Comparative transcriptome profiling of the fertile and sterile flower buds of a dominant genic male sterile line in sesame (Sesamum indicum L.)

Hongyan Liu, Mingpu Tan, Haijuan Yu, Liang Li, Fang Zhou, Minmin Yang, Ting Zhou and Yingzhong Zhao

Abstract

Background: Sesame (Sesamum indicum L.) is a globally important oilseed crop with highly-valued oil. Strong hybrid vigor is frequently observed within this crop, which can be exploited by the means of genic male sterility (GMS). We have previously developed a dominant GMS (DGMS) line W1098A that has great potential for the breeding of F1 hybrids. Although it has been genetically and anatomically characterized, the underlying molecular mechanism for male sterility remains unclear and therefore limits the full utilization of such GMS line. In this study, RNA-seq based transcriptome profiling was carried out in two near-isogenic DGMS lines (W1098A and its fertile counterpart, W1098B) to identify differentially expressed genes (DEGs) related to male sterility.

Results: A total of 1,502 significant DEGs were detected, among which 751 were up-regulated and 751 were down-regulated in sterile flower buds. A number of DEGs were implicated in both ethylene and JA synthesis & signaling pathway; the expression of which were either up- or down-regulated in the sterile buds, respectively. Moreover, the majority of NAC and WRKY transcription factors implicated from the DEGs were up-regulated in sterile buds. By querying the Plant Male Reproduction Database, 49 sesame homologous genes were obtained; several of these encode transcription factors (bHLH089, MYB99, and AMS) that showed reduced expression in sterile buds, thus implying the possible role in specifying or determining tapetal fate and development. The predicted effect of allelic variants on the function of their corresponding DEGs highlighted several Insertions/Deletions (InDels), which might be responsible for the phenotype of sterility/fertility in DGMS lines.

Conclusion: The present comparative transcriptome study suggested that both hormone signaling pathway and transcription factors control the male sterility of DGMS in sesame. The results also revealed that several InDels located in DEGs prone to cause loss of function, which might contribute to male sterility. These findings provide valuable genomic resources for a deeper insight into the molecular mechanism underlying DGMS.

Keywords: Sesame, Dominant genic male sterile, Transcriptome, Differentially expressed genes,
Background

Sesame (Sesamum indicum L.) is a globally important and ancient oilseed crop mainly consumed for high-quality oil [1, 2]. It has the highest oil content among the cultivated oil crops and is rich in natural antioxidants like sesamin and sesamol, which are known by their specific antihypertensive effects and anti-oxidative activity [3–5]. Although important, the seed yield of sesame is unstable and relatively low compared with rapeseed, peanut and soybean. Therefore, great efforts should be made to improve the seed yield of sesame.

Heterosis utilization is the most promising approach for yield improvement, since very strong hybrid vigor (>15 %) has been observed within this crop [6]. Heterosis can be effectively exploited either by cytoplasmic male sterility (CMS) or genic male sterility (GMS). So far, only recessive GMS has been successfully applied to the production of sesame F1 hybrids. However, this method might be constrained by certain drawbacks such as environmental sensitivity, incomplete sterility, and the timely removal of 50 % male-fertile plantlets from two-type lines for hybrid seeds production [7]. Recently, we have developed a novel dominant GMS line (DGMS) by crossing the wild species S. mulayanum L. (2n = 26) plants with the cultivated species S. indicum L. (2n = 26), which has great potential for the breeding of hybrid varieties. Cytological study showed that pollen abortion in the DGMS line (W1098A) began in pollen mother cells (PMC), continued throughout pollen development, and peaked at the late microspore stage. Moreover, the gene locus conditioning male sterile was delimited by two closely linked SSR markers SBM298 and GB50 [8]. However, the underlying molecular mechanism remains elusive.

The small diploid genome (~350 Mb) makes sesame an attractive species for genetic studies [9, 10]. Recently, the high-quality genome sequence of sesame was assembled, which contains ~27,148 predicted gene models, of which 91.7 % were anchored onto 16 pseudomolecules or linkage groups (LGs) [11]. Using forward and reverse genetic approaches, a growing number of genes have been identified that have vital roles in anther development. Consequently, the Plant Male Reproduction Database (PMRD, http://202.120.45.92/addb/), a comprehensive resource for genes and mutants related to plant male reproduction, has emerged [12].

Male sterility (MS) is associated with not only the lack of viable pollen, but also the failure of pollen release [13]. The importance of tapetal programmed cell death (PCD) for successful pollen formation has been highlighted by a number of MS mutants that fail to go through normal tapetal breakdown [13–15]. Archesporial cell number and tapetal cell fate is controlled by EXCESS MICROSPOROCYTES1 (EMS1), a leucine-rich repeat receptor like kinase, and a small secreted protein ligand, TAPETUM DETERMINANT1 (TPD1) [16]. Tapetal development is initiated by DYSFUNCTIONAL TAPETUM1 (DYT1) [17] and DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION1 (TDF1) [18], with tapetal maturation, pollen wall formation, and tapetal PCD involving ABORTED MICROSPORES (AMS) [19] and MALE STERILITY1 (MS1) [20]. The final stage of dehiscence involves jasmonic acid (JA)-induced gene expression and transcription factors associated with endothecium secondary thickening [13].

To elucidate the mechanism of MS more comprehensively, the transcriptomes of many higher plants have been sequenced, including Arabidopsis [21], buckwheat [22], cotton [23–25], watermelon [26], soybean [27], Brassica napus [28–30] and Brassica oleracea [31]. In this study, fertile and sterile flower buds from DGMS line with a length of ~2.5 mm were sampled for RNA-seq, representing the first study of the sesame DGMS transcriptome. The aim of this study is to identify differentially expressed genes (DEGs) associated with MS, and explore the different bioprocesses involved and their putative functions. These results will be helpful to elucidate the molecular mechanism for DGMS, and assist the breeding of sesame hybrid variety.

Results

Transcriptome profiling of fertile and sterile buds

We have previously demonstrated that male sterility mainly occurred at PMC stage in DGMS line [8]. Therefore, we sampled fertile and sterile buds at this stage, and prepared respective cDNA libraries. After sequencing with Illumina HiSeq 2000 platform, we obtained a total of 53,126,890 and 55,491,408 high quality pair-end reads from fertile and sterile flower buds, respectively, which were then cleaned and mapped to the sesame reference genome sequence containing 27,148 gene models [11]. In total, 83.54 % of the reads from fertile buds and 84.86 % from sterile buds were mapped to the reference genome, and the majority of which were uniquely mapped (Table 1). By sequences alignment, we found that a total of 22,373 and 22,788 genes were hit by the unique reads from fertile and sterile buds, respectively, which accounted for >82 % of the known gene models. The average length of genes in fertile buds was 1305 bp and it was 1297 bp for sterile buds. Most of these genes (74 % in sterile buds and 71 % sterile buds) showed very high level of gene coverage (90–100 %).

To gauge the relative level of gene expression in different tissues, we calculated the RPKM (Reads per Kilobase of exon model per Million mapped reads) value based on the uniquely mapped reads. The RPKM value for those genes detected in fertile buds ranged from 0.012 to 16683.020, with a mean of 40.974. Similarly, the minimum, maximum and average RPKM was 0.008, 33521.52
and 40,302 for genes in sterile buds. Thus, all the above genes were regarded to be expressed in either the fertile buds or the sterile buds, as indicated by a RPKM threshold ≥0.001. Unsurprisingly, most of these expressed genes (>95 %) were common between tissues; however, we also observed a small number of uniquely expressed genes (539 in fertile buds and 954 in sterile buds).

**Functional characterization of DEGs**

Using the criteria of at least two fold changes and false discovery rate (FDR)<0.001, we obtained 1,502 significant DEGs by comparing the genes expression levels between fertile and sterile buds, of which 751 were up-regulated and 751 down-regulated in sterile buds (Additional file 1: Table S2). Distribution of all DEGs across the sesame genome was then analyzed by anchoring gene sequences to the previously released 16 pseudomolecules (or LGs) that harbored 85.3 % of the sesame genome assembly [11]. By integrating the genome information available in public domain, we could assign the DEGs onto each LG. The results showed that LG4 had the least numbers of DEGs (4.47 %), following by LG11 with 4.76 %. In contrast, LG7 had the largest percentage of DEGs (6.83 %). Moreover, the percentage of up-regulated genes was nearly 2 folds that of down-regulated genes in LG16, LG8 and LG15. Also, LG2, LG10 and LG13 had higher percentage of up-regulated genes than down-regulated genes, while LG3, LG4, LG5, LG9, LG11 and LG12 showed an opposite trend. In addition, there were nearly equal numbers of up- and down- regulated genes in the rest of the four LGs (Fig. 1).

The putative function of each DEG was then characterized with both GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases. Due to the large numbers and the complex branch structure of GO categories, only the three most abundant functional groups, namely ‘Cellular Component’, ‘Molecular Function’ and ‘Biological Process’ were presented, as an example (Fig. 2). In the sub-category of ‘Cellular Component’, the largest numbers of genes were found to be associated with ‘cell part’, which can be further sub-divided into cascades of ‘intracellular’, ‘cytoplasmic vesicle’ and ‘intrinsic to membrane’. In the next main sub-category of ‘Molecular Function’, ‘ion binding’ and ‘catalytic’ were the most abundant cascades that have a respective of 71 and 19 genes. Moreover, ‘hydrolase activity acting on glycosyl bonds’ and ‘iron ion binding’ were the two dominant groups in the cascade of ‘catalytic’. Within the last sub-category ‘Biological Process’, ‘cellular process’ and ‘metabolic process’ were the two most prevalent cascades that can represent the typical activities of biological processes. Specifically, the most intriguing GO terms in ‘cellular process’ were found to be ‘meiosis I’ and ‘pollen wall assembly’, suggesting their active roles in MS. It was noted that ‘DNA recombination’ was highlighted in the cascade ‘metabolic process’.

In the KEGG analysis, a total of 34 pathways were enriched, of which 13 were inferred from both up- and down- regulated genes, and the rest were inferred from either down- or up- regulated genes alone (Table 2). It was showed that most of the genes are involved in ‘Metabolic pathways’ and ‘Biosynthesis of secondary metabolites’. Interestingly, there were at least 6 genes (SIN_1006103, SIN_1017099, SIN_1014074, SIN_1023392, SIN_1015497 and SIN_1014349) annotated as ‘Meiosis-yeast’ or ‘Oocyte meiosis’ in the list of genes down-regulated in sterile buds, consistent with the GO annotation results. In the ‘Biosynthesis of secondary metabolites’ pathway, the number of up-regulated genes was nearly 3 times that of down-regulated genes. Also, many more up-regulated genes were annotated as ‘Polycyclic aromatic hydrocarbon degradation’ and ‘alpha-Linolenic acid metabolism’. By contrast, many genes down-regulated in sterile buds were enriched in ‘Ascorbate and aldarate metabolism’ and ‘Glycerophospholipid metabolism’. There were also 14 up-regulated genes involved in the pathway of ‘Flavonoid biosynthesis’ (Table 2).

These findings were further supported by a more specific comparison of metabolic pathways by using MapMan [32]. All of the 1,502 DEGs identified between sterile and fertile buds were annotated in the TAIR database (http://www.arabidopsis.org). Consequently, 1,445 DEGs were found to be homologs of 1,240 Arabidopsis genes (Additional file 2: Table S3). To dissect the putative functions of the 1,445 DEGs that are likely to be associated

| Table 1 Summary of mapping transcriptome reads to reference sequence of sesame | Fertile Buds | Sterile Buds |
|---|---|---|
| Quantity | Percentage % | Quantity | Percentage % |
| Total Reads | 53,126,890 | 100.00 | 55,491,408 | 100.00 |
| Total Mapped Reads | 44,382,564 | 83.54 | 47091778 | 84.86 |
| Unique Match | 43,380,846 | 81.66 | 46141578 | 83.15 |
| Perfect Match | 36,189,950 | 68.12 | 37126510 | 66.90 |
| <5 bp Mismatch | 8,192,614 | 15.42 | 9965268 | 17.96 |
with MS phenotype, we fully visualized the Arabidopsis homologous genes with MapMan and inferred a candidate pathway network (Fig. 3).

In the network, the most significant changes in transcript abundance of genes were shown to be related to ‘Protein,’ ‘Targeting,’ ‘Hormones’ and ‘DNA.’ Moreover, the expression of genes implicated in ‘Ethylene and JA synthesis’ were up-regulated in sterile buds, while those genes involved in ‘Signaling pathway’ were down-regulated in the DGMS sterile buds. In addition, the DEGs involved in ‘Lipid (FA synthesis),’ ‘Redox (Ascorbate & Glutathion)’ and ‘Energy (transport p- and v-ATPases)’ were all down-regulated, whereas those in ‘Second Metabolism (Flavonoids),’ ‘Cell Wall (Modification),’ and ‘Energy (Fermentation)’ were up-regulated in sterile buds, if compared to those in fertile buds. Among the differentially expressed transcription factors within the ‘RNA TF’ group, all of the NAC, trihelix and WRKYs (except one WRKY) were up-regulated, whereas C2C2(Zn) DOF, CCAAT and SET were down-regulated. Furthermore, in the ‘Signalling’ category, two MAP kinase-coding genes were down-regulated in the sterile buds (Fig. 3; Additional file 2: Table S3).

Identification of male-sterility/male-reproduction related genes

To gain a deeper insight into the molecular mechanism underlying MS, we queried the sesame DEGs in the PMRD which contains 548 Arabidopsis male-sterility/male-reproduction related genes. Forty nine homologous genes related to plant male reproduction were retrieved; several of these genes encode transcription factors (e.g. bHLH089, MYB99 and AMS). The transcription factor encoding genes showed reduced expressions in sterile buds, implicating their important roles in specifying/determining tapetal fate and development (Table 3).
Allelic variants of DEGs

To gain a better understanding of the DEGs, we further predicted the effect of allelic variants on the function of their target genes using SnpEff predictor. A total of 1,057 Insertion/Deletions (InDels) were detected in 982 genes expressed in fertile buds, of which 52 reside within 48 DEGs (some genes have two InDels) (Additional file 3: Table S4). Similarly, 1,432 InDels were detected in 1,354 genes expressed in sterile buds, and 86 InDels were located within 83 DEGs (Additional file 4: Table S5). Together, we identified 138 InDels within 131 genes that were differentially expressed either in fertile or sterile buds. Of the 138 InDels identified, 62 were located in 57 genes that were up-regulated in sterile buds, and 76 were located in 68 genes that were down-regulated in sterile buds (Additional files 5: Table S6 and 6: Table S7).

Specifically, in the list of up-regulated genes, a number of transcription factor encoding genes such...
as SIN_1002610 (Ethylene-responsive transcription factor ERF106), SIN_1024026 (NAC2), SIN_1019334 (WRKY 28) and SIN_1011023 (WRKY 33) were found. Some genes encoding ‘Brassinosteroid-regulated protein BRU1’ (SIN_1022411), ‘COP9 signalosome complex subunit 2’ (SIN_1015172) and ‘Defensin J1-2’ (SIN_1021298) were also highlighted (Additional file 5: Table S6). In the list of down-regulated genes, SIN_1008339 (E3 ubiquitin-protein ligase MARCH1), SIN_1010740 (L-ascorbate oxidase homolog), SIN_1026145 (Pollen-specific protein SF3), SIN_1005014 (Protein disulfide-isomerase 5–3) and SIN_1010051 (Sugar transport protein 8) were of interest in that they were likely to be related with pollen development (Additional file 6: S7).

A subset of 21 genes containing InDels that were predicted to cause loss of function (LOF) and/or codon change (CC) was selected for further analysis (Table 4). Of these, InDels likely to cause CC (termed ‘CC-type’) were detected in 6 genes at sterile alleles, and in other 6 genes at fertile alleles. Moreover, LOF-type InDels were also detected in 6 fertile alleles and 7 sterile alleles, which showed a higher expression level in fertile buds and sterile buds, respectively (marked with asterisk; Table 4). Thus, it seemed that LOF-Type InDel might lead to the increase of transcript abundance in which it resides. This observation was further confirmed by the fact that in the 11 genes up-regulated in fertile buds, the majority (80 %) of the InDels were detected in sterile alleles.

In particular, some genes such as SIN_1025190 (SCP18, Serine carboxypeptidase), SIN_1017245 (F3PH, Flavonoid 3′-monooxygenase) and SIN_1018350 (IPT, Adenylate isopentenyltransferase) with both LOF-type and CC-type InDels in sterile alleles, were up-regulated in sterile buds. Moreover, the gene encoding a kinase (SIN_1004626) with both LOF- and CC- types of InDels in fertile allele was up-regulated in fertile buds (down-regulated in sterile buds). Interestingly, in another gene, SIN_1005818 (HMGB9, High mobility group B protein 9), InDel was detected in both alleles, with putative disruptive_inframe_deletion in sterile allele and LOF in fertile allele. The expression of this gene was down-regulated in fertile buds but up-regulated in sterile buds (Table 4, Additional file 7: Table S8). Taken together, a large number of sequence variants were detected in these DEGs, and their effects on transcript abundances were not conclusive.

**Real-time quantitative PCR validation**

To verify the RNA-Seq results, we chose an alternative strategy for both the up- and down-regulated DEGs. Twenty genes were randomly selected for validation by Real-time quantitative PCR (qRT-PCR) using the same RNA samples that was used for RNA-Seq. Primer sets were designed to span exon-exon junctions (Additional file 8: Table S1). Results showed that
Table 3  Sesame DEGs homologous to Arabidopsis male-sterility/reproduction genes

| Query IDa | log2FC | Subject ID | E-value | Score | Symbol | Description |
|-----------|--------|------------|---------|-------|--------|-------------|
| SIN_1005144 | -10.88 | AT4G39910  | 1E-22   | 90.1  | ATUBP3 | ubiquitin-specific protease 3 (UBP3) |
| SIN_1012023 | -8.47  | AT1G69940  | 2E-140  | 407   | AtpPME1 | PPM1 |
| SIN_1020647 | -4.94  | AT1G65470  | 2E-139  | 418   | ARF17  | auxin response factor 17 (ARF17) |
| SIN_1015972 | -4.78  | AT4G02780  | 4E-103  | 315   | QRT2   | QUARTET 2 (QRT2) |
| SIN_1008904 | -2.36  | AT1G77850  | 3E-139  | 418   | ARF17  | auxin response factor 17 (ARF17) |
| SIN_1001388 | -2.07  | AT4G21330  | 0       | 129   | DYT1   | DYSFUNCTIONAL TAPETUM 1 (DYT1) |
| SIN_1007044 | -1.94  | AT1G65470  | 2E-79   | 264   | FAS1   | FASCIATA 1 (FAS1) |
| SIN_1008202 | -1.61  | AT1G77850  | 0       | 367   | ATXR6, SDG34 Arabidopsis Trithorax-Related Protein 6 (ATXR6) |

| Upregulated, homologs of MS gene (cloned) |
| SIN_1004507 | 1.00  | AT4G34850  | 0       | 624   | LAP5   | LESS ADHESIVE POLLEN 5 (LAPS) |
| SIN_1013503 | 1.36  | AT1G65470  | 0       | 367   | ATXR6, SDG34 Arabidopsis Trithorax-Related Protein 6 (ATXR6) |

| Downregulated, homologs of MR gene with mutant evidence |
| SIN_1026151 | -1.02 | AT2G14760  | 2E-68   | 221   | C2H2-like zinc finger protein |
| SIN_1006103 | -2.68 | AT3G51430  | 0.015   | 335   | AGP6   | arabinogalactan protein 6 (AGP6) |

| Downregulated, homologs of MR gene with GO evidence |
| SIN_100609 | -2.46  | AT3G47870  | 1E-37   | 137   | LBD27, SCP | LOB domain-containing protein 27 (LBD27) |
| SIN_10107022 | -1.69 | AT1G78130  | 0       | 665   | UNE2   | unfertilized embryo sac 2 (UNE2) |
| SIN_1014074 | -1.63  | AT1G01690  | 2E-48   | 170   | ATPRD3, PRD3 | putative recombination initiation defects 3 (PRD3) |
| SIN_1017944 | -1.59  | AT1G64110  | 0       | 1150  | P-loop containing nucleoside triphosphate hydrolases |
| SIN_10155329 | 2.22  | AT3G23240  | 3E-31   | 114   | ATERF1, ERF1 | ethylene response factor 1 (ERF1) |
| SIN_1009935 | 2.28  | AT3G61230  | 6E-104  | 303   | PLIM2c | GATA type zinc finger transcription factor |
| SIN_1006818 | 1.84  | AT1G02205  | 0       | 711   | CER1   | ECERIFERUM 1 (CER1) |
| SIN_1011100 | 1.84  | AT3G39400  | 1.6     | 26.2  | BIN1, BRI1, DWF2 | BRASSINOSTEROID INSENSITIVE 1 (BRI1) |
| SIN_1006818 | 1.97  | AT4G20050  | 0       | 558   | QRT3   | QUARTET 3 (QRT3) |
| SIN_1005329 | 2.22  | AT3G23240  | 3E-31   | 114   | ATERF1, ERF1 | ethylene response factor 1 (ERF1) |
| SIN_1009935 | 2.28  | AT3G61230  | 6E-104  | 303   | PLIM2c | GATA type zinc finger transcription factor |
| SIN_1004029 | 2.53  | AT3G51430  | 0.015   | 335   | AGP6   | arabinogalactan protein 6 (AGP6) |
although genes expression fold changes detected by qRT-PCR, in most cases, were higher than those by RNA-Seq, the trends were similar between these two methods, thus confirming the accuracy and reliability of RNA-Seq. As an example, the expression patterns of 12 randomly selected Male-sterility/male-reproduction genes were listed in Table 5, which demonstrated that the expression levels revealed by qRT-PCR and RNA-Seq were highly correlated ($r = 0.762$, $P < 0.01$, $n = 12$).

**Discussion**

We presented here, to our knowledge, the first study of sesame DGMS at transcriptome level. Transcript abundances from both fertile and sterile buds were acquired

| Table 3 | Sesame DEGs homologous to Arabidopsis male-sterility/reproduction genes (Continued)|
| --- | --- |
| Upregulated, homologs of MR gene with GO evidence |
| SIN_1014884 | 1.00 | AT4G34050 | 3E-163 | 456 | CCaOMT1 caffeoyl coenzyme A O-methyltransferase 1 (CCaOMT1) |
| SIN_1002228 | 1.03 | AT3G28470 | 2E-39 | 144 | TDF1, ATMYB35 Defective In Meristem Development And Function 1 (TDF1) |
| SIN_1026357 | 1.30 | AT5G23530 | 2E-101 | 305 | AtCCE18, CXE18 carboxyesterase 18 (CCE18) |
| SIN_1012979 | 1.41 | AT5G03700 | 5E-155 | 454 | D-mannose binding lectin protein |
| SIN_1002445 | 1.57 | AT2G37040 | 0 | 1104 | ATPAL1, PAL1 PHE ammonia lyase 1 (PAL1) |
| SIN_1019202 | 1.66 | AT3G95350 | 0 | 608 | LAP3 Calcium-dependent phosphotriesterase |
| SIN_1019338 | 2.29 | AT3G17220 | 3E-08 | 50.4 | ATPMEI2, PMEI2 pectin methylesterase inhibitor 2 (PMEI2) |
| SIN_1007695 | 7.68 | AT2G19070 | 0 | 540 | SHT spermidine hydroxycinnamoyl transferase (SHT) |
| SIN_1005049 | 7.69 | AT4G27290 | 0 | 873 | S-locus lectin protein kinase family protein |

| Table 4 | 21 DEGs with InDels prone to cause loss of function or codon change |
| --- | --- |
| Gene ID | InDel | Codon-change | log$_2$(S/F) | Annotation |
| Allele from sterile buds |
| SIN_1025190 | c.1151_1151 + 1insCG | p.Gly385fs | 10.09 | SCP18_ARATH Serine carboxypeptidase-like 18 |
| SIN_1017245 | c.435dupA | p.Leu146fs | 3.51 | F3PH_ARATH Flavonoid 3’-monooxygenase |
| SIN_101325 | c.507_507 + 1insT | p.Thr297fs | 1.39 | IPT_HUMLU Adenylate isopentenyltransferase |
| SIN_1022131 | c.465 + 1_465 + 2insA | p.His381fs | 3.85 | AB22G_ARATH ABC transporter G family |
| SIN_1007649 | c.886-2_886-1insT | p.Leu294fs | 1.47 | BGAL_MALDO Beta-galactosidase |
| SIN_1001108 | c.934-2_934-1insCC | p.Ala296fs | 4.32 | E13B_WHEAT Glucan endo-1,3-beta-glucosidase |
| SIN_1010915 | c.507_507 + 1insTTTGAAGAT | p.Glu182fs | 1.28 | DHI1L_XENTR Hydroxysteroid 11-beta-dehydrogenase 1 |
| SIN_1025700 | c.318_319insCTGGAA | p.Val106_A107insLE | 1.00 | SCL32_ARATH Scarecrow-like protein 32 |
| Allele from fertile buds |
| SIN_1005818 | c.394 + 2delT | p.Ala25fs | 1.49 | HMG9_ARATH High mobility group B protein 9 |
| SIN_1004626 | c.74delC | p.Arg25fs | 1.30 | AMERL_ARATH AMER1 domain-containing protein |
| SIN_1019529 | c.18_20dupAGG | p.Gly7dup | 1.54 | DUF3774 Wound-induced protein |

* means InDel may cause Loss of function (LOF), as predicted by SnpEff. See Additional file 7: Table S8 for full information

[57x516]although genes expression fold changes detected by qRT-PCR, in most cases, were higher than those by RNA-Seq, the trends were similar between these two methods, thus confirming the accuracy and reliability of RNA-Seq. As an example, the expression patterns of 12 randomly selected Male-sterility/male-reproduction genes were listed in Table 5, which demonstrated that the expression levels revealed by qRT-PCR and RNA-Seq were highly correlated ($r = 0.762$, $P < 0.01$, $n = 12$).

**Discussion**

We presented here, to our knowledge, the first study of sesame DGMS at transcriptome level. Transcript abundances from both fertile and sterile buds were acquired

| Table 3 | Sesame DEGs homologous to Arabidopsis male-sterility/reproduction genes (Continued)|
| --- | --- |
| Upregulated, homologs of MR gene with GO evidence |
| SIN_1014884 | 1.00 | AT4G34050 | 3E-163 | 456 | CCaOMT1 caffeoyl coenzyme A O-methyltransferase 1 (CCaOMT1) |
| SIN_1002228 | 1.03 | AT3G28470 | 2E-39 | 144 | TDF1, ATMYB35 Defective In Meristem Development And Function 1 (TDF1) |
| SIN_1026357 | 1.30 | AT5G23530 | 2E-101 | 305 | AtCCE18, CXE18 carboxyesterase 18 (CCE18) |
| SIN_1012979 | 1.41 | AT5G03700 | 5E-155 | 454 | D-mannose binding lectin protein |
| SIN_1002445 | 1.57 | AT2G37040 | 0 | 1104 | ATPAL1, PAL1 PHE ammonia lyase 1 (PAL1) |
| SIN_1019202 | 1.66 | AT3G95350 | 0 | 608 | LAP3 Calcium-dependent phosphotriesterase |
| SIN_1019338 | 2.29 | AT3G17220 | 3E-08 | 50.4 | ATPMEI2, PMEI2 pectin methylesterase inhibitor 2 (PMEI2) |
| SIN_1007695 | 7.68 | AT2G19070 | 0 | 540 | SHT spermidine hydroxycinnamoyl transferase (SHT) |
| SIN_1005049 | 7.69 | AT4G27290 | 0 | 873 | S-locus lectin protein kinase family protein |

| Table 4 | 21 DEGs with InDels prone to cause loss of function or codon change |
| --- | --- |
| Gene ID | InDel | Codon-change | log$_2$(S/F) | Annotation |
| Allele from sterile buds |
| SIN_1025190 | c.1151_1151 + 1insCG | p.Gly385fs | 10.09 | SCP18_ARATH Serine carboxypeptidase-like 18 |
| SIN_1017245 | c.435dupA | p.Leu146fs | 3.51 | F3PH_ARATH Flavonoid 3’-monooxygenase |
| SIN_1018350 | c.889_890insT | p.Thr297fs | 1.39 | IPT_HUMLU Adenylate isopentenyltransferase |
| SIN_101325 | c.507_507 + 1insTTTGAAGAT | p.Glu182fs | 1.28 | AB22G_ARATH ABC transporter G family |
| SIN_1022131 | c.465 + 1_465 + 2insA | p.His381fs | 3.85 | DHI1L_XENTR Hydroxysteroid 11-beta-dehydrogenase 1 |
| SIN_1007649 | c.886-2_886-1insT | p.Leu294fs | 1.47 | BGAL_MALDO Beta-galactosidase |
| SIN_1001108 | c.934-2_934-1insCC | p.Ala296fs | 4.32 | E13B_WHEAT Glucan endo-1,3-beta-glucosidase |
| SIN_1010915 | c.507_507 + 1insTTTGAAGAT | p.Glu182fs | 1.28 | DHI1L_XENTR Hydroxysteroid 11-beta-dehydrogenase 1 |
| SIN_1025700 | c.318_319insCTGGAA | p.Val106_A107insLE | 1.00 | SCL32_ARATH Scarecrow-like protein 32 |
| Allele from fertile buds |
| SIN_1005818 | c.394 + 2delT | p.Ala25fs | 1.49 | HMG9_ARATH High mobility group B protein 9 |
| SIN_1004626 | c.74delC | p.Arg25fs | 1.30 | AMERL_ARATH AMER1 domain-containing protein |
| SIN_1019529 | c.18_20dupAGG | p.Gly7dup | 1.54 | DUF3774 Wound-induced protein |

* means InDel may cause Loss of function (LOF), as predicted by SnpEff. See Additional file 7: Table S8 for full information
by RNA-Seq using the Illumina sequencing platform. We then mapped the high quality transcriptome reads onto the sesame reference genome and identified more than 22 thousands expressed genes, of which only 1,502 genes (~6.6%) were differentially expressed in either sterile or fertile buds, suggesting that a limited number of key genes are responsible for MS (cloned) and MR related (male-reproduction related genes, with GO evidence), thus demonstrating that all these genes accounted for nearly one half of the genes within each regulated category, and the rest of genes were annotated as MR related (male-reproduction related genes, with GO evidence), thus demonstrating that all these genes might be good candidates responsible for MS (Table 3). This can be explained by the fact that the sesame MS homologs in Arabidopsis, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

From DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1024646 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned to the sesame MS genes (cloned) list (Table 3). Previous study also showed that RBOHE (At1g19230) was an anther-preferential or tapetum-enriched gene, and functional loss of RBOHE resulted in delayed tapetal degeneration, thus the expression of RBOHE was reduced in dyt1 and tdf1 [33]. Consistent with this, we found that the RBOHE homologs in sesame, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

Apart from DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1007695 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned to the sesame MS genes (cloned) list (Table 3). Previous study also showed that RBOHE (At1g19230) was an anther-preferential or tapetum-enriched gene, and functional loss of RBOHE resulted in delayed tapetal degeneration, thus the expression of RBOHE was reduced in dyt1 and tdf1 [33]. Consistent with this, we found that the RBOHE homologs in sesame, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

Apart from DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1007695 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned to the sesame MS genes (cloned) list (Table 3). Previous study also showed that RBOHE (At1g19230) was an anther-preferential or tapetum-enriched gene, and functional loss of RBOHE resulted in delayed tapetal degeneration, thus the expression of RBOHE was reduced in dyt1 and tdf1 [33]. Consistent with this, we found that the RBOHE homologs in sesame, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

Apart from DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1007695 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned to the sesame MS genes (cloned) list (Table 3). Previous study also showed that RBOHE (At1g19230) was an anther-preferential or tapetum-enriched gene, and functional loss of RBOHE resulted in delayed tapetal degeneration, thus the expression of RBOHE was reduced in dyt1 and tdf1 [33]. Consistent with this, we found that the RBOHE homologs in sesame, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

Apart from DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1007695 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned to the sesame MS genes (cloned) list (Table 3). Previous study also showed that RBOHE (At1g19230) was an anther-preferential or tapetum-enriched gene, and functional loss of RBOHE resulted in delayed tapetal degeneration, thus the expression of RBOHE was reduced in dyt1 and tdf1 [33]. Consistent with this, we found that the RBOHE homologs in sesame, SIN_1024646 and SIN_1007549, also displayed significantly reduced expression in sterile buds (log2S/F = −1.7 and −0.9), if compared to fertile buds (Additional file 1: Table S2). Therefore, RBOHE may have a similar function in sesame DGMS.

Apart from DYT1 mentioned above, QRT2 (QUARTET2) was also in the MS genes (cloned) list (Table 3). Three QRT genes including QRT2 are required for the degradation of pollen mother cell wall when microspores are released from their tetrads [12]. Furthermore, QRT2 are required for anther dehiscence. In the process of floral abscission which is co-regulated by JA, ethylene and abscisic acid (ABA), QRT2 is regulated by ethylene and ABA [34]. Moreover, anther dehiscence-related polygalacturonase activity is likely to be regulated by JA, ethylene and ABA [13]. In this study, the reduced expression of QRT2 was coupled with the up-regulation of genes involved in ethylene synthesis.

There were 17 up-regulated sesame genes with homologs in Arabidopsis (8 homologous to MS genes and 9 to MR genes, Table 3). Of these, the expression level of SIN_1007695 (sperrmidine hydroxyccinnamoyl transferase, SHT) showed >200 fold increase in sterile buds, which was reminiscent of SHT expressed in the tapetum of Arabidopsis anthers [35]. Moreover, SHT was assigned
into ‘cluster 81’ by the online tool of FlowerNet [36], which includes several genes such as KCS10, GH31 and ATA7; their homologs in sesame (i.e. SIN_1007525, SIN_1025709 and SIN_1002500) were co-up-regulated in sterile buds (Additional file 1: Table S2), implying their possible involvement in MS. This ‘cluster 81’ also contained TSMT1 (tapetum-specific methyltransferase1), which encodes a cation-dependent CCoAOMT-like protein involved in phenylpropanoid polyamine conjugate biosynthesis and has a role in the stamen/pollen development of Arabidopsis [37]; the rest of genes with unknown functions are likely to play roles in pollen exine and lipid biosynthesis, based on their description in AtEnsembl [36]. Therefore, it would be worthy of investigating the rest genes within this cluster to get a clear view of their function.

JA is specifically required for anther dehiscence during anther development [38]. Mutations in genes that participate in JA biosynthesis and perception cause a failure or delay in anther dehiscence and pollen inviability which result in male sterility [39]. Examples of such genes include the DEFECTIVE IN ANOTHER DEHISCEENCE 1 (DAD1), which encodes a phospholipase A1 that catalyses the initial step of JA biosynthesis; AOS, a gene that encodes allene oxide synthase; DEHISCENCE 1 (DDE1)/OPR3, which encodes the OPR protein 12-oxo-phytodienoic acid reductase in the JA synthesis pathway [40]. Defects in all stages of the JA pathway appear to cause similar phenotypes of reduced filament elongation and a lack of dehiscence. Delayed dehiscence or non-dehiscence phenotypes have been observed in mutants defective in JA biosynthetic enzymes [13]. In this study, SIN_1016850 (homolog of PLA15, Phospholipase A1-Igamma1) was significantly up-regulated in sterile buds, whereas the homologs of allene oxide synthase encoding genes did not show differences (data not shown). However, SIN_1022877 and SIN_1022878, which are homologs of OPR1 (12-oxophytodienoate reductase 1) in Arabidopsis, displayed obvious down-regulation in sterile buds (Additional file 1: Table S2). These data strongly indicated that genes involved in JA pathway are also responsible for MS in sesame.

Plant gene expression regulation is a complicated network. Through specific interactions with cis-acting target elements, transcription factors can regulate a series of relevant down-stream targets, which play an important role in plant development and the response to environmental stress. Arabidopsis ANOTHER IN-DEHISCEENCE FACTOR (AIF), a NAC-like gene, acts as a repressor that controls anther dehiscence by regulating genes in the jasmonate biosynthesis [38]. In fact, for the annotated NACs in Swissprot, all of the 9 sesame homologs were up-regulated in sterile buds, which strengthen the role of NACs in the regulation of MS (Fig. 3, Additional file 1: Table S2). Furthermore, 11 of the 12 WRKYs that were significantly up-regulated in sterile buds, were annotated as the orthologs of WRKY33 (Fig. 3, Additional files 1: Table S2). WRKY33 proteins are evolutionarily conserved with a critical role in broad plant stress responses, and Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses [41]. Moreover, genes involved in redox homeostasis, salicylic acid (SA) signaling, ethylene-JA-mediated cross-communication and camalexin biosynthesis were identified as direct targets of WRKY33 [42]. Furthermore, the down-regulation of JA-associated responses appears to involve direct activation of several jasmonate ZIM-domain genes, encoding repressors of the JA-response pathway, by loss of WRKY33 function and by additional SA-dependent WRKY factors. In the present study, the co-expression behavior of NACs and WRKYs suggested their pivotal roles in regulating the sesame MS (Fig. 3, Additional file 1: Table S2).

To understand the impact of sequence variation on gene expression, the effects of allelic variants on the function of their target genes were predicted using SnpEff. Interestingly, 6 InDels were found in fertile alleles, which were up-regulated in fertile buds (and the wild-type sterile allele had lower level of expression in sterile buds); and 7 InDels were found in sterile alleles, which were up-regulated in sterile buds (Table 4). This observation suggested that the causal effect of sequence variation on transcript abundance was not so straightforward, but rather confound. This can be explained by the way that most of the InDels were detected in coding regions rather than in the promoter regions, in which it can directly affect the transcript abundance. Occasionally, we also identified InDels showing a transcriptional-regulatory function, in which the transcript abundance was decreased by the existing of causative InDels. For example, two genes (SIN_1025700 and SIN_1005818) with InDels in sterile alleles caused a decrease of transcript abundances in sterile buds, and another two genes (SIN_1004703 and SIN_1019529) with InDels in fertile alleles led to the down-regulation of genes in fertile buds, thus demonstrating a cis-acting fashion.

As suggested by Rutley and Twell [43], transcriptome studies of the male gametophyte have not only increased our knowledge and understanding, but also improved the efficacy of experimental strategies by informing experimental design (such as by gene selection for reverse genetics) and through query-based and co-expression analysis. The present investigation provided many DEGs and a number of candidate genes that can be used to elucidate the molecular mechanism underlying sesame DGMS through transgenic verification in future.
Conclusions
This study provided a set of 1,502 genes differentially expressed in the fertile and sterile buds of sesame DGMS lines based on transcriptome profiling. Half of these genes were up-regulated in sterile buds, demonstrating a complex expression pattern. Regarding the genes implicated in ethylene and JA synthesis & signaling, the expression of which were up- and down-regulated in the sterile buds, respectively. Furthermore, the majority of NAC and WRKY transcription factors were up-regulated in sterile buds.

Moreover, 49 sesame genes with homologs in Arabidopsis related with male-sterility/male-reproduction showed reduced expression in sterile flower. Some of these genes encode transcription factors (bHLH089, MYB99, and AMS) that possibly have a role in specifying or determining tapetal fate and development. Furthermore, the predicted effect of allelic variants on the function of target gene highlighted several InDels, which might contribute to fertility determination.

Methods

Plant materials and RNA preparation
The sesame plant materials used in this study include the newly developed DGMS line W1098A and its fertile counterpart W1098B, which differed from each other only by pollen fertility [8]. These two lines were both cultivated in the experimental fields of the Oil Crops Research Institute, CAAS (Wuhan, Hubei Province, China). Buds with a length of ~2.5 mm were separately stripped from each of five male sterile and fertile plants and bulked for transcriptomic profiling. The fertile bulk and the sterile bulk of buds were immediately snap-frozen in liquid nitrogen and then stored at −80 °C freezer until use. Total RNA was isolated from bulks of sterile buds and fertile buds with TRIzol reagent (Gibco-BRL) according to the manufacturer’s instruction. Then two cDNA libraries were constructed from sterile and fertile buds, as previously described in sesame [44]. Briefly, approximately 5 mg of mRNA was fragmented, converted to cDNA, and PCR amplified according to the Illumina RNA-Seq protocol (Illumina, Inc. San Diego, CA). Sequence reads were generated using the Illumina Genome AnalyzerII (SanDiego, CA) and Illumina HiSeq 2000 platform (San Diego, CA) at the Beijing Genomics Institute (Shen Zhen, China).

Identification of Differentially Expressed Genes
The clean reads were mapped to the reference genome sequence of *S. indicum* (http://ocri-genomics.org/Sinbase/) [11] using SOAP aligner/soap2 (an improved ultrafast tool for short read alignment) [45]. RPKM were used to gauge the relative transcript abundance for each gene. Using the DEGseq program, significantly differential gene expression was identified between the fertile and sterile buds libraries [46]. The FDR was used to determine the threshold *p*-value. In this study, a stringent of FDR ≤ 0.001 and | log2 (Fold change ratio of sterile/fertile) | ≥ 1.00 was used as the threshold to select a significantly different expressed gene.

Characterization of genetic variations
Characterization of the sequence variants such as InDels was performed using SnpEff version 4.1 [47] by referring to sesame genome annotation downloaded from the Sinbase (http://ocri-genomics.org/Sinbase) according to Wang et al. [48]. Sequence variants (InDels, frame shift, stop gained, stop lost and non synonymous coding) that potentially have high impact on transcript/protein were predicted according to the method described by Saeed et al. [49].

GO and KEGG Pathway Enrichment Analysis
The DEGs were used for GO and pathway enrichment analysis. A corrected *P* ≤ 0.05 was selected as the threshold of significance to determine enrichment in the gene sets [50]. Functional classes inferred from DEGs were assigned according to GO mapping provided by the ensemble database. The Blast2GO program (https://www.blast2go.com/) was used to obtain GO annotations for the all DEGs [51]. Then, the results were submitted to WEGO (http://wego.genomics.org.cn) to generate a GO classification graph of all DEGs [52].

KEGG pathway analysis was based on the comparative results between our mapped genes and the current KEGG database [53]. MapMan (version 3.5.1 R2) was also used to annotate the DEGs onto metabolic pathways.

Confirmation of candidate DEGs by qRT-PCR
To validate the DEGs detected by RNA-seq, 20 DEGs were randomly selected from 52 common differentially expressed genes in two libraries and then subjected to qRT-PCR analysis, according to Qi et al. [54]. Gene-specific primers were designed with the online tool Primer3 [55] based on the selected unigenes sequences (Additional file 8: Table S1). Reactions were performed with the SYBR Green Real time PCR Master Mix (TOYOBO, Japan) in a Bio-Rad CFX96 instrument. For each sample, three replicates were run for each gene in a 96-well plate. The relative expression level of each gene was determined using the 2^ΔΔCt method [56]. All data are expressed as mean ± standard deviation.

Additional files

**Additional file 1: Table S2.** Expressions and annotations of the 1502 differentially expressed unigenes in sesame. (XLSX 338 kb)

**Additional file 2: Table S3.** Blast results of sesame 1445 DEGs in MapMan pathway analysis. (XLSX 264 kb)
Additional file 3: Table S4. 52 DEGs with InDels detected in fertile buds. (XLSX 28 kb)

Additional file 4: Table S5. 86 DEGs with InDels detected in sterile buds. (XLSX 37 kb)

Additional file 5: Table S6. 62 InDels in 57 DEGs up-regulated in sterile buds. (XLSX 29 kb)

Additional file 6: Table S7. 77 InDels in 68 DEGs down-regulated in sterile buds. (XLSX 33 kb)

Additional file 7: Table S8. The predicted effects of InDels in DEGs. The sequence variations (i.e. InDels) were first detected between sterile and fertile alleles and their potential effects (i.e. causing loss of gene function or codon change) were then predicted by the software SnpEff. (XLSX 15 kb)

Additional file 8: Table S1. List of primer sequences for qRT-PCR. (XLSX 11 kb)

Abbreviations
ABA: Ethylene and abscisic acid; AIF: ANOTHER INDEHISCENCE FACTOR; AMS: ABBREVIATED MICROSPORES; AOS: Allele oxidase syntheate; CMI: Cytoplasmic male sterility; DD1: DEFECTIVE IN ANOTHER DEHISCENCE 1; DDE1: DEHISCENCE 1 (also known as OPR3); DEGs: Differentially expressed genes; DGMS: Dominant genic male sterile line; DYT1: DYSFUNCTIONAL TAPETUM1; EMS1: EXCESS MICROSPOROCYTES1 (also known as EXTRA SPOROGENOUS CELLS); FDR: False discovery rate; FDR: False discovery rate; GO: Gene Ontology; GO: Gene Ontology; G20: Gene Ontology; G20: Gene Ontology; HMG89: High mobility group b protein 9; InDel: Insertion/Deletion; IPT: Adenylate isopentenyltransferase; KEGG: Kyoto encyclopedia of genes and genomes; LG: Linkage group; LG: Linkage groups; LOF: Loss of function; MS: Male sterility; MS: Male sterility; MS1: MALE STERILITY1; OP1: 12-oxophytodienoate reductase 1; PCR: Programmed cell death; PLA15: Phospholipase A1-2; PMR: Pollen mother cell; PMRD: Plant Male Reproduction Database; RFP: Respiratory burst oxidase homologues; RBOHE: RESPIRATORY BURST OXIDASE HOMOLOGUE E; RPKM: Reads per Kilobase of exon model per Million mapped reads; SHT: Spermidine hydroxycinnamoyl transferase; SHT: Spermidine hydroxycinnamoyl transferase; TDF1: DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION1; TDF2: TAPETAL DEVELOPMENT AND FUNCTION1; TDP1: TAPETAL DETERMINANT1

Acknowledgements
We thank colleagues in BGI (Beijing Genome Institute, China) for valuable materials for sequencing. LL and MT carried out bioinformatics analysis of the raw RNA-Seq data used in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under the accession number SRP076254 (http://trace.ncbi.nlm.nih.gov/Traces/study/?acc=SRP076254).

Authors’ contributions
HL, MT and YZ designed the research; HL and YZ prepared the plant materials for sequencing; LL and MT carried out bioinformatics analysis of the raw RNA-Seq data; HY, FZ, MY and TZ performed the qPCR experiments and statistical analyses; HL and MT interpreted the data and wrote the manuscript. All authors read and approved the final manuscript.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Ethics approval and consent to participate
Not applicable.

Author details
1Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, Hubei 430062, China. 2College of Life Sciences, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China.

Received: 29 June 2016 Accepted: 27 October 2016
Published online: 10 November 2016

References
1. Karatzis K, Stamatelopoulos K, Lykka M, Mantzouratou P, Skalidi S, Zakopoulos N. Sesame oil consumption exerts a beneficial effect on endothelial function in hypertensive men. Eur J Prev Cardiol. 2013;20:202–8.
2. Periasamy S, Hsu DZ, Chang PC, Liu MY. Sesame oil attenuates nutritional fibrosis steatohepatitis by modulating matrix metalloproteinases-2, 9 and PPAR-gamma. J Nutr Biochem. 2014;25:337–44.
3. Anilkumar KR, Pal A, Khanum F, Bawa AS. Nutritional, medicinal and industrial uses of sesame (Sesamum indicum L) seeds—an overview. Agric Conspec Sci (ACS). 2010;75(4):159–68.
4. Uzun B, Anslan C, Furat S. Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (Sesamum indicum L). J Am Oil Chem Soc. 2008;85:1135–42.
5. Erbas M, Sekerci H, Gul S, Furat S, Yol E, Uzun B. Changes in total antioxidant capacity of sesame (Sesamum indicum L) by variety. Asian J Chem. 2009;21:5549–55.
6. Murty DS. Heterosis, combining ability and reciprocal effects for agronomic and chemical characters in Sesamum. Theor Appl Genet. 1975;45:294–9.
7. Zheng YZ, Zhang HY, Mei HK, Zhang TD, Wei SL. Advances in Chinese hybrid sesame research. J Henan Agric Sci. 2003;11:17–9 (In Chinese with English Abstract).
8. Liu HY, Zhou XA, Wu K, Yang MM, Zhao YZ. Inheritance and molecular mapping of a novel dominant genic male-sterile gene in Sesamum indicum L. Moll Breed. 2015:35:9. doi:10.1007/s11032-015-0189-5.
9. Zhang H, Mao H, Wang L, Qu L, Liu H, Wang Q, Yue M. Genome sequencing of the important oilseed crop Sesamum indicum L. Genome Biol. 2013:14(1):401.
10. Wei X, Liu K, Zhang Y, Feng Q, Wang L, Zhao Y, Li D, Zhao Q, Zhu X, Xu L, Li W, Fan D, Gao Y, Lu Y, Zhang X, Tang X, Zhu C, Zhu L, Zhang R, Tian Q, Wen Z, Weng Q, Han B, Huang X, Zhang X. Genetic discovery for oil production and quality in sesame. Nat Commun. 2015:6:8609.
11. Wang L, Yu S, Tong C, Zhao Y, Liu Y, Song C, Zhang Y, Zhang X, Wang Y, Hua W, Li D, Li D, Li F, Yu J, Xu F, Han X, Huang S, Tai S, Wang J, Xu X, Li Y, Liu S, Vanshenky RK, Wang J, Zhang X. Genome sequencing of the high oil crop sesame provides insights into oil biosynthesis. Genome Biol. 2014(15):R39.
12. Cui X, Wang Q, Yin W, Xu H, Wilson ZA, Wei C, Pan S, Zhang D. PMRD: a curated database for genes and mutants involved in plant male reproduction. BMC Plant Biol. 2012;12:215.
13. Wilson ZA, Song L, Taylor B, Yang C. The final split: the regulation of anther dehiscence. J Exp Bot. 2011;62(5):1633–49.
14. Kawanabe T, Arikuzumi T, Kawai-Yamada M, Uchiyama H, Toriyama K. Abolition of the tapetum suicide program ruins microsporogenesis. Plant and Cell Physiol. 2006;47:784–7.
15. Parish RW, Li SF. Death of a tapetum: a programme of developmental altruism. Plant Sci. 2010;178:73–89.
16. Ja G, Liu X, Owen HA, Zhao D. Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. Proc Natl Acad Sci U S A. 2008;105:2220–5.
17. Zhang W, Sun YL, Timofejeva L, Chen C, Grossniklaus U, Ma H. Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM (DYT1) encoding a putative bHLH transcription factor. Development. 2006;133:3085–95.
18. Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, Guan YF, Yang ZN. Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. 2008;55:266–77.
19. Xu J, Yang C, Yuan Z, Zhang D, Gondwe MY, Ding Z, Li D, Wang D, Zhang D, Liu C, 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):3530–40.

20. Yang C, Vizcay-Barrena G, Conner K, Wilson ZA. Identification of tapetum-specific genes by digital gene-expression profiling. BMC Genomics. 2013;14:97.

21. Chen C, Farmer AD, Langley RI, Mudge J, Crow JA, May GD, Hunteley J, Smith AG, Retzel EF. Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. 2010;10:280.

22. Logacheva MD, Kastlanov AS, Vinogradov DM, Samigullin TH, Gelfand MS, Makeev VI, Pennin AK. De novo sequencing and characterization of floral transcriptome in two species of buckwheat (Fagopyrum). BMC Genomics. 2011;12:30.

23. Wei M, Song M, Fan S, Yu S. Transcriptomic analysis of differentially expressed genes during anther development in genetic maize male sterile and wild type cotton by digital gene-expression profiling. BMC Genomics. 2013;14:97.

24. Wu Y, Min L, Wu Z, Yang L, Zhu L, Yang X, Yuan D, Guo X, Zhang X. Polygalacturonases required for cell separation during reproductive development in Arabidopsis. Plant Physiol. 2015;173(4):1717–29.

25. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2 ΔΔCT method. Methods. 2001;25:402–8.

26. Kehoe A, Bennett J. OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, a regulator of cell separation. Plant Cell Physiol. 2009;50(1):136–46.

27. Smith AG, Retzel EF. Meiosis-specific gene discovery in plants: RNA-Seq analysis between the cytoplasmic male sterile line NJCMS1A and its fertile lines in watermelon. BMC Genomics. 2015;16:914.

28. Yan X, Dong C, Yu J, Liu W, Jiang C, Liu J, Hu Q, Fang X, Wei W. Comparative transcriptome analysis of differentially expressed genes in floral buds and flowers of male sterile and fertile lines in watermelon. BMC Genomics. 2015;16:914.

29. An H, Yang Z, Yi B, Wen J, Shen J, Tu J, Ma C, Fu T. Comparative transcriptomic profiling of the fertile and sterile flower buds of pol CMS in 8 Napus BMC Genomics. 2014;15:258.

30. Zhang L, Wang X, Yang J, Guo X, Li Y, Yang X, Wang J. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 2008;36(2):1547–51.

31. Birkenbihl RP, Diesel C, Sonnichsen F, Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. Plant Physiol. 2012;159(1):266–85.

32. Zhou J, Wang J, Zheng Z, Fan B, Yu JQ, Chen Z. Characterization of the promoter and extended C-terminal domain of Arabidopsis WRKY33 and functional analysis of tomato WRKY33 homologues in plant stress responses. J Exp Bot. 2015;66(15):4567–83.

33. Rutley N, Twell D. A decade of pollen transcriptomics. Plant Reprod. 2015;28:73–89.

34. Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X. Characterization of the sesame (Sesamum indicum L) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics. 2011;12:451.

35. Liu R, Yu C, Li Y, Lam TW, Yu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25(15):1966–8.

36. Lang F, Feng Z, Wang X, Wang Z. DEGseq: an R package for identifying differentially expressed genes from RNASeq data. Bioinformatics. 2010;26:36–8.

37. Cingolani P, Platts A, Le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.

38. Wang L, Yu J, Li D, Zhang X. Sinbase: an integrated database to study genomics, genetics and comparative genomics in Sesamum indicum. Plant Cell Physiol. 2015;56(1):e2.

39. Saeed B, Baranwal VK, Khurana P. Comparative transcriptomics and comprehensive marker resource development in mulberry. BMC Genomics. 2016;17:98.

40. Peng YH, Shih CF, Yang JY, Tan CM, Hsu WH, Huang YP, Liao PC, Yang CH. A RING-type E3 ligase controls anther dehiscence by activating the jasmonate biosynthetic pathway gene DEFECTIVE IN ANOTHER DEHISCENCE1 in Arabidopsis. Plant J. 2013;74(2):310–27.

41. Defective pollen wall contributes to male sterility in the male sterile line 1355A of cotton. Sci Rep. 2015;5:9608.

42. Bennett J. OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, a regulator of cell separation. Plant Cell Physiol. 2009;50(1):136–46.

43. Rutley N, Twell D. A decade of pollen transcriptomics. Plant Reprod. 2015;28:73–89.

44. Chen C, Farmer AD, Langley RI, Mudge J, Crow JA, May GD, Hunteley J, Smith AG, Retzel EF. Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated Arabidopsis male meiocytes. BMC Plant Biol. 2010;10:280.

45. Smith AG, Retzel EF. Meiosis-specific gene discovery in plants: RNA-Seq analysis between the cytoplasmic male sterile line NJCMS1A and its maintainer NJCMS1B in Soybean (Glycine max (L.) Merr.). PLoS One. 2015;10(5):e0126771.

46. Wang L, Feng Z, Wang X, Wang Z. DEGseq: an R package for identifying differentially expressed genes from RNASeq data. Bioinformatics. 2010;26:36–8.

47. Cingolani P, Platts A, Le Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.

48. Wang L, Yu J, Li D, Zhang X. Sinbase: an integrated database to study genomics, genetics and comparative genomics in Sesamum indicum. Plant Cell Physiol. 2015;56(1):e2.

49. Saeed B, Baranwal VK, Khurana P. Comparative transcriptomics and comprehensive marker resource development in mulberry. BMC Genomics. 2016;17:98.

50. Gao Y, Xu H, Shen Y, Wang J. Transcriptomic analysis of rice (Oryza sativa) endosperm using the RNA-Seq technique. Plant Mol Biol. 2013;81(4–5):363–8.

51. Conesa A, Gózzi S, Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics. 2008;2008:619832. doi:10.1155/2008/619832.

52. Ye J, Feng L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J. WEKO: a web tool for plotting GO annotations. Nucleic Acids Res. 2006;34(Web Server issue):W293–7.

53. Kehoe A, Bennett J. OsTDL1A binds to the LRR domain of rice receptor kinase MSP1, and is required to limit sporocyte numbers. Plant J. 2008;54(3):375–87.

54. Ogawa M, Kay P, Wilson S, Swain SM. ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are Polygalacturonases required for cell separation during reproductive development in Arabidopsis. Plant Cell. 2009;21(1):216–33.

55. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods. 2001;25:402–8.

56. Peng YH, Shih CF, Yang JY, Tan CM, Hsu WH, Huang YP, Liao PC, Yang CH. A RING-type E3 ligase controls anther dehiscence by activating the jasmonate biosynthetic pathway gene DEFECTIVE IN ANOTHER DEHISCENCE1 in Arabidopsis. Plant J. 2013;74(2):310–27.

57. Birkenbihl RP, Diesel C, Sonnichsen F, Arabidopsis WRKY33 is a key transcriptional regulator of hormonal and metabolic responses toward Botrytis cinerea infection. Plant Physiol. 2012;159(1):266–85.

58. Zhou J, Wang J, Zheng Z, Fan B, Yu JQ, Chen Z. Characterization of the promoter and extended C-terminal domain of Arabidopsis WRKY33 and functional analysis of tomato WRKY33 homologues in plant stress responses. J Exp Bot. 2015;66(15):4567–83.

59. Rutley N, Twell D. A decade of pollen transcriptomics. Plant Reprod. 2015;28:73–89.