences. Therefore, unknown proteins or isoforms were not considered in this study.

Statistical methods for detecting male sterility related genes

Two types of statistical tests were employed for detecting DEGs, Fisher's exact test and ANODEV. First, the Fisher's exact test is widely used for RNA-seq analysis without replications to compare gene expressions between two groups. Employing this statistical test is easiest way for detecting DEGs in each organs, under our study design. This experiment had a 2×2 factorial arrangement, since the generated RNA-seq data had two factors with two levels each, breeding lines (male sterile line: DAH3615-MS and male fertile line: DAH3615) and organ information (floral bud and mature flower). In this situation, 2-way ANODEV model can be employed for detecting DEGs considering experimental design as follows (1):

$$Expression_{ij} = \mu + breeding\ line_i + organ_j \qquad (1)$$

where i = {"male fertile line (DAH3615)" and "male sterile line (DAH3615_MS)"}
and j = {"floral bud", "flower"}

From the model, statistical test was performed about group effect (comparison between male sterile DAH3615-MS and male fertile DAH3615) for all genes, respectively, when organ effect was considered. Both statistical tests were performed in edgeR for detecting DEGs [95], which results were attached in Additional file 1. A gene was considered significant at FDR adjusted *P*-value of < 0.01.

qRT-PCR experiment for technically validating detected DEGs from RNA-seq analysis

In order to technically validate the results of RNA-seq, we performed qRT-PCR Eco™ Real-Time PCR system (Illumina, CA, USA). The experiment was performed using 3 biological replicates from each group, namely male fertile floral bud, male fertile flower, male sterile floral bud, and male sterile flower (n = 12). Of the detected DEGs from RNA-seq analysis, 21 genes were selected using the results from previous studies on male sterility-related genes. Representative samples of two developmental stages, floral buds and flowers, from each of DAH3615-MS and DAH3615 were used to validate RNA-seq results by qRT-PCR. PCR reaction mixture was consisted of 1 µl cDNA, 10 µl pre-mix, 1 µl evergreen fluorescence dye (SolGent, Korea), and 500 nM of each primer (except for 18S reran that 250nM of each primer were used). Cycling conditions were as follows: 95 °C for 12 min, 40 cycles at 95 °C for 10 s, and 60 °C for 30 s. Watermelon 18S reran was used as an internal control for normalizing mRNA and 20 candidate genes were selected and performed by qRT-PCR for evaluating DEGs. Primers were designed by Primer3 tool [96] using

Rhee *et al. BMC Genomics* (2015) 16:914

Page 12 of 14

gene sequences from the Curcurbit Genomics Database and presented in (Additional file 4). The relative expression levels of selected genes were calculated using the 2$^{-\Delta\Delta CT}$ transformation method (Additional file 2). The expression level of each transcript from the floral buds of DAH3615-MS was used as a calibrator. All reactions were performed in triplicate. The statistical analysis for qRT-PCR, which can be shown in Additional file 4.

Availability of data and materials

Our generated RNA-seq raw data supporting the results of this article is available in the Gene Expression Omnibus (GEO) repository. The accession number is GSE69073 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE69073).

Additional files

> **Additional file 1: RNA-seq analysis results.** The spreadsheet file contains statistical analysis results including FDR adjusted *P*-values of the 2-way ANODEV and Fisher's exact test as well as gene annotation. (XLSX 1795 kb)
>
> **Additional file 2: Gene expression in the qRT-PCR.** Normalized gene expression is included in the file. (XLSX 11 kb)
>
> **Additional file 3: Quality control result for RNA-seq experiment.** The file contains QC result of the RNA-seq experiment. (XLSX 8 kb)
>
> **Additional file 4: Detailed information for the qRT-PCR experiment.** Statistical analysis results and primer information, which is included in this file. (DOCX 21 kb)

Abbreviations

ADF: Actin depolymerizing factor; AGP: Arabinogalactan proteins; ANODEV: Analysis of deviance; CaM: Calmodulin; CBL: CALCINEURIN B-LIKE; CGMS: Cytoplasmic-genic male-sterility; CMS: Cytoplasmic male sterility; DEGs: Differentially expressed genes; EF-1α: Elongation factor-1 alpha; ES: Enrichment scores; FDR: False discovery rate; GMS: Genic male sterility; GO: Gene ontology; GWAS: Genome-wide association studies; HTSeq: High throughput sequencing; LEA: Late embryogenesis abundant; NGS: Next-generation sequencing; PCD: Programmed cell death; PG: Polygalacturonase; qRT-PCR: Quantitative reverse transcription polymerase chain reaction; RhoGDI: Rho GDP-dissociation inhibitor protein; RNS-seq: RNA-sequencing; SNPs: Single nucleotide polymorphisms.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SL performed RNA isolation for RNA-seq and conducted cDNA synthesis and qRT-PCR validation. MS performed global bioinformatics analysis. MS and SL wrote the manuscript. YJ helped with RNA isolation and generation of plant materials. SC helped with revised the manuscript and supervised the in-silico analysis. GPL revised the manuscript and supervised the experiment. All authors read and edited the manuscript.

Author details

[1]Department of Integrative Plant Science, Chung-Ang University, Ansung 456-756, Republic of Korea. [2]Interdisciplinary Program in Bioinformatics, Seoul National University, Kwan-ak St. 599, Kwan-ak Gu, Seoul 151-741, Republic of Korea. [3]CHO&KIM genomics, Main Bldg. #514, SNU Research Park, Seoul National University Mt.4-2, NakSeoungDae, Gwanakgu, Seoul 151-919, Republic of Korea.

References

1. Vedel F, Pla M, Vitart V, Gutierres S, Chetrit P, Depaepe R. Molecular-basis of nuclear and cytoplasmic male-sterility in higher-plants. Plant Physiol Bioch. 1994;32(5):601–18.
2. Anwar A, Mehdi M. Role of genetically engineered system of male sterility in hybrid production of vegetables. J Phytology. 2009;1(6):448–60.
3. Jeong HJ, Kang JH, Zhao M, Kwon JK, Choi HS, Bae JH, et al. Tomato Male sterile 1035 is essential for pollen development and meiosis in anthers. J Exp Bot. 2014;65(22):6693–709.
4. Suzuki H, Rodriguez-Uribe L, Xu JN, Zhang JF. Transcriptome analysis of cytoplasmic male sterility and restoration in CMS-D8 cotton. Plant Cell Rep. 2013;32(10):1531–42.
5. Wei MM, Song MZ, Fan SL, Yu SX. Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. BMC Genomics. 2013;14:97.
6. Liu C, Ma N, Wang PY, Fu N, Shen HL. 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. 2013;8(6):e65209.
7. Yan XH, Dong CH, Yu JY, Liu WH, Jiang CH, Liu J, et al. Transcriptome profile analysis of young floral buds of fertile and sterile plants from the self-pollinated offspring of the hybrid between novel restorer line NR1 and Nsa CMS line in Brassica napus. BMC Genomics. 2013;14:26.
8. An H, Yang ZH, Yi B, Wen J, Shen JX, Tu JX, et al. Comparative transcript profiling of the fertile and sterile flower buds of pol CMS in B. napus. BMC Genomics. 2014;15:258.
9. Li JJ, Han SH, Ding XL, He TT, Dai JY, Yang SP, et al. Comparative transcriptome analysis between the cytoplasmic male sterile line njcms1a and its maintainer njcms1b in soybean (Glycine max (L.) Merr.). PLoS One. 2015;10(5):e0126771.
10. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat Genet. 2013;45(1):51–8.
11. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, Cucumis sativus L. Nat Genet. 2009;41(12):1275–81.
12. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, et al. The genome of melon (Cucumis melo L.). Proc Natl Acad Sci U S A. 2012;109(29):11872–7.
13. Ray DT, Sherman JD. Desynaptic chromosome behavior of the gms mutant in watermelon. J Hered. 1988;79(5):397–9.
14. Watts V. A marked male-sterile mutant in watermelon. Proc Amer Soc Hort Sci:. 1962;1962:498–505.
15. Watts V. Development of disease resistance and seed production in watermelon stocks carrying msg gene. Proc Am Soc Hortic Sci. 1967;91:579. Amer Soc Horticultural Science 701 North Saint Asaph Street, Alexandria, Va 22314–1998.
16. Huang HX, Zhang XQ, Wei ZC, Li QH, Li X. Inheritance of male-sterility and dwarfism in watermelon [Citrullus lanatus (Thunb.) Matsum. and Nakai]. Sci Hortic-Amsterdam. 1998;74(3):175–81.
17. Zhang X, Wang M. A genetic male-sterile (ms) watermelon from China. Rep-Cucurbit Genet Coop. 1990;13:45–6.
18. Dyutin K, Sokolov S. Spontaneous mutant of watermelon with male sterility. Tsitol Genet. 1990;24(2):56–7.
19. Bang H, King SR, Liu W. A new male sterile mutant identified in watermelon with multiple unique morphological features. Rep-Cucurbit Genet Coop. 2005;28:47.
20. Rudich J, Zamski E. Citrullus lanatus. Handb Flower. 1985;2:272–4.
21. Smyth DR, Bowman JL, Meyerowitz EM. Early flower development in arabidopsis. Plant Cell. 1990;2(8):755–67.
22. Wilson ZA, Zhang DB. From Arabidopsis to rice: Pathways in pollen development. J Exp Bot. 2009;60(5):1479–92.

Rhee *et al. BMC Genomics* (2015) 16:914

Page 13 of 14

23. Goldberg RB, Beals TP, Sanders PM. Anther development: Basic principles and practical applications. Plant Cell. 1993;5(10):1217–29.

24. Scott RJ, Spielman M, Dickinson HG. Stamen structure and function. Plant Cell. 2004;16:S46–60.

25. Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu YC, Lee PY, et al. Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex Plant Reprod. 1999;11(6):297–322.

26. Schiefthaler U, Balasubramanian S, Sieber P, Chevalier D, Wisman E, Schneitz K. Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 1999;96(20):11664–9.

27. Ito T, Wellmer F, Yu H, Das P, Ito N, Alves-Ferreira M, et al. The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCYTELESS. Nature. 2004;430(6997):356–60.

28. Liu XD, Huang J, Parameswaran S, Ito T, Seubert B, Auer M, et al. The SPOROCYTELESS/NOZZLE gene is involved in controlling stamen identity in Arabidopsis. Plant Physiol. 2009;151(3):1401–11.

29. Yuan L, Sundaresan V. Spore formation in plants: Sporocyteless and more. Cell Res. 2015;25(1):7–8.

30. Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ. The Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J. 2001;28(1):27–39.

31. Ito T, Shinozaki K. The MALE STERILITY1 gene of Arabidopsis, encoding a nuclear protein with a PHD-finger motif, is expressed in tapetal cells and is required for pollen maturation. Plant Cell Physiol. 2002;43(11):1285–92.

32. Ito T, Nagata N, Yoshiba Y, Ohme-Takagi M, Ma H, Shinozaki K. Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell. 2007;19(11):3549–62.

33. Yang C, Vizcay-Barrena G, Conner K, Wilson ZA. MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell. 2007;19(11):3530–48.

34. Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW. AtMYB32 is required for normal pollen development in Arabidopsis thaliana. Plant J. 2004;40(6):979–95.

35. Millar AA, Gubler F. The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are MicroRNA-regulated genes that redundantly facilitate anther development. Plant Cell. 2005;17(3):705–21.

36. Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Li H, et al. Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J. 2007;52(3):528–38.

37. Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, et al. Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. 2008;55(2):266–77.

38. Jung KH, Han MJ, Lee YS, Kim YW, Hwang IW, Kim MJ, et al. Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. Plant Cell. 2005;17(10):2705–22.

39. Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H. Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development. 2006;133(16):3085–95.

40. Honys D, Twell D. Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol. 2004;5(11):R85.

41. Pina C, Pinto F, Feijo JA, Becker JD. Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol. 2005;138(2):744–56.

42. Guo SG, Zheng Y, Joung JG, Liu SQ, Zhang ZH, Crasta OR, et al. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. BMC Genomics. 2010;11:569.

43. Wu T, Qin ZW, Zhou XY, Feng Z, Du YL. Transcriptome profile analysis of floral sex determination in cucumber. J Plant Physiol. 2010;167(11):905–13.

44. Guo S, Liu J, Zheng Y, Huang M, Zhang H, Gong G, et al. Characterization of transcriptome dynamics during watermelon fruit development: sequencing, assembly, annotation and gene expression profiles. BMC Genomics. 2011;12:454.

45. Grassi S, Piro G, Lee JM, Zheng Y, Fei Z, Dalessandro G, et al. Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. BMC Genomics. 2013;14:781.

46. Fan M, Huang Y, Zhong Y, Kong Q, Xie J, Niu M, et al. Comparative transcriptome profiling of potassium starvation responsiveness in two contrasting watermelon genotypes. Planta. 2014;239(2):397–410.

47. Wang Z, Hu H, Goertzen LR, McElroy JS, Dane F. Analysis of the Citrullus colocynthis transcriptome during water deficit stress. PLoS One. 2014;9(8):e104657.

48. Lu GY, Guo SG, Zhang HY, Geng LH, Song FM, Fei ZJ, et al. Transcriptional profiling of watermelon during its incompatible interaction with Fusarium oxysporum f. sp niveum. Eur J Plant Pathol. 2011;131(4):585–601.

49. Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell R, et al. Differential gene expression and alternative splicing between diploid and tetraploid watermelon. J experiment botany. 2015;66(5):1369–1385.

50. Jagadeeswaran G, Nimmakayala P, Zheng Y, Gowdu K, Reddy UK, Sunkar R. Characterization of the small RNA component of leaves and fruits from four different cucurbit species. BMC Genomics. 2012;13:329.

51. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20. btu170.

52. Anders S, Pyl PT, Huber W. HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics. 2014;31(2):166–9. btu638.

53. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2008;4(1):44–57.

54. Zheng BB, Wu XM, Ge XX, Deng XX, Grosser JW, Guo WW. Comparative transcript profiling of a male sterile cybrid pummelo and its fertile type revealed altered gene expression related to flower development. PLoS One. 2012;7(8):e43758.

55. Varnier AL, Mazeyrat-Gourbeyre F, Sangwan RS, Clement C. Programmed cell death progressively models the development of anther sporophytic tissues from the tapetum and is triggered in pollen grains during maturation. J Struct Biol. 2005;152(2):118–28.

56. Taylor AA, Horsch A, Rzepczyk A, Hasenkampf CA, Riggs CD. Maturation and secretion of a serine proteinase is associated with events of late microsporogenesis. Plant J Cell Mol Biol. 1997;12(6):1261–71.

57. Lou P, Kang JG, Zhang GY, Bonnema G, Fang ZY, Wang XW. Transcript profiling of a dominant male sterile mutant (Ms-cd1) in cabbage during flower bud development. Plant Sci. 2007;172(1):111–9.

58. Sheoran IS, Ross AR, Olson DJ, Sawhney VK. Differential expression of proteins in the wild type and 7B-1 male-sterile mutant anthers of tomato (Solanum lycopersicum): a proteomic analysis. J Proteomics. 2009;71(6):624–36.

59. Jagadish SVK, Muthurajan R, Oane R, Wheeler TR, Heuer S, Bennett J, et al. Physiological and proteomic approaches to address heat tolerance during anthesis in rice (Oryza sativa L.). J Exp Bot. 2010;61(1):143–56.

60. Gorguet B, Schipper D, van Lammeren A, Visser RGF, van Heusden AW. ps-2, the gene responsible for functional sterility in tomato, due to non-dehiscent anthers, is the result of a mutation in a novel polygalacturonase gene. Theor Appl Genet. 2009;118(6):1199–209.

61. Sander L, Child R, Ulvskov P, Albrechtsen M, Borkhardt B. Analysis of a dehiscence zone endo-polygalacturonase in oilseed rape (Brassica napus) and Arabidopsis thaliana: evidence for roles in cell separation in dehiscence and abscission zones, and in stylar tissues during pollen tube growth. Plant Mol Biol. 2001;46(4):469–79.

62. Rhee SY, Osborne E, Poindexter PD, Somerville CR. Microspore separation in the quartet 3 mutants of Arabidopsis is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. Plant Physiol. 2003;133(3):1170–80.

63. Lyu M, Liang Y, Yu Y, Ma Z, Song L, Yue X, et al. Identification and expression analysis of BoMF25, a novel polygalacturonase gene involved in pollen development of Brassica oleracea. Plant Reprod. 2015;28(2):121–32.

64. Huang L, Cao J, Ye W, Liu T, Jiang L, Ye Y. Transcriptional differences between the male-sterile mutant bcms and wild-type Brassica campestris ssp. chinensis reveal genes related to pollen development. Plant Biol. 2008;10(3):342–55.

65. Jiang JJ, Yao LN, Yu YJ, Lv ML, Miao Y, Cao JS. PECTATE LYASE-LIKE10 is associated with pollen wall development in Brassica campestris. J Integr Plant Biol. 2014;56(11):1095–105.

66. Jiang JJ, Yao LN, Yu YJ, Liang Y, Jiang JX, Ye NH, et al. PECTATE LYASE-LIKE 9 from Brassica campestris is associated with intine formation. Plant Sci. 2014;229:66–75.

Rhee *et al. BMC Genomics* (2015) 16:914

Page 14 of 14

67. Kulikauskas R, McCormick S. Identification of the tobacco and Arabidopsis homologues of the pollen-expressed LAT59 gene of tomato. Plant Mol Biol. 1997;34(5):809–14.

68. Wu YZ, Qiu X, Du S, Erickson L. PO149, a new member of pollen pectate lyase-like gene family from alfalfa. Plant Mol Biol. 1996;32(6):1037–42.

69. Chen XY, Liu L, Lee E, Han X, Rim Y, Chu H, et al. The Arabidopsis Callose Synthase Gene GSL8 Is Required for Cytokinesis and Cell Patterning. Plant Physiol. 2009;150(1):105–13.

70. Enns LC, Kanaoka MM, Torii KU, Comai L, Okada K, Cleland RE. Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. Plant Mol Biol. 2005;58(3):333–49.

71. Dong XY, Hong ZL, Sivaramakrishnan M, Mahfouz M, Verma DPS. Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in Arabidopsis. Plant J. 2005;42(3):315–28.

72. Bosch M, Cheung AY, Hepler PK. Pectin methylesterase, a regulator of pollen tube growth. Plant Physiol. 2005;138(3):1334–46.

73. Steinhorst L, Kudla J. Calcium - a central regulator of pollen germination and tube growth. Bba-Mol Cell Res. 2013;1833(7):1573–81.

74. Zhou L, Lan W, Chen B, Fang W, Luan S. A calcium sensor-regulated protein kinase, CALCINEURIN B-LIKE PROTEIN-INTERACTING PROTEIN KINASE19, is required for pollen tube growth and polarity. Plant Physiol. 2015;167(4):1351–60.

75. Fu Y, Wu G, Yang ZB. Rop GTPase-dependent dynamics of tip-localized F-actin controls tip growth in pollen tubes. J Cell Biol. 2001;152(5):1019–32.

76. Xu C, Liu Z, Zhang L, Zhao C, Yuan S, Zhang F. Organization of actin cytoskeleton during meiosis I in a wheat thermo-sensitive genic male sterile line. Protoplasma. 2013;250(1):415–22.

77. Xie CT, Wei DM, Tian HQ. [Advances in cell biological researches on male sterility of higher plants]. J Plant Physiol Mol Biol. 2006;32(1):17–23.

78. Baltz R, Evrard JL, Domon C, Steinmetz A. A lim motif is present in a pollen-specific protein. Plant Cell. 1992;4(12):1465–6.

79. Ye J, Xu M. Actin bundler PLIM2s are involved in the regulation of pollen development and tube growth in Arabidopsis. J Plant Physiol. 2012;169(5):516–22.

80. Bou Daher F, van Oostende C, Geitmann A. Spatial and temporal expression of actin depolymerizing factors ADF7 and ADF10 during male gametophyte development in Arabidopsis thaliana. Plant Cell Physiol. 2011;52(7):1177–92.

81. Maciver SK, Hussey PJ: The ADF/cofilin family: actin-remodeling proteins. Genome Biol 2002, 3(5).

82. Wang CS, Hsu SW, Hsu YF. new insights into desiccation-associated gene regulation by Lilium longiflorum ASR during pollen maturation and in transgenic Arabidopsis. Int Rev Cel Mol Bio. 2013;301:37–94.

83. Mogami N, Shiota H, Tanaka I. The identification of a pollen-specific LEA-like protein in Lilium longiflorum. Plant Cell Environ. 2002;25(5):653–63.

84. Coen ES, Meyerowitz EM. The war of the whorls - genetic interactions controlling flower development. Nature. 1991;353(6339):31–7.

85. Kobayashi A, Sakamoto A, Kubo K, Rybka Z, Kanno Y, Takatsuji H. Seven zinc-finger transcription factors are expressed sequentially during the development of anthers in petunia. Plant J. 1998;13(4):571–6.

86. Chen CB, Marcus A, Li WX, Hu Y, Calzada JPV, Grossniklaus U, et al. The Arabidopsis ATK1 gene is required for spindle morphogenesis in male meiosis. Development. 2002;129(10):2401–9.

87. Ursin VM, Irvine JM, Hiatt WR, Shewmaker CK. Developmental analysis of elongation factor-1-alpha expression in transgenic tobacco. Plant Cell. 1991;3(6):583–91.

88. Wang DX, Oses-Prieto JA, Li KH, Fernandes JF, Burlingame AL, Walbot V. The male sterile 8 mutation of maize disrupts the temporal progression of the transcriptome and results in the mis-regulation of metabolic functions. Plant J. 2010;63(6):939–51.

89. Yang W, Boss WF. Regulation of phosphatidylinositol 4-kinase by the protein activator pik-A49 - activation Requires phosphorylation of pik-A49. J Biol Chem. 1994;269(5):3852–7.

90. Izawa T, Fukata Y, Kimura T, Iwamatsu A, Dohi K, Kaibuchi K. Elongation factor-1 alpha is a novel substrate of Rho-associated kinase. Biochem Bioph Res Co. 2000;278(1):72–8.

91. Toueille M, Saint-Jean B, Castroviejo M, Benedetto JP. The elongation factor 1A: A novel regulator in the DNA replication/repair protein network in wheat cells? Plant Physiol Bioch. 2007;45(2):113–8.

92. Terefe D, Tatlioglu T. Isolation of a partial sequence of a putative nucleotide sugar epimerase, which may involve in stamen development in cucumber (Cucumis sativus L.). Theor Appl Genet. 2005;111(7):1300–7.

93. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

94. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

95. Chen Y, Lun AT, Smyth GK. Differential expression analysis of complex rna-seq experiments using edgeR. In: Datta S, Nettleton DS, editors. Statistical analysis of next generation sequencing data. New York: Springer; 2014. p. 51–74.

96. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115–5.