

RESEARCH ARTICLE

Open Access



De novo transcriptome assembly from flower buds of dioecious, gynomonoecious and chemically masculinized female *Coccinia grandis* reveals genes associated with sex expression and modification

Ravi Suresh Devani¹, Sangram Sinha², Jayeeta Banerjee¹, Rabindra Kumar Sinha², Abdelhafid Bendahmane³ and Anjan Kumar Banerjee^{1*} 

Abstract

Background: *Coccinia grandis* (ivy gourd), is a dioecious member of Cucurbitaceae having heteromorphic sex chromosomes. Chromosome constitution of male and female plants of *C. grandis* is 22A + XY and 22A + XX respectively. Earlier we showed that a unique gynomonoecious form of *C. grandis* (22A + XX) also exists in nature bearing morphologically hermaphrodite flowers (GyM-H). Additionally, application of silver nitrate (AgNO₃) on female plants induces stamen development leading to the formation of morphologically hermaphrodite flowers (Ag-H) despite the absence of Y-chromosome. Due to the unavailability of genome sequence and the slow pace at which sex-linked genes are identified, sex expression and modification in *C. grandis* are not well understood.

Results: We have carried out a comprehensive RNA-Seq study from early-staged male, female, GyM-H, and Ag-H as well as middle-staged male and GyM-H flower buds. A de novo transcriptome was assembled using Trinity and annotated by BLAST2GO and Trinotate pipelines. The assembled transcriptome consisted of 467,233 'Trinity Transcripts' clustering into 378,860 'Trinity Genes'. Female_Early_vs_Male_Early, Ag_Early_vs_Female_Early, and GyM-H_Middle_vs_Male_Middle comparisons exhibited 35,694, 3574, and 14,954 differentially expressed transcripts respectively. Further, qRT-PCR analysis of selected candidate genes validated digital gene expression profiling results. Interestingly, ethylene response-related genes were found to be upregulated in female buds compared to male buds. Also, we observed that AgNO₃ treatment suppressed ethylene responses in Ag-H flowers by downregulation of ethylene-responsive transcription factors leading to stamen development. Further, GO terms related to stamen development were enriched in early-staged male, GyM-H, and Ag-H buds compared to female buds supporting the fact that stamen growth gets arrested in female flowers.

Conclusions: Suppression of ethylene responses in both male and Ag-H compared to female buds suggests a probable role of ethylene in stamen suppression similar to monoecious cucurbits such as melon and cucumber. Also, pollen fertility associated GO terms were depleted in middle-staged GyM-H buds compared to male buds indicating the necessity of Y-chromosome for pollen fertility. Overall, this study would enable identification of new sex-biased genes for further investigation of stamen arrest, pollen fertility, and AgNO₃-mediated sex modification.

Keywords: *Coccinia grandis*, Dioecious, Gynomonoecious, De novo transcriptome, Stamen arrest, Silver nitrate, Ethylene, Pollen fertility

* Correspondence: akb@iiserpune.ac.in

¹Biology Division, Indian Institute of Science Education and Research (IISER), Pune, Pune, Maharashtra, India

Full list of author information is available at the end of the article



Background

Monoecy, dioecy, and hermaphroditism are the three major sexual forms observed among the flowering plants. Ninety (90%) of angiosperms are found to be hermaphrodite (both male and female organs are in the same flower), while 5% plant species exhibit monoecy (male and female flowers are on the same plant) and remaining 5% show dioecy (male and female flowers are in separate plant) [1]. Dioecism provides a unique opportunity to study the genetic basis of sex determination. *Silene latifolia* (Caryophyllaceae), *Rumex acetosa* (Polygonaceae), *Carica papaya* (Caricaceae), *Spinacia oleracea* (Chenopodiaceae) and *Populus* (Salicaceae), have been well characterized to understand the mechanism of sex determination [2–5]. However, the molecular mechanism and the genes that govern sex determination are not well understood.

Coccinia grandis (L.) Voigt, a dioecious member of Cucurbitaceae family having an inferior ovary has received comparatively less attention. Members of Cucurbitaceae family exhibit variety of sexual forms [6]. Apart from its rich medicinal value, *C. grandis*, commonly known as ivy gourd, is also used as a vegetable. *Coccinia grandis* bears male and female unisexual flowers on separate plants. Similar to *Silene latifolia* (Caryophyllaceae), the sex in *Coccinia grandis* is determined by the presence of Y-chromosome [7–9]. The chromosome constitution of male and female plants is 22A + XY and 22A + XX respectively, where Y-chromosome is larger than the X-chromosome [10–12]. The male flower consists of three convoluted (bitheous) stamens [13, 14] and lacks female reproductive organs; however, the female flower consists of three rudimentary stamens surrounding the three fused carpels with an inferior ovary [14]. There are two ways by which unisexual flower development can be achieved. One of the ways is when both male and female sex organ primordia are initiated at early stages of flower development, but at later stages the opposite sex organs are aborted as in *Silene latifolia* [15]. Another way is that organ primordia of the opposite sex organs do not develop at all as shown in *Thalictrum dioicum* [16]. Also, there are flowers, wherein the inappropriate sex organs are retained in rudimentary form (instead of getting aborted) as in *Rumex* and *C. grandis* [2, 14]. Additionally, *Coccinia grandis* shows sex modification upon application of AgNO_3 leading to the development of stamens in female flower (such flower will be referred to as Ag-H) as described in our previous report [14]. Ag^+ has been long known to be an inhibitor of ethylene response [17]. It has been suggested that the binding of Ag^+ to the ethylene receptor inhibits the conformational change, which maintains the receptor in the active conformation [18]. Application of silver compounds such as silver nitrate (AgNO_3) or silver thiosulphate ($\text{Ag}_2\text{S}_2\text{O}_3$)

masculinizes monoecious plants such as *Cucumis sativus* as well as female plants of dioecious species such as *Silene latifolia* and *Cannabis sativa* [19–21]. However, the mechanism of action by which Ag^+ induces stamen development is not known till date [21].

Despite the interesting discovery of sex chromosomes in dioecious plants more than 50 years ago, the mechanism of sex determination remains poorly understood [22, 23]. This is primarily because of the slow pace at which sex-linked genes were identified from dioecious species (one to two genes/year) [24]. However, the improvement in NGS technology has already started changing the situation by accelerating the rate of sex-linked gene identification. The NGS-based approach has a big advantage that it does not require prior knowledge of the gene sequences to be investigated. Recently, an NGS-based RNA-Seq approach was applied to *Silene latifolia*, which was the first report that demonstrated the phenomenon of dosage compensation in plants [24]. A comparative transcriptomics approach was applied to papaya, a trioecious species, to identify the candidate genes for sex determination. This study led to the identification of 312 unique tags that were specifically mapped to the primitive sex chromosome (X or Y^{h}) sequences in papaya [5]. A genome-wide transcriptional profiling of apical tissue of a gynoecious mutant (Csg-G) and the monoecious wild-type (Csg-M) of cucumber was also performed to isolate genes involved in sex determination. This study revealed that genes involved in plant hormone signaling pathways, such as *ACS*, *Asr1*, *CsIAA2*, *CS-AUX1*, and *TLP*, and their crosstalk might play a critical role in the sex determination. Authors have also predicted the regulation of some transcription factors, including *EREBP-9*, in sex determination [25]. In another study, transcriptome sequencing was carried out from cucumber flower buds of two near-isogenic lines, WI1983G, a gynoecious plant which bears only pistillate flowers and WI1983H, a hermaphroditic plant which bears only bisexual flowers [26]. This study identified differentially expressed genes as well as putative SSR and SNP markers between flowers of two different sexes. T Akagi, IM Henry, R Tao and L Comai [27] sequenced genomic DNA, mRNA as well as small RNA from flower buds of persimmon and identified a Y-chromosome-encoded small RNA, *OGI*, that targets a homeodomain transcription factor *MeGI* regulating pollen fertility in a dosage-dependent manner. A recent de novo transcriptomics study in garden *Asparagus* identified 570 differentially expressed genes, where genes involved in pollen microspore and tapetum development were shown to be specifically expressed in males and supermales in contrast to females [28].

In order to identify sex-biased genes in *C. grandis* and to elucidate the mechanism of AgNO_3 mediated sex modification, a comprehensive RNA-Seq study from early-staged male (M), female (F), GyM-H and Ag-H as

well as middle-staged male and GyM-H flower buds was carried out. De novo transcriptome was assembled to identify *C. grandis* homologs of various flower development genes. Digital expression profiling was undertaken to identify sex-biased genes which might play a pivotal role in the arrest of stamen development in female flowers, genes that promote anther development in female flowers upon AgNO₃ treatment and genes controlling pollen fertility in male flowers.

Methods

Flower bud collection and RNA isolation

Clones of wild-type male, female and gynomonoecious (GyM) forms of *C. grandis* were grown in the experimental plot at IISER Pune, India. Gynomonoecious (GyM) plant bears pistillate flowers (GyM-F) as well as morphologically hermaphrodite flowers (GyM-H) (Herbarium Voucher: Tripura University Campus, Karmakar, 433). Foliar spray of 35 mM AgNO₃ solution to the basal leaves on some of the female plants led to the development of morphologically hermaphrodite flowers (Ag-H) as per our earlier observation [14]. Ag-H flower buds were morphologically similar to GyM-H flowers. Flower buds from male (M), female (F), GyM-H and Ag-H were harvested separately in liquid nitrogen and categorized into early and middle stages based on our previous study [14] (Fig. 1). In early-staged male flower buds, only stamens are present with no sign of carpel initials. Whereas, early-staged female flowers (stages 3–4, Fig. 1) have both carpel and stamen primordia. Stamen growth in female flowers gets arrested around stages 4–5. In the hermaphrodite flowers of gynomonoecious plant, however, both stamens and carpels develop simultaneously during early as well as

middle stages of development. Our selection of early-staged flower buds was carried out such that the event of stamen inhibition in female flowers can be analysed. Whereas, middle-staged flower buds were chosen such that meiosis-stage and pollen maturation event can be investigated. Total RNA was isolated by TRIzol reagent (Invitrogen) following the manufacturer's instructions. RNA quality was assessed using an Agilent Bioanalyzer RNA nanochip, and RNA samples with RIN > 8.0 were used for library preparation.

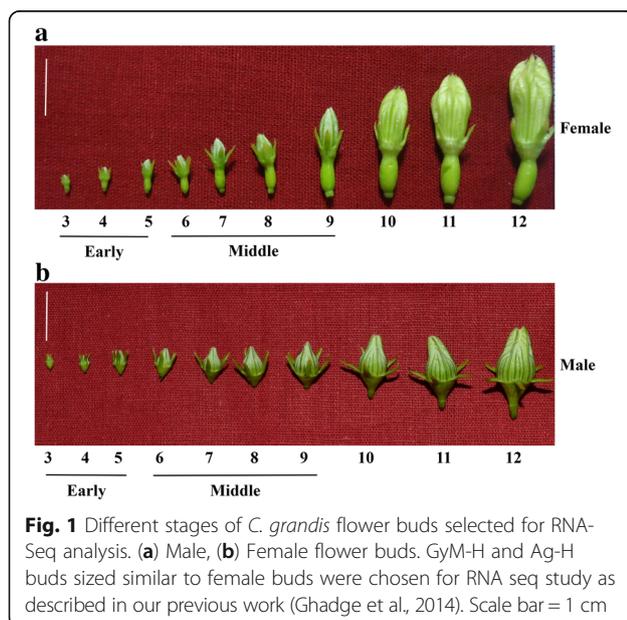
RNA-Seq library preparation and sequencing

Library preparation was performed at Genotypic Technology's Genomics facility, Bangalore using Illumina TruSeq RNA Sample Preparation Kit according to the manufacturer's specifications. RNA sequencing libraries were prepared in duplicate for early-staged male (M), female (F), GyM-H and Ag-H flower buds, as well as middle-staged male (M) and GyM-H flower buds. The quality of all the twelve libraries and insert size distribution was assessed using an Agilent High Sensitivity Bioanalyzer Chip. Libraries showed a peak in the range of 250–1000 bp. The effective sequencing insert size was 130–880 bp, and the inserts were flanked by adaptors whose combined size was 120 bp. Libraries were quantified using Qubit and sequenced on Illumina NextSeq 500 platform, producing 2 X 150-nucleotide paired-end reads. RNA-Seq data generated in this study has been deposited in the NCBI SRA study SRP111347.

Pre-processing of Illumina reads and de novo transcriptome assembly

Raw RNA-Seq reads were processed using Trimmomatic v0.33 for trimming adapters as well as low-quality bases from ends of the reads [29]. Poor quality reads with average Phred quality score < 20 and reads with length < 36 were also filtered out. The resulting set of good quality reads were then assembled with Trinity v2.1.1 software using default parameters [30, 31].

The quality of the resulting assembly was assessed by various methods. First of all, RNA-Seq read representation of the assembly was checked using bowtie2 [32]. Ex90N50 transcript contig length (the contig N50 value based on the set of transcripts representing 90% of the expression data) was computed using *contig_ExN50_statistic.pl* script bundled with Trinity. Then the representation of full-length reconstructed protein-coding genes was studied. The assembled transcripts were compared with Swiss-Prot using BLAST and the hits were analyzed using a perl script *blast_outfmt6_group_segments.tophit_coverage.pl*, provided with the Trinity package. BUSCO (Benchmarking Universal Single-Copy Orthologs) was used to explore completeness of the transcriptome according to conserved ortholog content [33]. Finally, TransRate was used to



compare the assembly to the publicly available *Cucumis sativus* protein-coding primary transcript sequences [34].

Annotation of the de novo-assembled transcripts

The de novo-assembled transcripts were compared with the viridiplantae sequences from nr and Swiss-Prot database using BLASTX with an e-value threshold of $1e^{-3}$ [35]. BLAST output generated from this comparison was loaded into BLAST2GO for mapping GO terms to the transcripts and annotation [36]. Enzyme codes and KEGG pathway mapping were also carried out. ANNEX (Annotation Expander) was used to enhance the annotations. Finally, GO-Slim mapping was applied to get a broad overview of the ontology content.

In addition to BLAST2GO facilitated annotation, Trinotate pipeline was used to carry out comprehensive functional annotation of the transcripts leveraging various annotation databases (eggNOG/GO/KEGG databases) [31]. Trinotate pipeline also included identification of open reading frames, homology search against Swiss-Prot and TrEMBL. Protein domain identification was carried out using HMMER/PFAM. Protein signal peptide and transmembrane domains were predicted by signalP and tmHMM respectively.

Transcript quantification and differential expression analysis

align_and_estimate_abundance.pl script from Trinity package was applied to align cleaned reads from each library to the de novo transcriptome using bowtie and to estimate the transcript abundance using RSEM [37]. *abundance_estimates_to_matrix.pl* script was used to construct a matrix of counts and a matrix of normalized expression values. *PtR* script was used to generate correlation matrix and Principal Component Analysis (PCA) plot for comparing replicates across all the samples. Differential expression analysis was carried out with two biological replicates from the count matrix using *run_DE_analysis.pl* with edgeR as the method of choice [38]. *analyze_diff_expr.pl* script was used to examine GO enrichment and to extract all transcripts that had *p*-values at most $1e^{-3}$ and were at least 2^2 fold differentially expressed. The DE features were partitioned into clusters with similar expression patterns by *define_clusters_by_cutting_tree.pl* script with Ptree method.

Validation of differentially expressed genes by qRT-PCR

For expression analysis, qRT-PCR was carried out using aliquots of the same RNA samples that were used for RNA sequencing. Two micrograms (2 μ g) of total RNA was used for complementary DNA (cDNA) synthesis by SuperScript IV reverse transcriptase (Invitrogen) using an oligo(dT) primer. *CgACT2* gene was used as reference gene for normalization. qRT-PCR was performed on BIO-RAD CFX96 machine with gene-specific

forward and reverse primers (Additional file 1: Table S1). The reactions were carried out using Takara SYBR Premix Ex Taq II (Takara Bio Inc.) and incubated at 95 °C for 3 min followed by 40 cycles of 95 °C for 15 s, 58 °C for 15 s and 72 °C for 15 s. PCR specificity was checked by melting curve analysis, and data were analysed using the $2^{-\Delta\Delta CT}$ method [39].

Results

RNA sequencing, Trinity-based de novo transcriptome assembly and annotation using BLAST2GO and Trinotate

A total of 306,575,536 paired-end reads (150 bp) were obtained after sequencing all the twelve libraries on the Illumina NextSeq 500 platform. Subsequently, 186,399,131 good quality paired-end reads were used for de novo assembly of *Coccinia grandis* flower bud transcriptome using Trinity software package with default parameters (Table 1). The resulting assembly consisted of 467,233 ‘Trinity Transcripts’ clustering into 378,860 ‘Trinity Genes’ with an N50 value of 881 bp (Table 2, Additional file 2). The transcripts of 200–399 bp size were found to be most abundant in the length distribution of assembled transcripts (Fig. 2). However, a higher proportion of transcripts with length around 1000–2000 bp had a BLAST hit compared to the proportion of smaller transcripts (Fig. 2). Cleaned reads were mapped back to the transcriptome using bowtie2 with ~70% or more reads from each library aligning concordantly (Table 1). An Ex90N50 statistic calculated using 80,806 transcripts from the assembly (ignoring the rest of the transcripts with poor read coverage) was found to be 1784 bp (Additional file 3: Figure S1).

Altogether, 8916 unique BLAST hits in the Swiss-Prot database were represented by nearly full-length transcripts, having more than 70% alignment coverage, and 12,315 hits showed more than 50% alignment coverage (Additional file 4: Table S2). BUSCO output [C:89.8%(S:14.5%,D:75.3%),F:5.0%,M:5.2%,n:1440] showed that out of 1440 BUSCOs for Plants dataset, 1293 full length BUSCOs were detected in our de novo-assembled *Coccinia grandis* flower bud transcriptome indicating 89.8% completeness. Finally using TransRate, we were able to detect *C. grandis* homologs for 84% (18,039) of protein-coding primary transcripts of *C. sativus* of which, 13,430 reference sequences had at least 95% of their bases covered by a CRB-BLAST hit (Additional file 5: Table S3).

Coccinia grandis flower bud transcripts were compared to plant protein sequences of the nr and Swiss-Prot databases resulting in 259,200 and 136,663 transcripts having at least one hit from the respective database. Species distribution analysis of the BLAST hits showed that majority of these hits were from *Arabidopsis* and rice for Swiss-Prot database whereas for nr database most top hits were from cucumber and melon (Fig. 3a,b).

Table 1 RNA sequencing read counts and alignment statistics for all the samples used for de novo transcriptome assembly

Sample name	Raw Reads	Cleaned reads	% Read pairs mapped concordantly
Male Early A	21,719,110	13,216,446	76.33%
Male Early B	22,080,977	12,149,633	74.84%
Female Early A	25,607,195	15,634,598	76.34%
Female Early B	25,433,955	15,316,996	75.94%
GyM-H Early A	27,936,206	17,378,706	76.67%
GyM-H Early B	27,617,808	17,107,330	76.69%
Ag-H Early A	26,147,527	15,140,736	71.02%
Ag-H Early B	25,392,540	15,128,539	69.35%
Male Middle A	25,502,209	15,490,027	76.70%
Male Middle B	25,837,400	15,919,727	75.27%
GyM-H Middle A	28,465,770	17,441,923	76.90%
GyM-H Middle B	24,834,839	16,474,470	77.43%
Total	306,575,536	186,399,131	

The number of transcripts annotated with various GO terms of biological process, molecular function, and cellular component categories are provided in Fig. 3c.

Trinotate v3 pipeline was also used simultaneously for comprehensive functional annotation of the *Coccinia grandis* flower bud transcripts. Details regarding the Swiss-Prot/TrEMBL BLAST hits, GO, KEGG and eggNOG mappings can be found in the Additional file 6: Table S4. HMMER/PFAM predicted protein domains, as well as information regarding signal peptides and transmembrane domains could also be found in the Additional file 6: Table S4. Taken together, we have assembled a good quality transcriptome for early- and middle-staged flower buds of *Coccinia grandis* and comprehensively annotated the transcripts using well-established BLAST2GO and Trinotate pipelines.

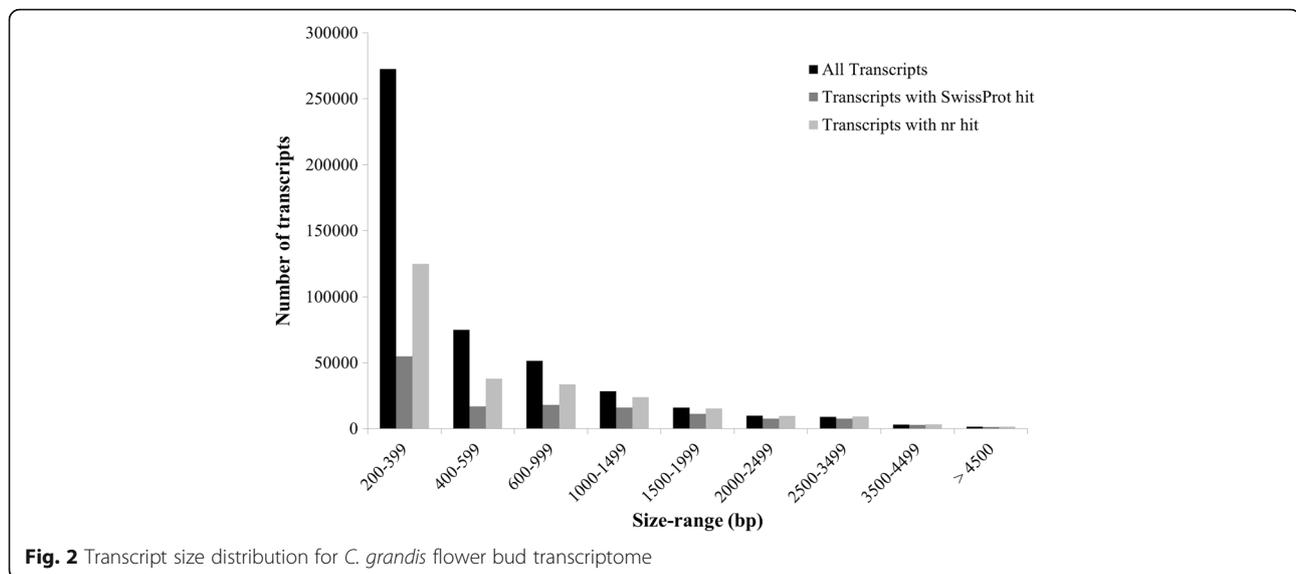
Differential expression analysis reveals probable factors for pollen fertility and sex modification

Following transcriptome assembly and annotation, differential expression analysis was carried out. First of all, RSEM was used for transcript abundance estimation.

Table 2 Assembly statistics for *C. grandis* flower bud transcriptome

Parameter	Assembly statistics
Number of Trinity Transcripts'	467,233
Number of Trinity Genes'	378,860
Percent GC	38.96
Median contig length (bp)	347
Average contig length (bp)	606.45
N50 (bp)	881
Total assembled bases	283,354,298

Following which, we checked for the correlation between the replicates for all the samples using *PtR* script. PCA analysis and correlation matrix showed a good correlation between the replicate sets for each of the six samples (Fig. 4). EdgeR was used to identify the differentially expressed transcripts for all the pairwise comparisons between the six samples (Table 3; Fig. 5; Additional file 7: Figure S2). Differentially expressed transcripts at a minimum fold change of 2² with *p*-values at most 1e-3 were extracted and GO enrichment analysis was performed (Additional file 8: Table S5, Additional file 9: Table S6). Among all the comparisons, few interesting ones such as *Ag_Early_vs_Female_Early*, *Female_Early_vs_Male_Early*, and *GYM-H_Middle_vs_Male_Middle* had 3574, 35,694 and 14,954 differentially expressed transcripts respectively (Table 3, Fig. 5). The DE features were partitioned into clusters with similar expression patterns (Fig. 6; Additional file 10: Figure S3). In the context of anther development, we identified several GO terms (GO:0080110, GO:0010208, GO:0010584, GO:0009555, GO:0055046, GO:0048658, GO:0048653) differentially enriched in male buds compared to female buds at an early stage of floral development. qRT-PCR was done to validate the results of differential expression analysis for a few interesting *Coccinia* homologs of *AMS* (ABORTED MICROSPORES), *CER3* (ECERIFERUM 3), *DEX1* (DEFECTIVE IN EXINE FORMATION 1), *DYT1* (DYSFUNCTIONAL TAPETUM 1), *EIL1* (ETHYLENE INSENSITIVE 3-like 1), *EMSI* (EXCESS MICROSPORO CYTES 1), *FER* (FERONIA), *MMD1* (MALE MEIOCYTE DEATH 1), *MS1* (MALE STERILITY 1), *SHT* (Spermidine hydroxycinnamoyl transferase), *TPD1* (TAPETUM DETERMINANT 1) and *ZAT3* (Zinc finger protein ZAT3). Expression profiles for these genes deduced by qRT-PCR revealed similar patterns to that seen from the digital DE analysis results (Fig. 7). Also, we have found GO terms related to pollen fertility enriched in the male buds (GO:0080092, GO:0009846, GO:0009860) compared to GyM-H and Ag-H buds, which had sterile pollens. Accordingly, expression profile for homologs of a number of genes involved in pollen tube development such as *CSLD1* (Cellulose synthase-like protein D1; TRINITY_DN92683_c0_g1_i1), *CDPKO* (Calcium-dependent protein kinase 24; TRINITY_DN93671_c0_g1_i3), *PME4* (Pectin methylesterase 4; TRINITY_DN14239_c0_g1_i1), *PME37* (Pectin methylesterase 37; TRINITY_DN3663_c0_g1_i1), *PPME1* (POLLEN SPECIFIC Pectin methylesterase 1; TRINITY_DN66415_c0_g1_i1) and *PTR52* (Protein NRT1/PTR FAMILY 2.8; TRINITY_DN112735_c0_g14_i3) were analysed and found to be enriched in middle-staged male buds similar to our digital expression profiles based on RNA-Seq data (Fig. 8). Downregulation of ethylene signalling upon AgNO₃ treatment was evident as GO:0009723 (response to ethylene) and



GO:0009873 (ethylene-activated signaling pathway) were depleted in AgNO₃ treated plant (Additional file 9: Table S6). In order to validate this, we studied the expression profile of *Ethylene-responsive transcription factors*, *ERF5*, *ERF17*, and *EF102*. We found that all three *ERFs* were downregulated upon AgNO₃ treatment (Fig. 9).

Discussion

Genetic basis of sex determination and differentiation is not well studied in *C. grandis*. Identification and investigation of sex-linked genes would lead to better understanding of dioecy in plants and this can be achieved by whole genome sequencing approach. However, sex-determining genes are most likely linked to non-recombining regions of Y-chromosome, which are difficult to assemble from sequence data [40]. An alternative approach is to use comparative transcriptomics to identify sex-biased genes that could play a role in sex differentiation and determination [24]. Further, the presence of mutations and SNPs in sex-biased genes can provide insights regarding the evolution of dioecy. Using this approach, we have assembled and annotated a de novo transcriptome from the flower buds of dioecious, gynomonocious and AgNO₃ treated female *C. grandis*. We have identified differentially expressed genes which might be playing a role in stamen arrest of female flowers. Also, we have analysed the genes that were differentially expressed upon AgNO₃ treatment on female plants promoting stamen development. Finally, we have compared middle-staged male (bearing fertile pollens) and GyM-H buds (bearing sterile pollens) to study the genes involved in pollen maturation and fertility.

Differential expression of stamen developmental genes and arrest of stamen growth in female flowers

At the early stages (stages 3–4) of flower development in female *C. grandis*, both carpel and stamen organs are initiated simultaneously. However, stamen growth gets arrested during the course of development (stages 4–5) resulting in a female flower with rudimentary stamens. In contrast, no sign of carpel primordia was observed during the histological study of flower development in male *C. grandis* as described in our previous report [14]. The molecular players involved in stamen initiation and development process are well characterized in the hermaphrodite plant *Arabidopsis*. In order to identify the stage at which stamen growth gets arrested, *Coccinia grandis* homologs of *Arabidopsis* stamen development genes were identified from the de novo-assembled transcriptome. Among genes involved in stamen initiation, *Pistillata* (*CgPI*, TRINITY_DN71631_c0_g1_i1) was found to be expressed in a male-biased fashion (Additional file 8: Table S5). *Pistillata* has been shown to specify stamen identity in *Arabidopsis* [41] (Table 4). Further, EXCESS MICROSPOROCTES 1 (*EMS1*) has been shown to interact with TAPETUM DETERMINANT 1 (*TPD1*) regulating specification of reproductive as well as somatic cells in *Arabidopsis* anthers [42]. Differential expression analyses revealed that homologs of both *EMS1* (TRINITY_DN106236_c0_g4_i1) and *TPD1* (TRINITY_DN116795_c2_g1_i3) were enriched in male flowers compared to female flowers (Table 4; Additional file 8: Table S5; Fig. 7). *DYSFUNCTIONAL TAPETUM 1* (*DYT1*) plays an important role in tapetum development by regulating the expression of *DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION 1* (*TDF1*) in *Arabidopsis* [43]. Also, *DYT1* is known to interact with Basic helix-loop-

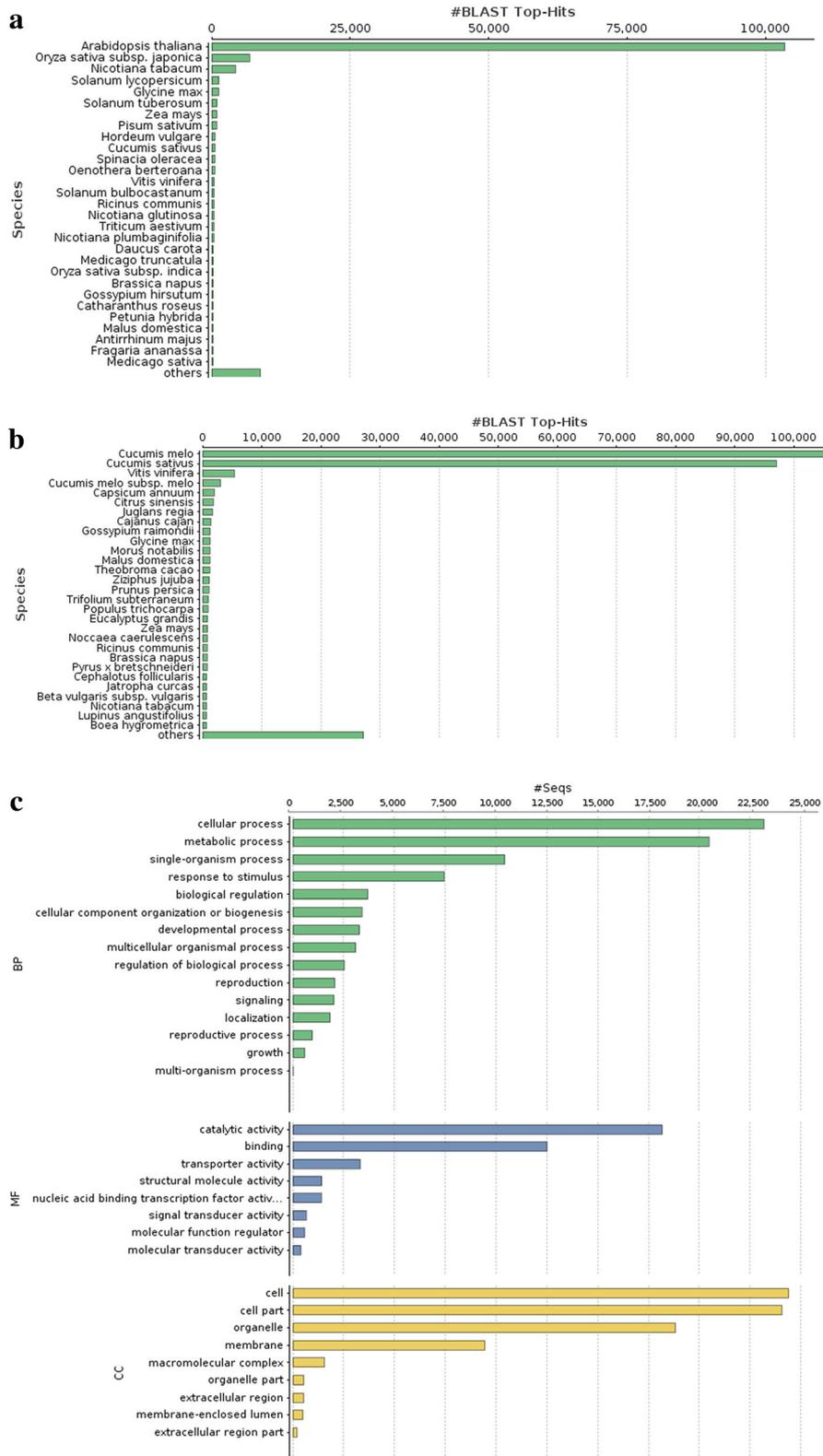
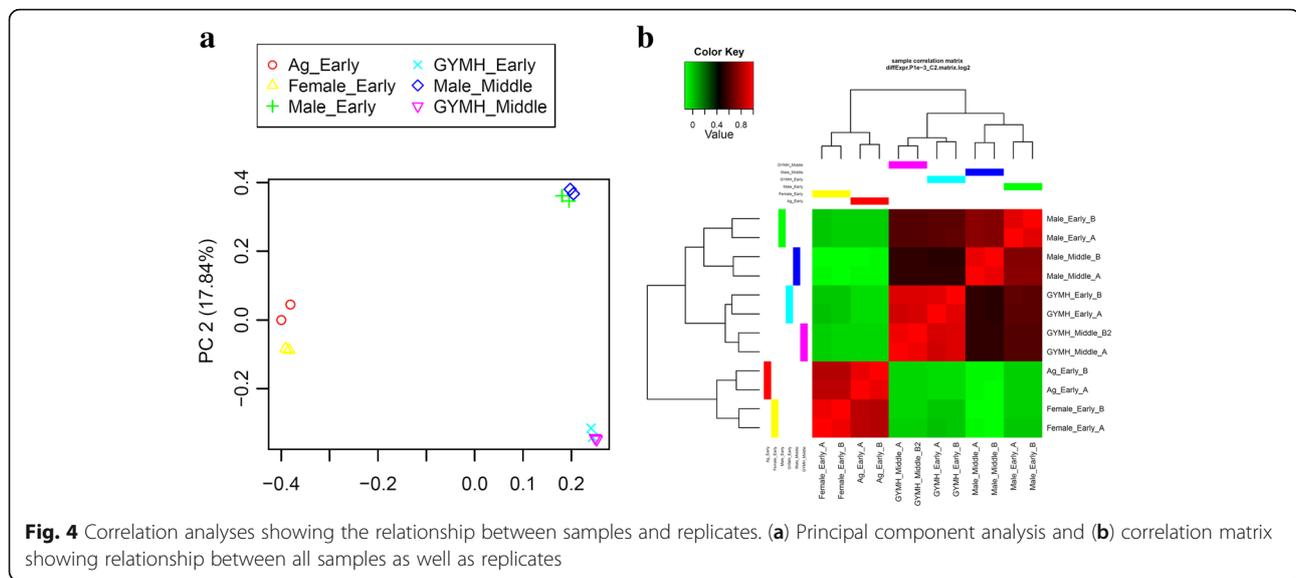


Fig. 3 BLAST2GO annotation of *C. grandis* flower bud transcriptome. **(a)** BLAST Top-Hits species distribution when compared with Swiss-Prot database, **(b)** BLAST Top-Hits species distribution when compared with nr database, **(c)** GO category distribution of *C. grandis* flower bud transcriptome



helix protein 89 (bHLH89) which is highly expressed in anthers and required for normal anther development and male fertility [44]. *TDF1* homolog (TRINITY_DN97604_c1_g7_i1) as well as *bHLH89* homolog (TRINITY_DN85771_c0_g1_i1) showed male-biased expression in *C. grandis* (Table 4; Additional file 8: Table S5). Differential regulation of these genes related to stamen development explains the possible cause for early stamen arrest in female flowers of *C. grandis*.

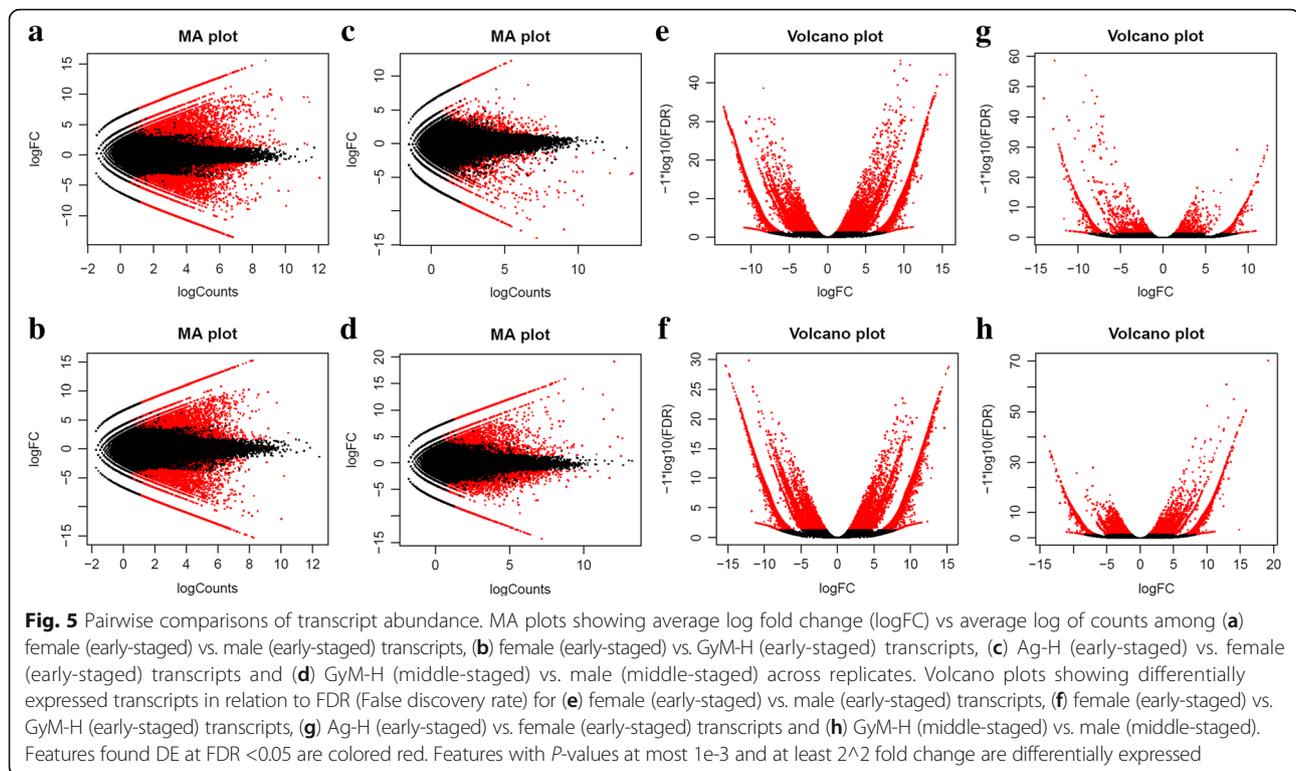
Table 3 Number of differentially expressed transcripts for each pairwise comparison between the flower types. Transcripts that had p-values at most $1e-3$ and were at least $2^{\wedge}2$ fold were considered as differentially expressed

Flower Buds Comparison	Number of DE transcripts
Ag_Early_vs_Female_Early	3574
Ag_Early_vs_GYM_H_Early	34,458
Ag_Early_vs_GYM_H_Middle	38,849
Ag_Early_vs_Male_Early	33,863
Ag_Early_vs_Male_Middle	36,923
Female_Early_vs_GYM_H_Early	31,886
Female_Early_vs_GYM_H_Middle	39,885
Female_Early_vs_Male_Early	35,694
Female_Early_vs_Male_Middle	40,477
GYM_H_Early_vs_GYM_H_Middle	816
GYM_H_Early_vs_Male_Early	8659
GYM_H_Early_vs_Male_Middle	11,576
GYM_H_Middle_vs_Male_Early	12,357
GYM_H_Middle_vs_Male_Middle	14,954
Male_Early_vs_Male_Middle	4427

According to recent reports from monoecious cucurbits like melon, cucumber, and watermelon, ethylene plays a major role in sex determination by inhibiting stamen development process [45–48]. We found that compared to male, GO:0009723 (response to ethylene) was enriched in female buds indicating a potential role of ethylene in sex determination of *C. grandis* (Additional file 9: Table S6).

AgNO₃ treatment on female plant releases the stamen inhibition

Female plants of *C. grandis* bear flowers with fused carpels and rudimentary stamens. Earlier, we have shown that foliar spray of 35 mM AgNO₃ on the female plant of *C. grandis* promotes further development of the rudimentary stamens [14]. In the current study, gene expression profiles for early-staged Ag-H flower buds were compared with female buds (Table 3; Fig. 5c, g). Ag⁺ ions are known to inhibit responses to ethylene, a gaseous plant hormone [17]. Also, silver compounds have been shown to induce maleness by promoting stamen development in many monoecious and dioecious species [19–21]. No other inhibitors of ethylene biosynthesis or signaling could induce the stamen development in *Silene latifolia*, suggesting that ethylene signaling might not be the only pathway that gets affected upon application of silver thiosulphate [21]. In contrast to *Silene latifolia*, AVG (aminoethoxyvinylglycine), an inhibitor of ethylene-biosynthesis has been shown to induce male flowers in gynoeious muskmelon similar to silver compounds [49]. Considering the role of 1-aminocyclopropane-1-carboxylate synthase (ACS, an enzyme involved in ethylene biosynthesis) in sex determination of many other members of Cucurbitaceae, an ethylene-mediated effect of



AgNO₃ seems more likely to be involved in the modification of sex in *C. grandis* [50].

In our study, GO:0009723 (response to ethylene) and GO:0009873 (ethylene-activated signaling pathway) were enriched in female buds compared to Ag-H buds (Additional file 9: Table S6). Transcripts for genes such as *Ethylene-responsive transcription factors*, *ERF5* (TRINITY_DN102355_c3_g13_i1), *ERF17* (TRINITY_DN80749_c0_g6_i1), *EF109* (TRINITY_DN8704_9_c0_g1_i1), *EF102* (TRINITY_DN90257_c1_g2_i1), *ERF99* (TRINITY_DN93821_c0_g1_i2), *ERF60* (TRINITY_DN93262_c1_g6_i2) and *ERF78* (TRINITY_DN985_03_c3_g1_i1) were downregulated in Ag-H buds indicating impaired ethylene signalling (Additional file 8: Table S5). Additionally, qRT-PCR based expression pattern analysis for *ERF5*, *ERF17* and *EF102* genes clearly showed the suppression of ethylene responses by AgNO₃ (Fig. 9).

Downregulation of ethylene signaling in Ag-H buds was correlated with the promotion of stamen growth. GO:0048655 (anther wall tapetum morphogenesis), GO:0048657 (anther wall tapetum cell differentiation), GO:0048658 (anther wall tapetum development) were seen to be enriched in early-staged Ag-H buds compared to female buds (Additional file 9: Table S6). *C. grandis* homologs of *MS1*, *MMD1* (TRINITY_DN109512_c4_g3_i1, TRINITY_DN108927_c0_g6_i1), *ZAT3* (TRINITY_DN10

8658_c0_g2_i1) and *AMS* (TRINITY_DN116105_c0_g2_i1) genes which play important roles in tapetum and pollen development of *Arabidopsis* flowers were upregulated upon AgNO₃ treatment indicating promotion of stamen growth [51–56] (Additional file 8: Table S5; Fig. 7). *MYB35* (TRINITY_DN92649_c0_g7_i1), which was proposed as a putative sex-determining gene in *Asparagus* was also found to be upregulated in Ag-H buds [57] (Additional file 8: Table S5). Apart from that, gene ontology terms related to pollen wall assembly (GO:0010208), pollen exine formation (GO:0010584), sporopollenin biosynthetic process (GO:0080110), pollen development (GO:0009555) and pollen sperm cell differentiation (GO:0048235) were also enriched in Ag-H buds (Additional file 9: Table S6). Further, we noticed that *Ethylene-responsive transcription factors* (ERFs) were not affected in GyM-H buds as compared to female buds suggesting that stamen development in GyM-H flower buds might be regulated by some other mechanism evading ethylene signaling inhibition.

Transcripts governing pollen fertility are depleted in GyM-H and Ag-H flower buds

C. grandis is one of the few species in which the presence of heteromorphic sex chromosomes is reported. The large Y-chromosome present in males might play a major role in sex determination. The GyM form of *C.*

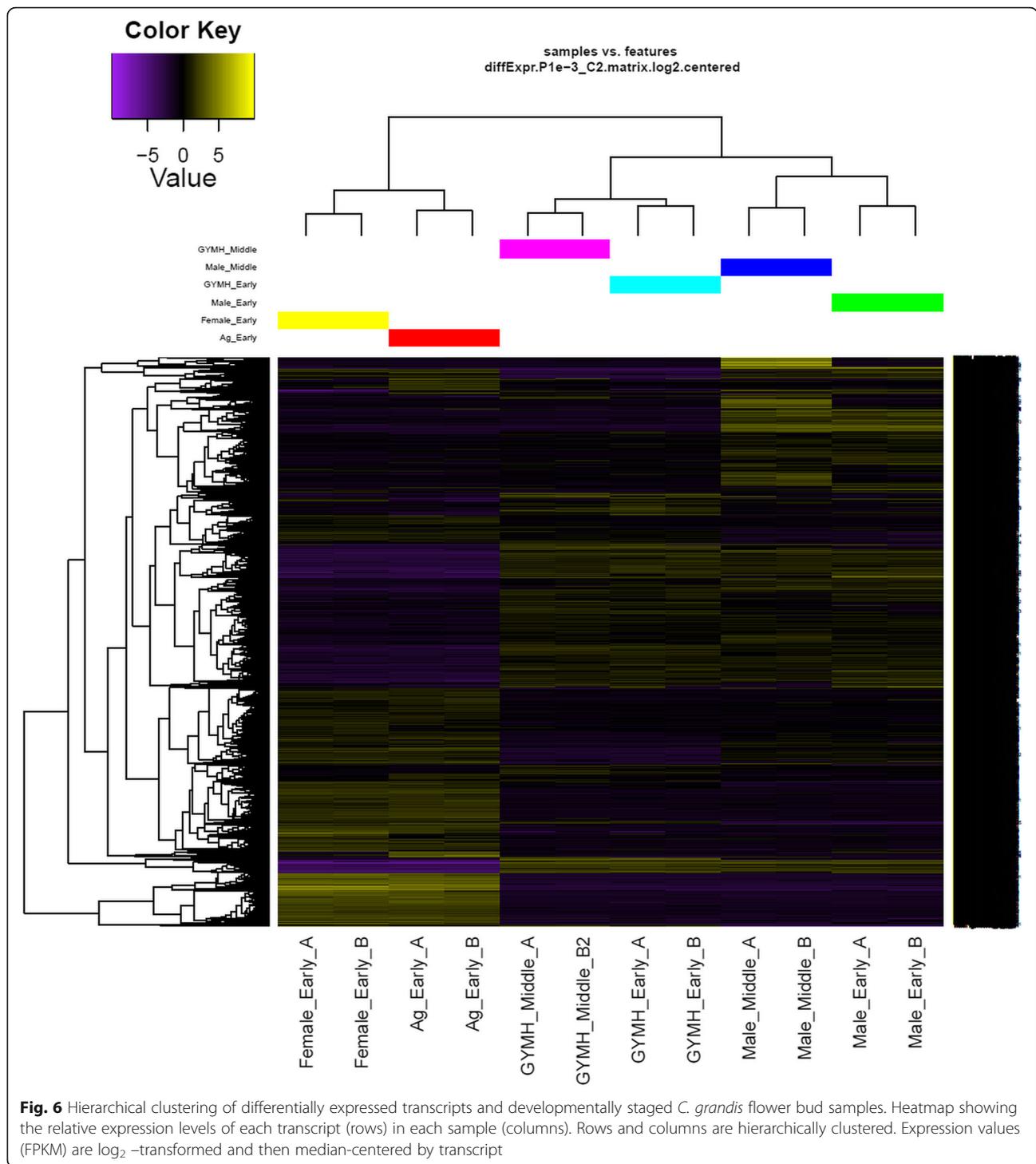
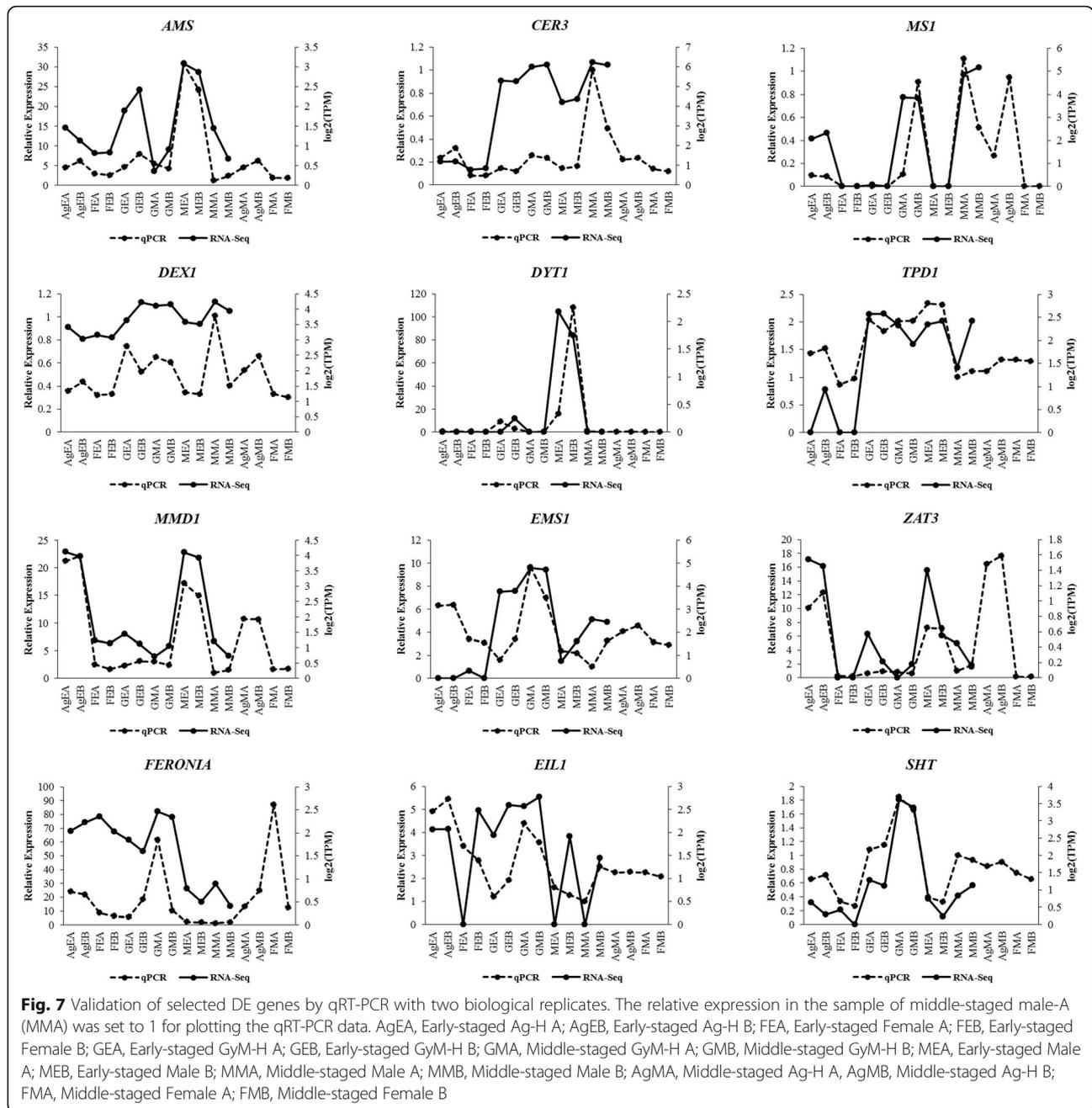


Fig. 6 Hierarchical clustering of differentially expressed transcripts and developmentally staged *C. grandis* flower bud samples. Heatmap showing the relative expression levels of each transcript (rows) in each sample (columns). Rows and columns are hierarchically clustered. Expression values (FPKM) are \log_2 -transformed and then median-centered by transcript

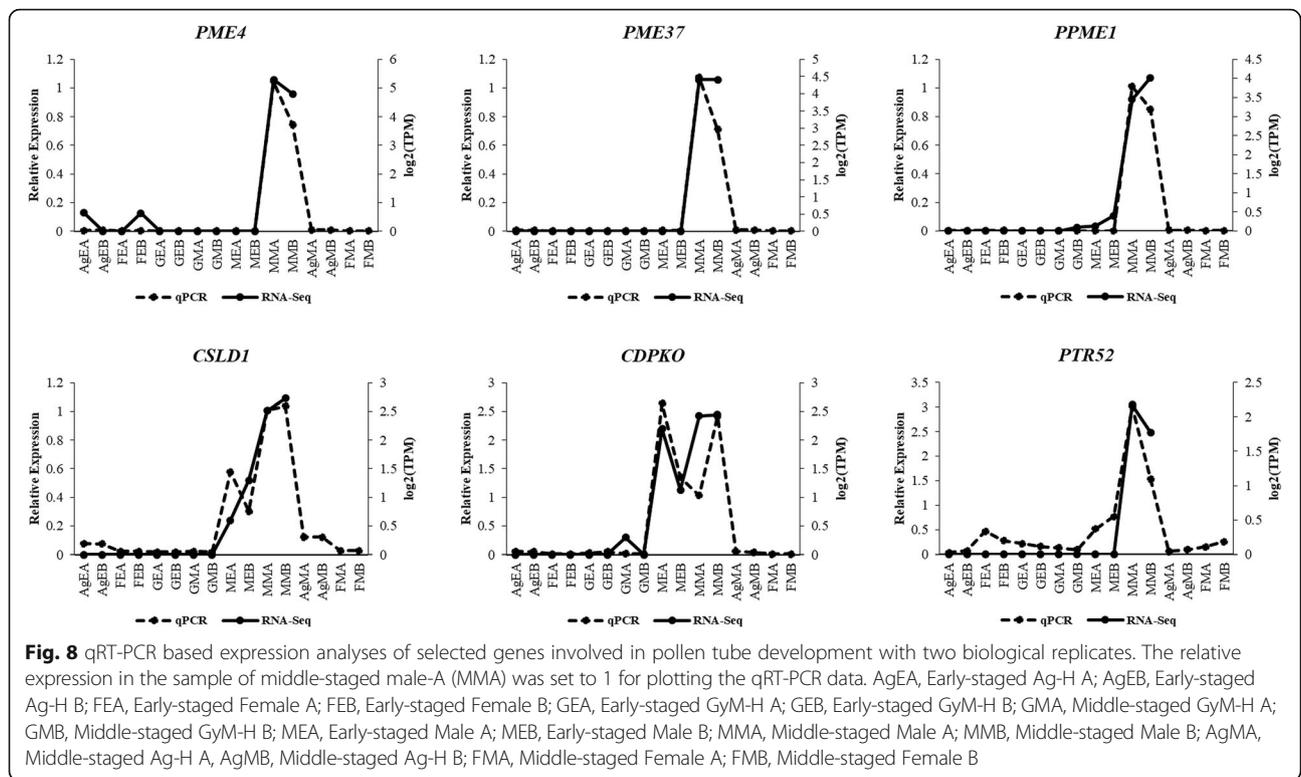
grandis included in the current study does not have Y-chromosome [14]. GyM-H flowers still develop full-sized stamens despite lacking Y-chromosome. Similarly, AgNO₃ treatment induces stamen development in female plants having XX sex chromosomes. However, the pollens from GyM-H and Ag-H flowers buds were found to be sterile unlike the pollens from male buds [14].

Differential expression analysis revealed that gene ontology terms for pollen tube (GO:0090406), pollen germination (GO:0009846), regulation of pollen tube growth (GO:0080092), pollen tube growth (GO:0009860) and microsporogenesis (GO:0009556) were enriched in middle-staged male buds compared to middle-staged GyM-H buds (Additional file 9: Table S6).



GAUTE plays an important role in pollen tube wall biosynthesis in *Arabidopsis* [58]. *TRINITY_DN111340_c1_g1_i6*, which showed similarity with *GAUTE* was enriched in male buds compared to GyM-H buds. Unlike most other plant cell walls, pollen tube wall does not contain callose or cellulose. Pectin methylsterases (PMEs) have been shown to play a very important role in the growth of pollen tubes [59–61]. *PME4* (*TRINITY_DN14239_c0_g1_i1*), *PME37* (*TRINITY_DN3663_c0_g1_i1*) and *PPME1* (*TRINITY_DN66415_c0_g1_i1*, *TRINITY_DN71598_c0_g2_i1*) were downregulated in GyM-H buds

compared to male buds (Additional file 8: Table S5). This could be a possible cause for pollens from GyM-H not forming pollen tubes. H Zhan, Y Zhong, Z Yang and H Xia [62] has shown that *IPMKB* (Inositol polyphosphate multikinase beta) is an important factor for pollen development. We have found that *TRINITY_DN96290_c0_g3_i2* transcript matching to *Arabidopsis IPMKB (Atlpk2beta)* was downregulated in GyM-H compared to male buds. Earlier, several reports have demonstrated that *MALE STERILITY 1 (MS1)* gene of *Arabidopsis* expresses in tapetal cells and plays



an important role in pollen maturation [51, 54, 55]. *C. grandis* homolog of *MS1*, TRINITY_DN109512_c4_g3_i1 was expressed in a male-biased manner (Additional file 8: Table S5; Fig. 7). Similarly, homologs of genes important for pollen tube growth such as *CSLD1* (TRINITY_DN92683_c0_g1_i1), *CDPKO* (TRINITY_DN93671_c0_g1_i3), *NRX1* (TRINITY_DN106708_c1_g2_i3), *PTR52* (TRINITY_DN112735_c0_g14_i3; TRINITY_DN112735_c0_g3_i1), *TAF6* (TRINITY_DN96231_c1_g1_i2) and *CALS5* (TRINITY_DN113564_c1_g1_i1) were enriched in male [63–68] (Fig. 8, Additional file 8: Table S5). Genes involved

in pollen exine formation such as *FACR2/MS2* (TRINITY_DN74585_c1_g5_i3), *EA6* (TRINITY_DN76274_c1_g1_i1), *C70A2/DEX2* (TRINITY_DN99059_c0_g1_i1) were also upregulated in male [69–71]. *EMS1* (TRINITY_DN89942_c0_g7_i1), *SERK1* (TRINITY_DN108624_c1_g7_i5), *JASON* (TRINITY_DN83440_c0_g1_i1), *RPK2* (TRINITY_DN113423_c0_g1_i4), which are essential for microsporogenesis and pollen maturation were observed to be expressed at significantly higher levels in middle-staged male buds compared to GyM-H buds. [72–75] (Additional file 8: Table S5; Fig. 7).

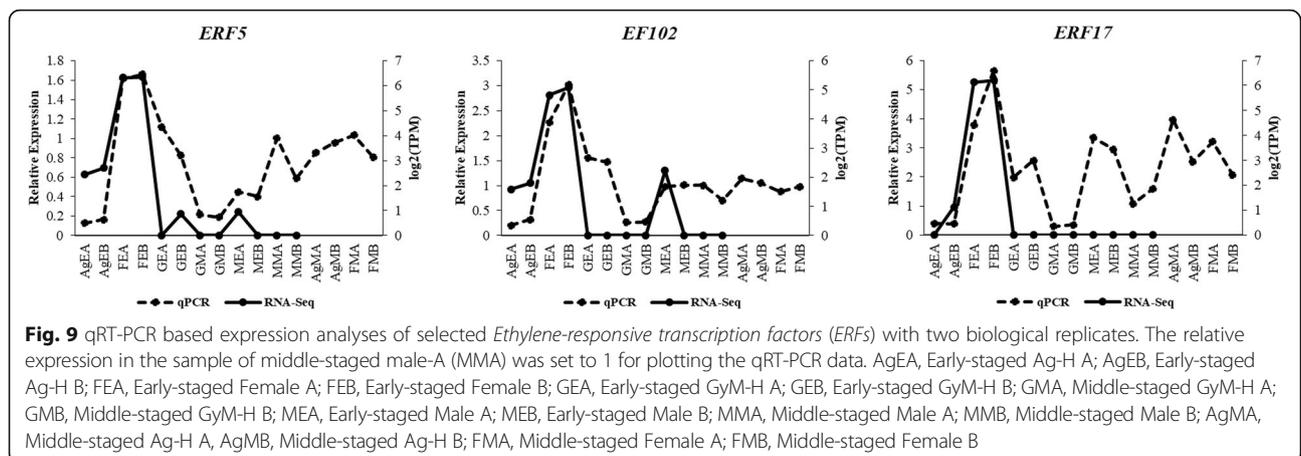


Table 4 Digital Expression profile for genes in anther developmental pathway

Stage of stamen development	Genes	Expression pattern
Stamen Primordia Initiation	AG	Unbiased
	CLV1/CLV2	Unbiased/male-biased
	PI	Male-biased
	AP3	Unclear homolog
	JAG	Male-biased
Archeporsial initiation	BAM1/BAM2	Unclear homolog
	SPL/NZZ	Unclear homolog
Tapetal Development	EMS1	Male-biased
	SERK1/2	Male-specific
	TPD1	Male-biased
	RPK2	Male-specific
	TDF	Male-biased
	DYT1	Unbiased
	bHLH89	Male-biased
	Mature Pollen Formation	AMS
MS1		Male-biased
MS2		Male-biased
MIA		Unbiased
LAP3		Unbiased
LAP5	Male-biased	

Expression profiling for homologs of *MS1*, *EMS1*, *DYT1*, *PME4*, *PME37*, *PPME1*, *CSLD1*, *CDPKO*, and *PTR52* was studied by qRT-PCR for all the tissue samples including middle-staged Ag-H buds (Figs. 7 and 8; Table 4). Transcripts for all these homologs were downregulated in Ag-H buds and GyM-H buds, suggesting a male-biased expression pattern implicating the reason for pollen sterility in Ag-H and GyM-H buds.

Conclusions

De novo-assembled transcriptome developed from RNA-Seq of different sexual phenotypes has enabled identification of *C. grandis* homologs of many genes known to be involved in flower development in species such as *Arabidopsis*, melon, and cucumber. We found out that many genes involved in stamen initiation, tapetal development, and pollen maturation were downregulated in female buds compared to male buds. Interestingly, ethylene response-related genes were upregulated in female buds compared to male buds indicating a probable role of ethylene in stamen suppression similar to monoecious cucurbits such as melon and cucumber. We speculate that the Y-chromosome might express genes that inhibit ethylene signaling or suppress the carpel development, the site

of ethylene production leading to the formation of stamens in male flowers. This was supported by the observation that AgNO₃ treatment suppressed ethylene responses and induced stamen development in female flowers of *C. grandis*. However, the pollens produced by Ag-H flowers were sterile indicating a decisive role of Y-chromosome in determining maleness. In accordance with this, the transcripts involved in pollen maturation, pollen germination, and pollen tube elongation were depleted in middle-staged GyM-H buds compared to male buds. This could be because of the absence of Y-chromosome in GyM plant. Altogether, differentially expressed genes identified in this study could shed light on the probable mechanisms of sex determination, differentiation, and modification in *C. grandis*.

Additional files

Additional file 1: Table S1. List of primers used in this study. (PDF 328 kb)

Additional file 2: De novo-assembled *Coccinia grandis* floral bud transcriptome. Available at <https://doi.org/10.6084/m9.figshare.5220874.v1>. (FASTA 4100 kb)

Additional file 3: Figure S1. ExN50 statistic for *C. grandis* flower de novo transcriptome assembly. (TIFF 1799 kb)

Additional file 4: Table S2. Distribution of percent length coverage for the top matching Swiss-Prot database entries. (PDF 7 kb)

Additional File 5: Table S3. *Coccinia grandis* flower bud transcriptome metrics calculated using TransRate. Protein-coding primary transcripts of *Cucumis sativus* were chosen as reference. (PDF 10 kb)

Additional file 6: Table S4. Annotation report for the de novo-assembled *C. grandis* flower bud transcriptome generated using Trinotate. Available at <https://doi.org/10.6084/m9.figshare.5217652.v1>. (XLSX 5870 kb)

Additional file 7: Figure S2. Pairwise comparisons of transcript abundance. MA plots showing average log fold change (logFC) vs average log of counts across replicates. Volcano plots showing differentially expressed transcripts in relation to FDR (False discovery rate). Features found DE at FDR <0.05 are colored red. Features with *P*-values at most 1e-3 and at least 2^Δ fold change are differentially expressed. (TIFF 5988 kb)

Additional file 8: Table S5. Expression matrix of transcripts with *P*-values at most 1e-3 and at least 2^Δ fold differential expression. (XLS 7379 kb)

Additional file 9: Table S6. Gene ontology (GO) enrichment analysis for differentially expressed transcripts in each pairwise comparison. (XLSX 1509 kb)

Additional file 10: Figure S3. Transcript clusters extracted from the hierarchical clustering with R. X-axis: samples; y-axis: median-centered log₂(FPKM). Grey lines, individual transcripts; blue line, average expression values per cluster. (TIFF 5037 kb)

Abbreviations

Ag-H flower buds: Hermaphrodite flowers from a plant treated with silver nitrate; AVG: Aminoethoxyvinylglycine; BUSCO: Benchmarking Universal Single-Copy Orthologs; CRB-BLAST: Conditional Reciprocal Best BLAST; edgeR: Empirical Analysis of Digital Gene Expression Data in R; eggNOG: evolutionary genealogy of genes: Non-supervised Orthologous Groups; FPKM: Fragments per kb per million reads; GO: Gene ontology; GyM: Gynomonocious; GyM-H flower buds: Hermaphrodite flowers from a gynomonocious plant; HMMER: Tool used for searching sequence databases for sequence homologs; KEGG: Kyoto encyclopedia of genes and genomes; NGS: Next-generation sequencing; Nr: Non-redundant protein database; PFAM: database of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs); qRT-

PCR: Quantitative real-time PCR; RIN: RNA integrity number; RSEM: RNA-Seq by Expectation-Maximization; Swiss-Prot: Annotated protein sequence database; TrEMBL: Computer-annotated supplement to Swiss-Prot database

Acknowledgements

We thank Mr. Nitish Lahigude for maintaining plants. Computational resource provided by Prof. Sanjeev Galande's lab at IISER Pune is thankfully acknowledged.

Funding

R.S.D. acknowledges research fellowship obtained from CSIR, New Delhi. Financial Support from DBT, Govt. of India (Grant No-BT/PR16399/NER/95/125/2015), and Director, IISER Pune is thankfully acknowledged.

Availability of data and materials

The raw sequencing data were deposited in the NCBI Short Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra/>) under the accession number SRP111347. De novo-assembled *Coccinia grandis* floral bud transcriptome is available at <https://doi.org/10.6084/m9.figshare.5220874.v1>. Annotation Report for the de novo-assembled *C. grandis* flower bud transcriptome generated using Trinotate is available at <https://doi.org/10.6084/m9.figshare.5217652.v1>.

Author Contributions

RSD, AKB, and SS planned and designed the research, RSD performed experiments and analysed the data. RSD, AKB, and SS wrote the manuscript. AB, JB, and RKS helped in critical discussion, interpretation of data and edited the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Publisher's Note

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

Author details

¹Biology Division, Indian Institute of Science Education and Research (IISER), Pune, Pune, Maharashtra, India. ²Department of Botany, Tripura University, Suryamaninagar, Tripura, India. ³IPSG, INRA, CNRS, University Paris Sud, University of Evry, University Paris Diderot, University of Paris Saclay, Batiment 630, 91405 Orsay, France.

Received: 19 July 2017 Accepted: 30 November 2017

Published online: 12 December 2017

References

- Charlesworth D. Plant sex determination and sex chromosomes. *Heredity*. 2002;88(2):94–101.
- Ainsworth C, Crossley S, Buchanan-Wollaston V, Thangavelu M, Parker J. Male and female flowers of the dioecious plant sorrel show different patterns of MADS box gene expression. *Plant Cell*. 1995;7(10):1583–98.
- Pfent C, Pobursky KJ, Sather DN, Golenberg EM. Characterization of *SpAPETALA3* and *SpPSTILLATA*, B class floral identity genes in *Spinacia oleracea*, and their relationship to sexual dimorphism. *Dev Genes Evol*. 2005; 215(3):132–42.
- Yin TM, DiFazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M, et al. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res*. 2008; 18(3):422–30.
- Urasaki N, Tarora K, Shudo A, Ueno H, Tamaki M, Miyagi N, Adaniya S, Matsumura H. Digital transcriptome analysis of putative sex-determination genes in papaya (*Carica papaya*). *PLoS One*. 2012;7(7):e40904.
- Kouonon LC, Jacquemart A-L, Zoro Bi AI, Bertin P, Baudoin J-P, Dje Y. Reproductive biology of the andromonoecious *Cucumis melo* subsp. *agrestis* (Cucurbitaceae). *Ann Bot*. 2009;104(6):1129–39.
- Kumar LSS, Deodikar GB. Sex chromosomes of *Coccinia indica* Wight and Arn. *Curr Sci*. 1940;9(3):128–30.
- Bhaduri PN, Bose PC. Cyto-genetical investigations in some common cucurbits, with special reference to fragmentation of chromosomes as a physical basis of speciation. *J Genet*. 1947;48(2):237–56.
- Chakravorti AK. Cytology of *Coccinia indica* W. & a. With reference to the behaviour of its sex-chromosomes. *Proceedings: Plant Sciences*. 1948;27(3):74–86.
- Kumar LS, Visevshwaraiha S. Sex mechanism in *Coccinia indica* Wight and Arn. *Nature*. 1952;170(4321):330–1.
- Bhowmick BK, Jha TB, Jha S. Chromosome analysis in the dioecious cucurbit *Coccinia grandis* (L.) Voigt. *Chromosome Science*. 2012;15(1+ 2):9–15.
- Sousa A, Fuchs J, Renner SS. Molecular cytogenetics (FISH, GISH) of *Coccinia grandis*: a ca. 3 myr-old species of Cucurbitaceae with the largest Y/ autosome divergence in flowering plants. *Cytogenetic and genome research*. 2013;139(2):107–18.
- Bhaskute S, Makde K, Deshpande P. Staminal organization in *Coccinia grandis* (L.) Voigt. *Ann Bot*. 1986:415–8.
- Ghadge AG, Karmakar K, Devani RS, Banerjee J, Mohanasundaram B, Sinha RK, Sinha S, Banerjee AK. Flower development, pollen fertility and sex expression analyses of three sexual phenotypes of *Coccinia grandis*. *BMC Plant Biol*. 2014;14(1):325.
- Hardenack S, Ye D, Saedler H, Grant S. Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. *Plant Cell*. 1994;6(12):1775–87.
- Di Stilio VS, Kramer EM, Baum DA. Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae) – a new model for the study of dioecy. *Plant J*. 2005;41(5):755–66.
- Beyer EM. A potent inhibitor of ethylene action in plants. *Plant Physiol*. 1976;58(3):268–71.
- McDaniel BK, Binder BM. Ethylene receptor 1 (ETR1) is sufficient and has the predominant role in mediating inhibition of ethylene responses by silver in *Arabidopsis thaliana*. *J Biol Chem*. 2012;287(31):26094–103.
- Sarath G, Mohan Ram HY. Comparative effect of silver ion and gibberellic acid on the induction of male flowers on female *Cannabis* plants. *Experientia*. 1979; 35(3):333–4.
- Yin T, Quinn JA. Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am J Bot*. 1995:1537–46.
- Law TF, Lebel-Hardenack S, Grant SR. Silver enhances stamen development in female white campion (*Silene latifolia* [Caryophyllaceae]). *Am J Bot*. 2002; 89(6):1014–20.
- Blackburn KB. Sex chromosomes in plants. *Nature*. 1923;112(2819):687–8.
- Kihara H, Ono T. The sex-chromosomes of *Rumex acetosa*. *Zeitschrift für induktive Abstammungs-und Vererbungslehre*. 1925;39(1):1–7.
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA. Rapid *de novo* evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol*. 2012;10(4):e1001308.
- Wu T, Qin Z, Zhou X, Feng Z, Du Y. Transcriptome profile analysis of floral sex determination in cucumber. *J Plant Physiol*. 2010;167(11):905–13.
- Guo S, Zheng Y, Joung J-G, Liu S, Zhang Z, Crasta OR, Sobral BW, Xu Y, Huang S, Fei Z. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics*. 2010;11(1):384.
- Akagi T, Henry IM, Tao R, Comai L. A Y-Chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science*. 2014;346(6209):646–50.
- Harkess A, Mercati F, Shan HY, Sunseri F, Falavigna A, Leebens-Mack J. Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytol*. 2015;207(3):883–92.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;btu170.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 2011;29(7):644.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494–512.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.

33. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;btv351.
34. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* 2016;26(8):1134–44.
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
36. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 2008;36(10):3420–35.
37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics.* 2011;12(1):323.
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
39. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001;25(4):402–8.
40. Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA. SEX-DETECTOR: a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution.* 2016;8(8):2530–43.
41. Krizek BA, Meyerowitz EM. The *Arabidopsis* homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development.* 1996;122(1):11–22.
42. Jia 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.* 2008;105(6):2220–5.
43. Gu JN, Zhu J, Yu Y, Teng XD, Lou Y, Xu XF, Liu JL, Yang ZN. DYT1 directly regulates the expression of *TDF1* for tapetum development and pollen wall formation in *Arabidopsis*. *Plant J.* 2014;80(6):1005–13.
44. Zhu E, You C, Wang S, Cui J, Niu B, Wang Y, Qi J, Ma H, Chang F. The DYT1-interacting proteins bHLH010, bHLH089 and bHLH091 are redundantly required for *Arabidopsis* anther development and transcriptome. *Plant J.* 2015;83(6):976–90.
45. Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari M-A, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science.* 2008;321(5890):836–8.
46. Boualem A, Troadec C, Kovalski I, Sari M-A, Perl-Treves R, Bendahmane A. A conserved ethylene biosynthesis enzyme leads to andromonoecy in two *Cucumis* species. *PLoS One.* 2009;4(7):e6144.
47. Boualem A, Troadec C, Camps C, Lemhemdi A, Morin H, Sari M-A, Fraenkel-Zagouri R, Kovalski I, Dogimont C, Perl-Treves R, Bendahmane A. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science.* 2015;350(6261):688–91.
48. Boualem A, Lemhemdi A, Sari M-A, Pignoly S, Troadec C, Abou Choucha F, Solmaz I, Sari N, Dogimont C, Bendahmane A. The andromonoecious sex determination gene predates the separation of *Cucumis* and *Citrullus* genera. *PLoS One.* 2016;11(5):e0155444.
49. Owens K, Peterson C, Tolla G. Production of hermaphrodite flowers on gynoeious muskmelon by silver nitrate and aminoethoxyvinylglycine. *Hortscience.* 1980;15(5):654–5.
50. Zhang J, Boualem A, Bendahmane A, Ming R. Genomics of sex determination. *Curr Opin Plant Biol.* 2014;18:110–6.
51. 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.
52. Sorensen AM, Kröber S, Unte US, Huijser P, Dekker K, Saedler H. The *Arabidopsis* *ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. *Plant J.* 2003;33(2):413–23.
53. Yang X, Makaroff CA, Ma H. The *Arabidopsis* *MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis. *Plant Cell.* 2003;15(6):1281–95.
54. Ito T, Nagata N, Yoshida 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.
55. 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.
56. Borg M, Brownfield L, Khatib H, Sidorova A, Lingaya M, Twell D. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. *Plant Cell.* 2011;23(2):534–49.
57. Tsugama D, Matsuyama K, Ide M, Hayashi M, Fujino K, Masuda K. A putative MYB35 ortholog is a candidate for the sex-determining genes in *Asparagus officinalis*. *Sci Rep.* 2017;7.
58. Wang L, Wang W, Wang Y-Q, Liu Y-Y, Wang J-X, Zhang X-Q, Ye D, Chen L-Q. *Arabidopsis* galacturonosyltransferase (GAUT) 13 and GAUT14 have redundant functions in pollen tube growth. *Mol Plant.* 2013;6(4):1131–48.
59. Jiang L, Yang S-L, Xie L-F, Puah CS, Zhang X-Q, Yang W-C, Sundaresan V, Ye D. *VANGUARD1* encodes a pectin methylesterase that enhances pollen tube growth in the *Arabidopsis* style and transmitting tract. *Plant Cell.* 2005;17(2):584–96.
60. Tian G-W, Chen M-H, Zaltsman A, Citovsky V. Pollen-specific pectin methylesterase involved in pollen tube growth. *Dev Biol.* 2006;294(1):83–91.
61. Leroux C, Bouton S, Kiefer-Meyer M-C, Fabrice TN, Marek A, Guénin S, Fournet F, Ringli C, Pelloux J, Driouch A, Lerouge P, Lehner A, Mollet J-C. PECTIN METHYLESTERASE48 is involved in *Arabidopsis* pollen grain germination. *Plant Physiol.* 2015;167(2):367–80.
62. Zhan H, Zhong Y, Yang Z, Xia H. Enzyme activities of *Arabidopsis* inositol polyphosphate kinases AtIPK2a and AtIPK2b are involved in pollen development, pollen tube guidance and embryogenesis. *Plant J.* 2015;82(5):758–71.
63. Dong X, Hong Z, Sivaramkrishnan 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.
64. Lago C, Clerici E, Dreni L, Horlow C, Caporali E, Colombo L, Kater MM. The *Arabidopsis* TFIID factor AtTAF6 controls pollen tube growth. *Dev Biol.* 2005;285(1):91–100.
65. Bernal AJ, Yoo C-M, Mutwil M, Jensen JK, Hou G, Blaukopf C, Sørensen I, Blancaflor EB, Scheller HV, Willats WG. Functional analysis of the cellulose synthase-like genes *CSLD1*, *CSLD2*, and *CSLD4* in tip-growing *Arabidopsis* cells. *Plant Physiol.* 2008;148(3):1238–53.
66. Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLoS Genet.* 2009;5(8):e1000621.
67. Xie B, Wang X, Hong Z. Precocious pollen germination in *Arabidopsis* plants with altered callose deposition during microsporogenesis. *Planta.* 2010;231(4):809–23.
68. Zhao L-N, Shen L-K, Zhang W-Z, Zhang W, Wang Y, Wu W-H. Ca^{2+} -dependent protein kinase11 and 24 modulate the activity of the inward rectifying K^{+} channels in *Arabidopsis* pollen tubes. *Plant Cell.* 2013;25(2):649–61.
69. Morant M, Jørgensen K, Schaller H, Pinot F, Møller BL, Werck-Reichhart D, Bak S. CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell.* 2007;19(5):1473–87.
70. Dobritsa AA, Geanconteri A, Shrestha J, Carlson A, Kooyers N, Coerper D, Urbanczyk-Wochniak E, Bench BJ, Sumner LW, Swanson R, Preuss D. A large-scale genetic screen in *Arabidopsis* to identify genes involved in pollen exine production. *Plant Physiol.* 2011;157(2):947–70.
71. Xiong SX, Lu JY, Lou Y, Teng XD, Gu JN, Zhang C, Shi QS, Yang ZN, Zhu J. The transcription factors MS188 and AMS form a complex to activate the expression of *CYP703A2* for sporopollenin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 2016;88(6):936–46.
72. Zhao D-Z, Wang G-F, Speal B, Ma H. The *EXCESS MICROSPOROXYTES1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. *Genes Dev.* 2002;16(15):2021–31.
73. Albrecht C, Russinova E, Hecht V, Baaijens E, de Vries S. The *Arabidopsis thaliana* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASES1 and 2 control male sporogenesis. *Plant Cell.* 2005;17(12):3337–49.
74. Mizuno S, Osakabe Y, Maruyama K, Ito T, Osakabe K, Sato T, Shinozaki K, Yamaguchi-Shinozaki K. Receptor-like protein kinase 2 (RPK 2) is a novel factor controlling anther development in *Arabidopsis thaliana*. *Plant J.* 2007;50(5):751–66.
75. De Storme N, Geelen D. The *Arabidopsis* mutant *jason* produces unreduced first division restitution male gametes through a parallel/fused spindle mechanism in meiosis II. *Plant Physiol.* 2011;155(3):1403–15.