Functional & Integrative Genomics (2021) 21:251–263
https://doi.org/10.1007/s10142-021-00775-y

ORIGINAL ARTICLE

Genome-wide comparative transcriptome analysis of the A4-CMS line ICPA 2043 and its maintainer ICPB 2043 during the floral bud development of pigeonpea

Abhishek Bohra[1] · Abhishek Rathore[2] · Prasad Gandham[2] · Rachit K. Saxena[2] · S. J. Satheesh Naik[1] · Dibendu Dutta[1] · Indra P. Singh[1] · Farindra Singh[1] · Meenal Rathore[1] · Rajeev K. Varshney[2] · Narendra P. Singh[1]

Received: 29 August 2020 / Revised: 5 February 2021 / Accepted: 9 February 2021 / Published online: 26 February 2021

Abstract

Cytoplasmic male sterility (CMS) offers a unique system to understand cytoplasmic nuclear crosstalk, and is also employed for exploitation of hybrid vigor in various crops. Pigeonpea A4-CMS, a predominant source of male sterility, is being used for efficient hybrid seed production. The molecular mechanisms of CMS trait remain poorly studied in pigeonpea. We performed genome-wide transcriptome profiling of A4-CMS line ICPA 2043 and its isogenic maintainer ICPB 2043 at two different stages of floral bud development (stage S1 and stage S2). Consistent with the evidences from some other crops, we also observed significant difference in the expression levels of genes in the later stage, i.e., stage S2. Differential expression was observed for 143 and 55 genes within the two stages of ICPA 2043 and ICPB 2043, respectively. We obtained only 10 differentially expressed genes (DEGs) between the stage S1 of the two genotypes, whereas expression change was significant for 582 genes in the case of stage S2. The qRT-PCR assay of randomly selected six genes supported the differential expression of genes between ICPA 2043 and ICPB 2043. Further, GO and KEGG pathway mapping suggested a possible compromise in key bioprocesses during flower and pollen development. Besides providing novel insights into the functional genomics of CMS trait, our results were in strong agreement with the gene expression atlas of pigeonpea that implicated various candidate genes like sucrose-proton symporter 2 and an uncharacterized protein along with pectate lyase, pectinesterase inhibitors, L-ascorbate oxidase homolog, ATPase, β-galactosidase, polygalacturonase, and aldose 1-epimerase for pollen development of pigeonpea. The dataset presented here provides a rich genomic resource to improve understanding of CMS trait and its deployment in heterosis breeding in pigeonpea.

Keywords Cytoplasm · Male sterility · Hybrid · Gene expression · Pathway · RNA sequencing

Introduction

Cytoplasmic male sterility (CMS), an inability to produce functional pollen, is reported in more than 610 plant species (Chen and Liu 2014). To impart yield gains and stress resistance, this unique feature of CMS plants has been exploited for developing hybrids in various crops (Bohra et al. 2016). In pigeonpea, nine cytoplasms have been reported so far that confer male sterility, of which A2- and A4-CMSs have been applied for hybrid development due to stable male sterility and fertility restoration systems (Saxena et al. 2010; Bohra et al. 2020).

Defective cytoplasmic-nuclear crosstalk that interferes with the normal process of reproductive development and pollen formation results in CMS phenotypes (Horn et al. 2014). In this context, rearrangements in mitochondrial genomes have been identified as key drivers for CMS induction, and various research groups have associated aberrant mitochondrial open reading frames (ORFs) with CMS in different crops (Bohra et al. 2016). The CMS-associated genes are commonly identified by analyzing variations in mitochondrial genomes, transcriptomes, and proteomes of male sterile and fertile plants (Lv et al. 2020). In pigeonpea, analysis of whole mitochondrial genome sequences of A4-CMS line, corresponding

✉ Abhishek Bohra
abhi.omics@gmail.com

[1] ICAR-Indian Institute of Pulses Research (ICAR-IIPR), Kanpur 208024, India

[2] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India

B-line, and hybrid revealed a set of mitochondrial ORFs as putative candidates responsible for A4-CMS (Tuteja et al. 2013). Later, structure variation and expression analysis of the essential mitochondrial genes showed involvement of *nad* genes in A4-CMS induction in pigeonpea (Sinha et al. 2015).

Recent research has shown the participation of other players besides mitochondrial elements in development of non-functional male gametophyte in CMS plants (Huang et al. 2011). In view of this, RNA-sequencing of floral buds of CMS lines and their cognate maintainers has emerged as a powerful tool to gain new insights on global transcriptional networks underlying male sterility. For example, Illumina sequencing of flower buds of Ogura-type CMS turnip (male sterility is conferred by mitochondrial gene *orf138*) and its maintainer line revealed 5,117 differentially expressed genes (DEGs) (Lin et al. 2019), and the authors proposed that abnormalities in tapetum degeneration and pollen wall formation might present the crucial events leading to male sterility.

The availability of whole genome sequence greatly improves our ability to perform gene annotations and analysis of the genome-wide sets of transcriptome data (Wu et al. 2017). Reference-based analysis of whole transcriptome datasets of flower buds of CMS lines and isogenic maintainers or restorers has been performed recently in different crops such as maize (Liu et al. 2018), cotton (Hamid et al. 2019; Li et al. 2019; Liu et al. 2020), *Brassica* (Yan et al. 2013), soybean (Li et al. 2015), and pepper (Liu et al. 2013). In the current research, we performed deep sequencing of the floral bud transcriptomes of A4-CMS line (ICPA 2043) and its isogenic maintainer line (ICPB 2043) from two different stages of bud development. We then analyzed the candidate genes expressing differentially (DEGs) between the sterile and fertile lines from two stages (S1 and S2). Furthermore, we elucidated the biological functions of the DEGs and discuss about the impaired processes that eventually lead to abortive microsporogenesis in A4-CMS pigeonpea.

Material and methods

Plant material and isolation of RNA from floral buds

The seeds of the CMS line ICPA 2043 and maintainer line ICPB 2043 were provided by ICRISAT, India. The two lines were planted at main farm of ICAR-Indian Institute of Pulses Research (IIPR), Kanpur. Samples of flower buds from two different stages (<10 mm (S1) and >10 mm (S2)) were collected from CMS line and maintainer line and immediately frozen in liquid nitrogen. The collected samples were stored at −80 °C for RNA extraction and transcriptome sequencing. Two biological replicates were considered for each stage of the two genotypes. Total RNA was isolated from the samples by Trizol method. The quality of the isolated RNA was checked on 1% formaldehyde denaturing agarose gel and quantified using Qubit 2.0 (Thermo Fisher Scientific, USA).

Library preparation and sequencing

The libraries were prepared with input total RNA ~1 μg using Illumina TruSeq Stranded mRNA Library Preparation Kit as per the manufacturer's protocol. The amplified libraries were analyzed on Bioanalyzer 2100 (Agilent Technologies) using High Sensitivity (HS) DNA chip as per the manufacturer's instructions. After quantifying libraries with Qubit and the mean peak size from Bioanalyzer profile, libraries were loaded onto Illumina platform for cluster generation and paired-end sequencing.

Data analysis of RNA-Seq and differential gene expression analysis

Sequencing reads in the fastq format were quality checked using FastQC v. 0.11.8 tool (Andrews 2010). Then Illumina TruSeq adapters were trimmed and low-quality reads were discarded using Trimmomatic v. 0.36 by specifying parameters SLIDINGWINDOW:4:15 and minimum read length cutoff of 35 bp (Bolger et al. 2014). The trimmed reads were again quality checked and high-quality reads were aligned to the reference genome using Hisat2 v.2.1.0 (Kim et al. 2015) followed by read count estimation using featureCounts v.2.0.0 (Liao et al. 2014). The DEGs were identified using DESeq2 R package (Love et al. 2014). The significant DEGs were filtered for further analysis using adjusted p value < 0.1 and a cutoff of 2 in logFC values. EnhancedVolcano R package was used to generate volcano plot (Blighe et al. 2019)

Functional annotation of DEGs and KEGG pathway level analysis

Gene ontology (GO) annotations for the identified DEGs were taken from the annotations given in the original Pigeonpea genome assembly files and LegumeIP v. 2 database (Li et al. 2012). Further, DEGs were subjected to GO enrichment analysis to find enriched molecular functions, biological process, and cellular component using topGo R package (Alexa and Rahnenfuhrer 2019) and BiNGO v. 3.0.3 (Maere et al. 2005) plugin in Cytoscape v. 3.7.1 (Shannon et al. 2003). The KEGG Orthology-Based Annotation System (KOBAS v. 3.0) was used to calculate the enriched KEGG pathways from the DEGs (Xie et al. 2011).

Quantitative Real-Time PCR of DEGs

To validate the gene expression patterns from RNA-Seq data, qRT-PCR analysis was performed with ABI SYBR GREEN PCR reaction on an ABI Fast 7500 System (Applied Biosystems, Foster City, CA). Total RNA was extracted from flower bud tissue of ICPA 2043 and ICPB 2043, and 3 μg of RNA was used for first-strand cDNA synthesis using the High Capacity cDNA Reverse Transcription Kit (Thermo Fisher, USA) following the manufacturer's protocol. PCR conditions for qRT-PCR reactions were used as follows: cycle initiated with pre-incubation at 50 °C for 2 min, denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing and extension at 60 °C for 15 s and 72 °C for 1 min. The actin 1 was used as internal control. Each reaction was performed in two biological and two technical replicates along with no template control. Relative quantification (RQ) was determined by the following formula: $RQ=2^{-\Delta\Delta Ct}$ (Livak and Schmittgen 2001).

Results

Transcriptome sequencing

The present study involves transcriptome sequencing of two stages of unopened flower buds of CMS line ICPA 2043 and its isogenic maintainer line ICPB 2043. A total of 282.8 million paired reads amounting to 43.7 Gb were generated. After the raw data were trimmed, 129.72 million paired and 4.96 million unpaired clean reads for ICPA 2043 sample and 142.92 million paired and 4.97 million unpaired clean reads for ICPB 2043 sample were obtained. All clean reads were mapped to the pigeonpea reference genome v. 1.0 (Varshney et al. 2012) by Hisat software. The alignment rate of different samples and replicates ranged from 87.34 to 96.57% (Table 1). One sample showed poor matching with the reference genome (40.89%) and hence this sample was excluded from the downstream analysis. Supplementary Figure 1 depicts the sample distance matrix to visualize relationships among different replicates. The raw reads from this study have been deposited in the NCBI SRA database and are accessible through BioSample accessions: SAMN14981245, SAMN14981246, SAMN14981247, SAMN14981248, SAMN14981249, SAMN14981250, SAMN14981251, SAMN14981252 under the Bioproject PRJNA634080.

DEG analysis

We found a total of 22405 genes matching with the reference genome of pigeonpea with a total read count greater than 10 (a total of 48680 genes were there in the gff file provided with the original assembly; this file was used to get the read count estimates out of which 22405 genes are having counts >10 total). Subsequently, differentially expressed genes were computed between the two stages of ICPA 2043 and ICPB 2043 and between the genotypes ICPA 2043 and ICPB 2043 irrespective of the flower bud stage (Supplementary Table 1). Considering the threshold padj < 0.1 and fold change cutoff of 2 in logFC values, we could obtain 143 and 55 genes with differential expression between the two stages of ICPA 2043 and ICPB 2043, respectively (Figs. 1a, 2a and 1b, 2b). While comparing the two stages between ICPA 2043 and ICPB 2043, we obtained only 10 DEGs between the stage S1 of the two genotypes (Figs. 1c and 2c), whereas in the

Table 1 Statistical summary of sequencing data analysis

Sample name	Total alignments	Overall alignment rate	Successfully assigned alignments	Assigned	Unassigned Unmapped	Unassigned MultiMapping	Unassigned NoFeatures	Unassigned Ambiguity
ICPA 2043_ Stage S1_R1	33103480	90.83%	23103211	69.80%	2307419	4212866	2904093	575891
ICPA 2043_ Stage S1_R2	40183587	92.94%	28216692	70.20%	1993762	6135463	3107329	730341
ICPA 2043_ Stage S2_R1	34250277	93.01%	18694022	54.60%	1527930	10644101	2763665	620559
ICPA 2043_ Stage S2_R2	42090600	40.89%	12315902	29.30%	22761314	2521957	4284201	207226
ICPB 2043_ Stage S1_R1	52379409	90.57%	22982697	43.90%	2976759	21757652	3396418	1265883
ICPB 2043_ Stage S1_R2	45567342	96.57%	34487712	75.70%	918094	5871803	3404353	885380
ICPB 2043_ Stage S2_R1	54830774	91.56%	18988628	34.60%	2545861	29058835	2930290	1307160
ICPB 2043_ Stage S2_R2	36042631	87.34%	19957829	55.40%	3256174	9501807	2512444	814377

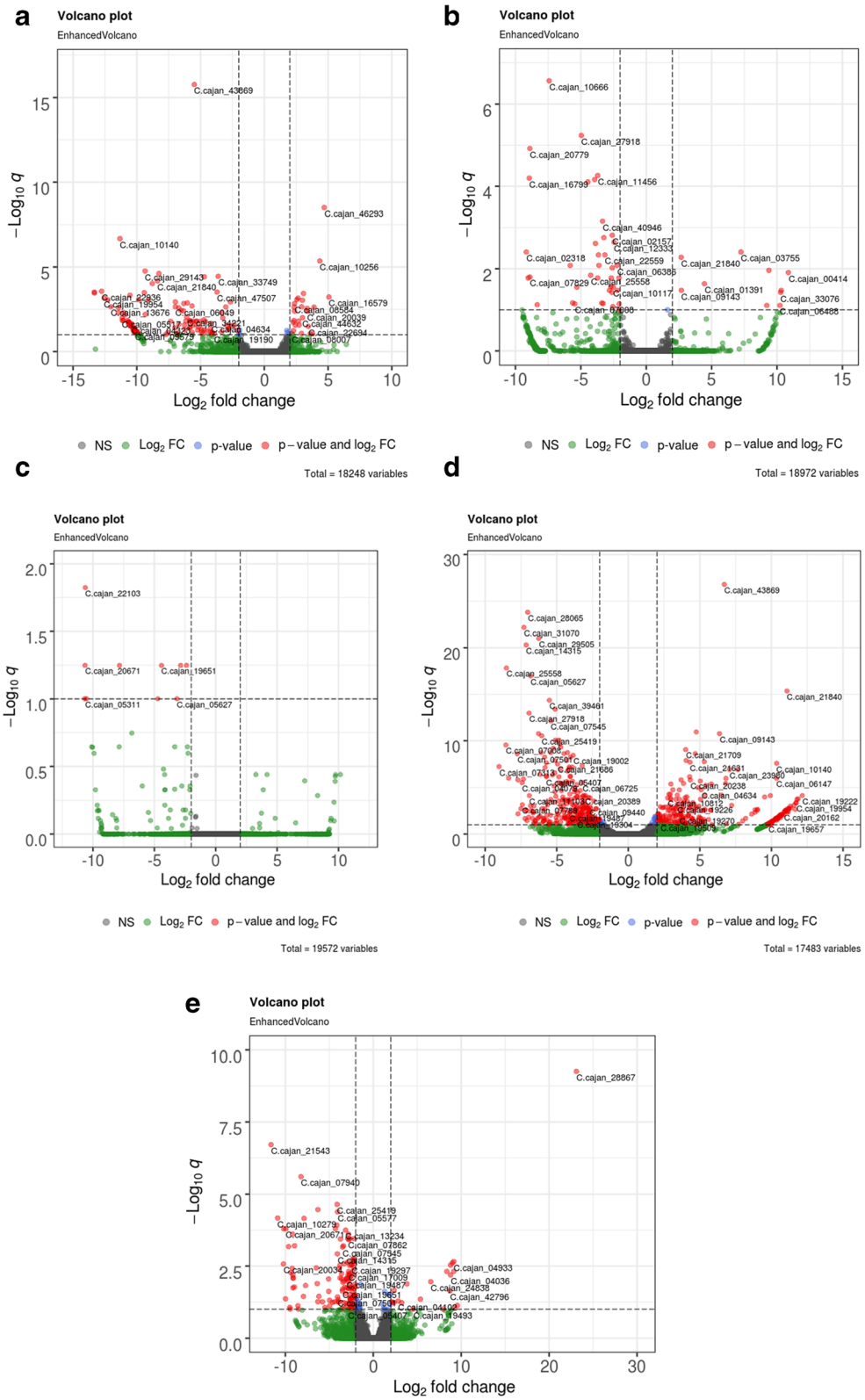

Fig. 1 Volcano plots showing expression abundance of genes during two stages of floral bud development in ICPA 2043 and ICPB 2043. The plots show DEGs between **a** ICPA 2043 stage S1 vs ICPA 2043 stage S2, **b** ICPB 2043 stage S1 vs ICPB 2043 stage S2, **c** ICPA 2043 stage S1 vs ICPB 2043 stage S1, **d** ICPA 2043 stage S2 vs ICPB 2043 stage S2, and **e** ICPA 2043 vs ICPB 2043. Upregulated genes are shown towards the right, while downregulated genes are on the left side. The most statistically significant DEGs are towards the top represented by red dots, while green dots represent less significant DEGs

case of stage S2, changes in expression level were recorded for 582 genes (Figs. 1d and 2d). One hundred fifty six genes showed differential expression between ICPA 2043 and ICPB 2043 irrespective of flower bud stage (Figs. 1e and 2e). Figure 3 shows number of DEGs between the two stages and the two genotypes.

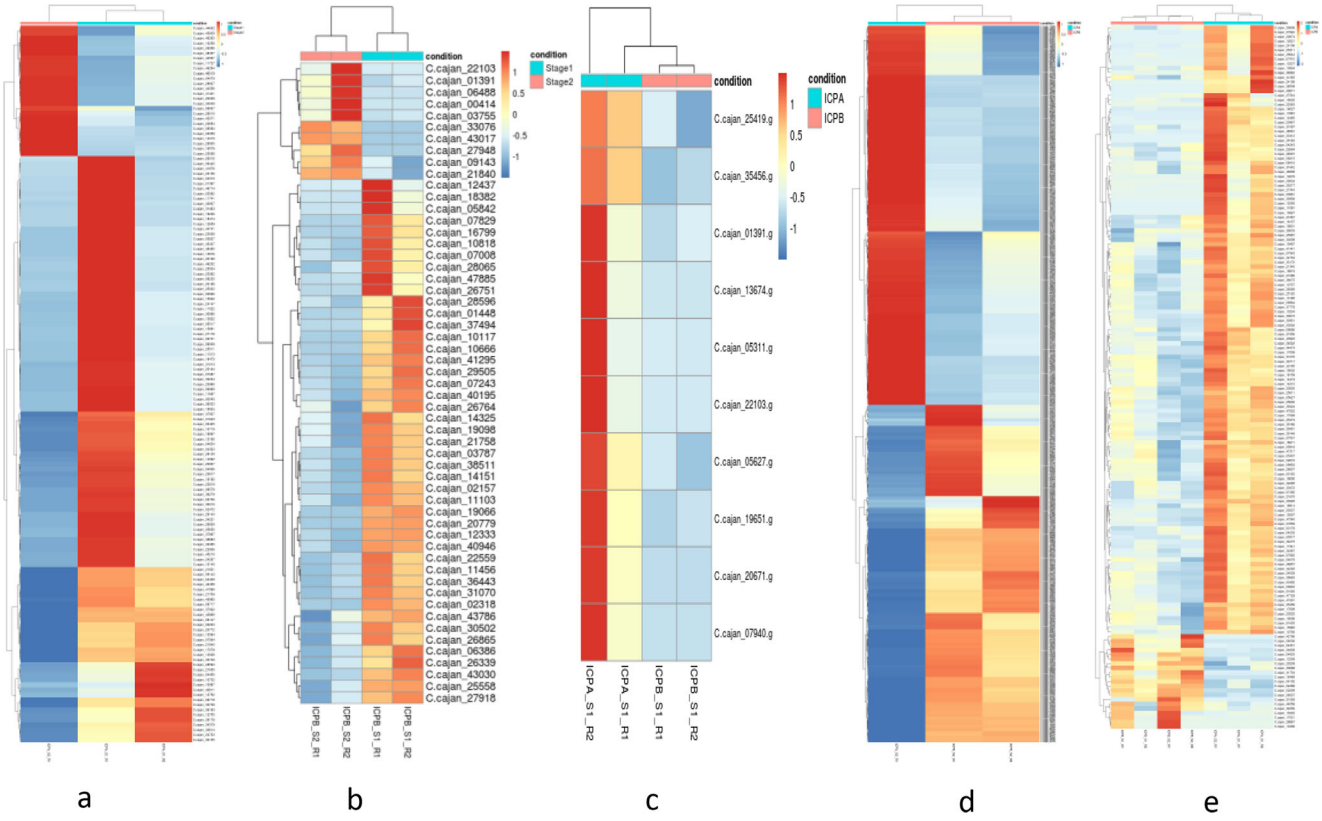

Fig. 2 Hierarchical cluster analysis of the genes showing differential expression between two stages of ICPA 2043 and ICPB 2043. **a** ICPA 2043 stage S1 vs ICPA 2043 stage S2, **b** ICPB 2043 stage S1 vs ICPB 2043 stage S2, **c** ICPA 2043 stage S1 vs ICPB 2043 stage S1, **d** ICPA 2043 stage S2 vs ICPB 2043 stage S2, **e** ICPA 2043 vs ICPB 2043

Functional annotations of the DEGs

Annotations of the DEGs were performed using topGO, BiNGO, and KEGG analyses (Supplementary Figures 2, 3). The gene ontology (GO) terms were enriched with p value < 0.05. Concerning the DEGs between the two stages (S1 and S2) of ICPA 2043, the downregulated genes in ICPA 2043 stage S2 were enriched with 96 GO terms belonging to 68 biological processes (BP), 23 molecular function (MF), and 5 cellular component (CC), whereas 28 GO terms enriched in the case of upregulated genes corresponded to 13 BP, 5 CC, and 10 MF. Similarly, 61 GO terms enriched from

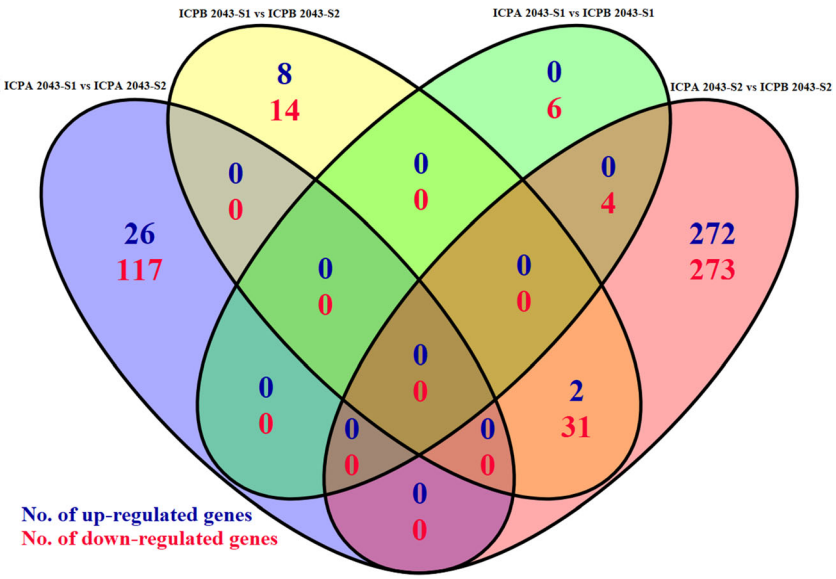

Fig. 3 Venn diagram showing differential expression of genes between the two stages and the two genotypes

downregulated genes in ICPB 2043 stage S2 belonged to 48 BP, 5 CC, and 8 MF when the DEGs were compared between the two stages (S1 and S2) of ICPB 2043. Seven GO terms were enriched from downregulated genes in ICPB 2043 stage S1 (6 BP and 1 CC) in comparison to the stage S1 of ICPA 2043. With regard to the gene expression patterns between the stage S2 of both lines (IPCA 2043 and ICPB 2043), the downregulated genes of ICPB 2043 had 188 GO terms associated with 140 BP, 11 CC, and 37 MF sub-ontologies, whereas 135 GO terms enriched from upregulated genes of ICPB 2043 belonged to 78 BP, 37 MF, and 20 CC categories (Supplementary Table 2).

Concerning dominance within each sub-ontology, "disaccharide transport (GO:0015766)" "sucrose transmembrane transporter activity (GO:0008515)," and "cell periphery (GO:0071944)" were most frequent in BP, MF, and CC, respectively, in the case of genes unregulated in stage S2 of ICPA 2043 in comparison to its stage S1. Similarly, "regulation of catalytic activity (GO:0050790)," "enzyme inhibitor activity (GO:0004857)," and extracellular region (GO:0005576) were dominant terms in the case of downregulated genes.

In the case of the genes upregulated in stage S2 of ICPB 2043 as compared to stage S2 of ICPA 2043, "negative regulation of catalytic activity (GO:0043086)," "enzyme inhibitor activity (GO:0004857)," and "cell wall (GO:0005618)" were the topmost terms in BP, MF, and CC, respectively, whereas "response to organo-nitrogen compound (GO:0010243)," "hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)," and "cell periphery (GO:0071944)" were most dominant in the case of downregulated genes (Fig. 4a and 4b).

Concerning the DEGs between the two genotypes irrespective of flower bud stage, majority of genes downregulated in ICPB 2043 were assigned to "response to chitin (BP; GO:0010200)," "chitin binding (MF; GO:0008061)" hydrolase activity, hydrolyzing O-glycosyl compounds (MF; GO:0004553), and "cell periphery (CC; GO:0071944)" sub-ontologies.

KEGG pathway enrichment analysis of the DEGs identified five categories in stage S2 including "Cellular Processes," "Environmental Information Processing," "Genetic Information Processing," and "Metabolism and Organismal Systems" (Supplementary Table 3). We did not observe any significant KEGG enrichment in ICPA 2043 stage S1 vs ICPB 2043 stage S1 comparison. At stage S2, within the metabolism category, the highest number of DEGs was related to "Metabolic pathways" followed by "Biosynthesis of secondary metabolites," "Pentose and glucuronate interconversions," "Amino sugar and nucleotide sugar metabolism," and "Phenylpropanoid biosynthesis." In the "Environmental Information Processing" category, DEGs were enriched in MAPK signaling pathway and Plant hormone signal transduction. Significantly enriched DEGs were associated with "Protein processing in endoplasmic reticulum" and "RNA transport" in the "Genetic Information Processing" category. Plant-pathogen interaction and circadian rhythm were enriched in "Organismal Systems" category. Representative images of KEGG enrichment analysis of the DEGs between stage S2 of ICPA 2043 and ICPB 2043 are shown in Fig. 5 a and b.

Confirmation of gene expression patterns by qRT-PCR

For qRT-PCR assay, we randomly selected six genes based on their differential expression between ICPA 2043 and ICPB 2043 regardless of the floral bud stage (Table 2). We observed differential expression between ICPA 2043 and ICPB 2043 for all six genes, *C.cajan_21543, C.cajan_39183*, *C.cajan_23412*, *C.cajan_20674*, *C.cajan_07940*, and *C.cajan_12465* (Fig. 6).

Discussion

Heterosis breeding has gained tremendous impetus in pigeonpea since discovery of male sterility systems based on genic (GMS) or cytoplasmic (CMS) elements. Understanding of the molecular mechanisms underlying CMS is crucial to increasing efficiency of hybrid breeding programs. In the present study, we examined the mechanisms of male sterility in A4-CMS pigeonpea by comparing the deep-sequenced transcriptomes of flower buds from CMS line and its isogenic fertile counterpart. Using pigeonpea genome as reference, we mapped up to 96.5% of clean reads generated from seven flower bud samples. Further, we observed dramatically enhanced number of DEGs in S2 stage as compared to the S1 stage. Our observation is consistent with the recent transcriptome-based studies involving CMS and fertile lines (Ding et al. 2018; Lv et al. 2020). These studies supported a gradual increase in DEGs with the preposition that the changes between the sterile and fertile lines tend to be more conspicuous during upward stages of flower development in plants.

We compared our results with the genes involved in flower-related network, as inferred from pigeonpea gene expression atlas CcGEA (Pazhamala et al. 2017). It is interesting to note that, in S2 stage of CMS line, we observed significantly lower expression (as compared to B line) for several previously reported genes such as encoding a sucrose-proton symporter 2 (*C.cajan_35396*) and an uncharacterized protein (*C.cajan_28171*). Importantly, the two genes present 'hub' genes, which are connected to 27 other genes in the co-expression network of flowering related genes constructed by Pazhamala et al. (2017). In addition, CMS line ICPA 2043 also showed downregulation of other genes of this co-

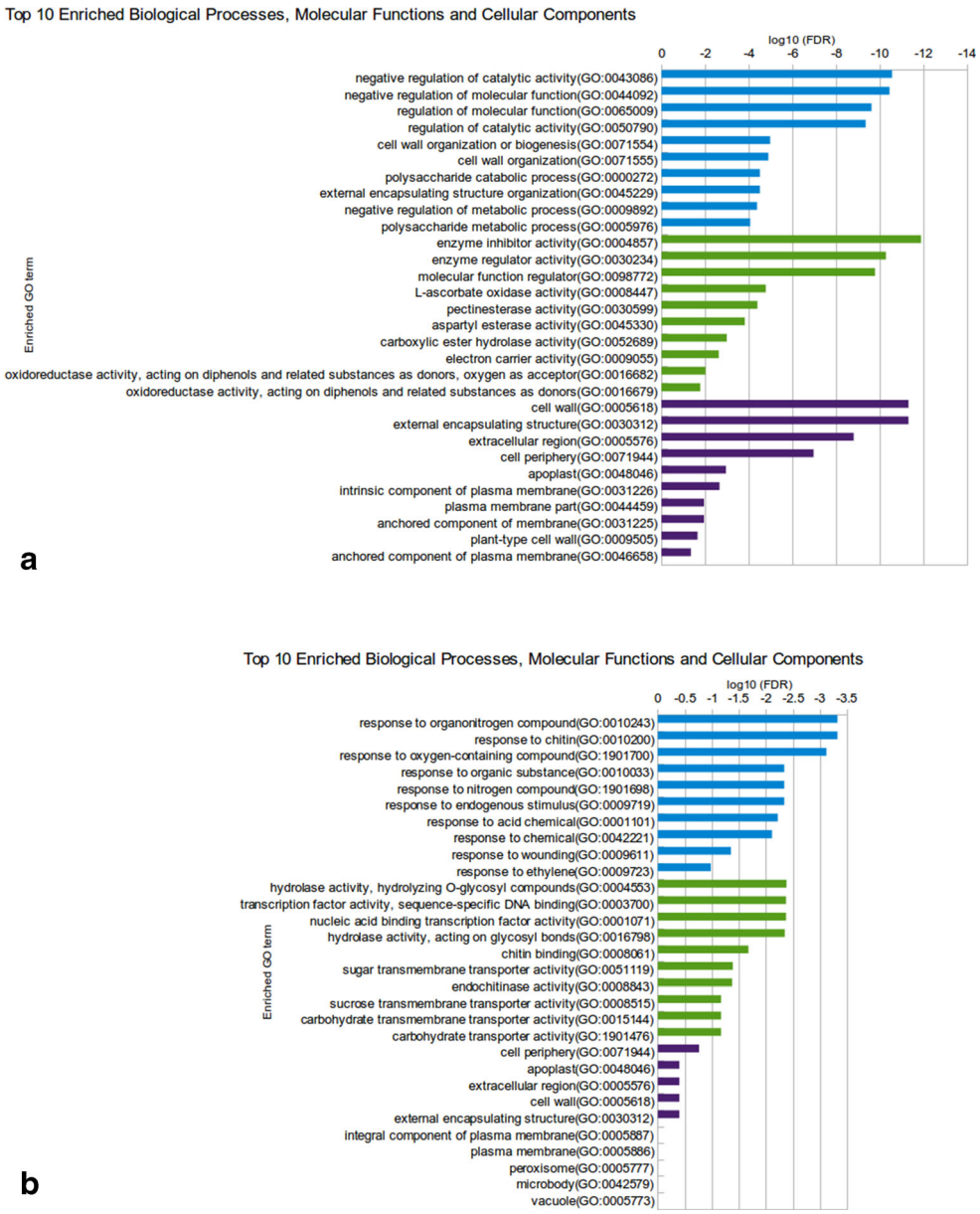

Fig. 4 Representative images showing top 10 enriched GO terms of DEGs **a** upregulated genes in stage S2 of ICPB 2043 in comparison to stage S2 of ICPA 2043, and **b** downregulated genes

expression network, including pollen specific SF3 (*C.cajan_11513*), olee1-like protein (*C.cajan_31667*), probable pectate lyase 3 (*C.cajan_44741*), pectinesterase inhibitors (*C.cajan_10140*, *C.cajan_04310*, *C.cajan_46391*), L-ascorbate oxidase homolog (*C.cajan_19226*), ATPase 6 (*C.cajan_45656*), β-galactosidase (*C.cajan_32927*), polygalacturonase (*C.cajan_04312*), and aldose 1-epimerase (*C.cajan_31220*). Of these, pollen-specific nature of genes like ascorbate oxidase homolog, β-galactosidase, and polygalacturonase has been evident from previous research (Masuko et al. 2006).

In S2 stage, downregulation was noted in CMS line for genes related to ATP synthesis including plasma membrane ATPases such as ATPase 9 (*C.cajan_37897*) and ATPase 6 (*C.cajan_45656*), cation-transporting P-type ATPase D (*C.cajan_20098*), V-type H+-transporting ATPase subunit E (*C.cajan_21260*), and V-type H+-transporting ATPase subunit D (*C.cajan_22965*). As reported by Wu et al. (2017), a compromised mitochondrial energy metabolism might not suffice a high ATP demand experienced during pollen formation, which could eventually lead to pollen abortion and CMS occurrence in plants. "Energy deficiency model" is among the widely accepted models that explain CMS formation in plants (Chen and Liu 2014).

Pathway mapping with KEGG suggested metabolism as the major activity in flower buds and our findings are in accordance with the recent transcriptome studies of CMS and maintainer line in other crops such as *Brassica napus* (An et al. 2014; Ding et al. 2018). Most of the DEGs in "Pentose and glucuronate interconversions" pathways such as *C.cajan_44741*, *C.cajan_26253*,

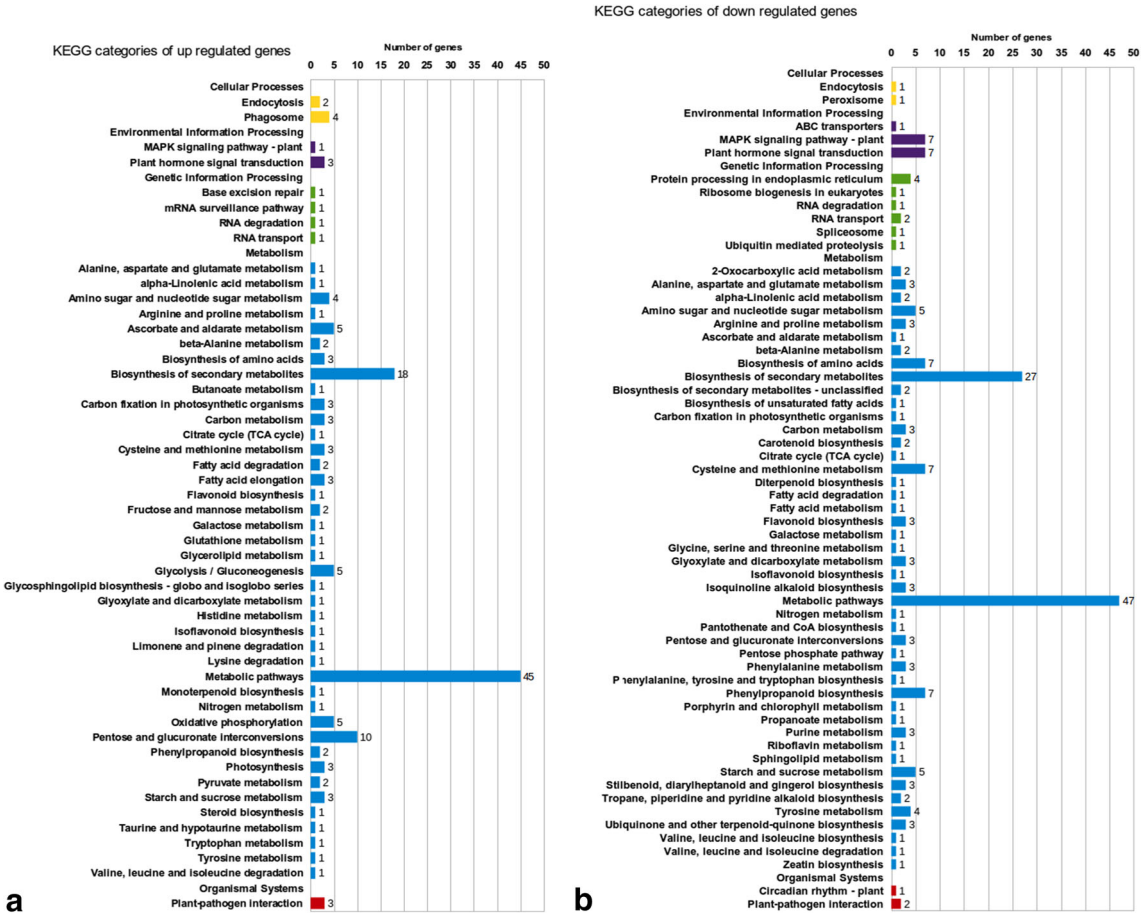

Fig. 5 The best represented KEGG pathways of DEGs in stage 2 of ICPB 2043 in comparison to stage 2 of ICPA 2043 **a** upregulated genes and **b** downregulated genes

C.cajan_46391, *C.cajan_01397*, *C.cajan_10140*, *C.cajan_06872*, *C.cajan_11741*, *C.cajan_48490*, *C.cajan_04312*, and *C.cajan_0431* were downregulated in CMS line. The role of "Pentose and glucuronate interconversion" in pollen development is well established (Ma et al. 2012; Pazhamala et al. 2017). Similarly, suppressed expression of genes engaged in "Pentose and glucuronate interconversion" and "Starch and sucrose metabolism" is shown to hamper normal pollen development in pepper (Liu et al. 2013; Lv et al. 2020). However, we observed higher expression levels of five genes enriched in "Starch and sucrose metabolism" in the S2 stage of the CMS line. Five genes involved in "Oxidative phosphorylation" viz. *C.cajan_22965*, *C.cajan_45656*, *C.cajan_10974*, *C.cajan_37897*, and *C.cajan_21260* showed suppressed expression in the CMS line. Earlier, several DEGs enriched in "Oxidative phosphorylation" were obtained based on

Table 2 Details of primer pairs selected for qRT-PCR assay

S. no.	Gene	Forward primer (5′-3′)	Tm (°C)	Reverse primer (5′-3′)	Tm (°C)	Product size (bp)
1	*C.cajan_21543*	AAGGGGTACGAGGATGTTCC	60.19	CAGCAGTGACACCATGATCC	60.121	172
2	*C.cajan_39183*	CGTGCAACTGAAGAGGTGAA	60.025	TAGTGCCAACTCCCACATCA	60.112	165
3	*C.cajan_23412*	GCATTAATCGATTTGCGGAC	60.432	TTCCTCCAGTCCATGGTAGC	60.073	157
4	*C.cajan_20674*	GCAAGGATAGCACTTTCGGA	60.352	GAAACAAAATGTTGCCCCAG	60.344	158
5	*C.cajan_07940*	CAGGAAGTTCCAATCCTTTCAA	60.455	ACTCCAACCCACAAATGGAA	60.21	161
6	*C.cajan_12465*	GACCATGGAAAGCCAATTCA	60.843	AGCATCAAAGCATCCCCTTT	60.959	154
7	Actin	TTCCCTGGCATTGCTGAT	60.161	TGCGATCCACATCTGCTG	60.546	159

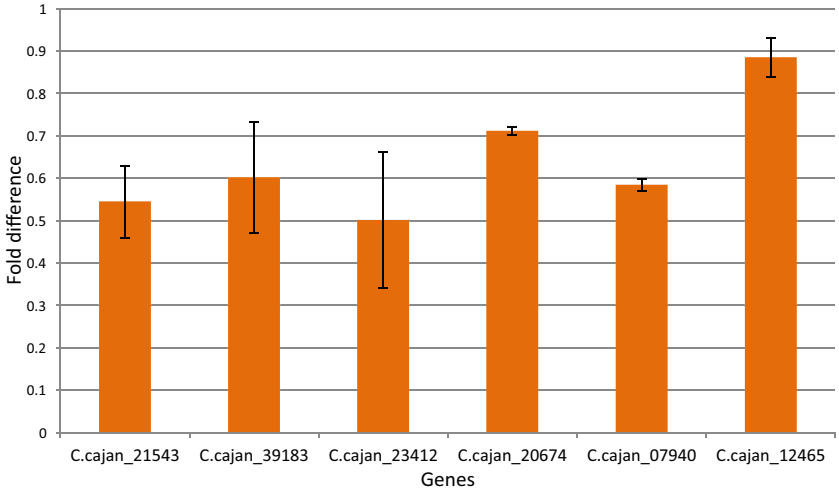

Fig. 6 Validation of the expression of six candidate genes by qRT-PCR

transcriptome sequencing of CMS and fertile lines in *Brassica napus* (Yan et al. 2013), welsh onion (Liu et al. 2016), pepper (Liu et al. 2013), and soybean (Li et al. 2015). We, however, observed higher expression of the gene *C.cajan_21670* encoding NADH dehydrogenase in the CMS line.

As shown by Li et al. (2015) in soybean, the process of pollen development is regulated through multiple signaling pathways, and Ca^{++} and calmodulin remain crucial to these pathways. In the current study, lower expression for four calmodulin-related genes in CMS line could explain the aberrations observed during the process of pollen formation.

Premature or delayed programmed cell death (PCD) causing aberrant tapetal degeneration results in abortive pollen in plants (Chen and Liu 2014). Aberrant ROS production serves as a signaling event to impair the perfectly timed nature of PCD during pollen development and maturation (Yan et al. 2013). In the current analysis, abundance of *C.cajan_38877* encoding cysteine protease might be responsible for abortive pollen in CMS pigeonpea. Cysteine protease expression has been related to regulation of tapetal PCD and pollen development in tobacco (Shukla et al. 2014) and Arabidopsis (Yang et al. 2014; Zhang et al. 2014). Similarly, decreased expression of several genes encoding arabinogalactan (AGP) proteins (*C.cajan_04588* (Fasciclin-like arabinogalactan protein 3), *C.cajan_18332* (Fasciclin-like arabinogalactan protein 10), *C.cajan_16162* (AtAGP14), *C.cajan_04588* (AtAGP16)) in CMS line could be responsible for pollen abortion. AGP proteins are crucial to pollen intine formation, and Lin et al. (2014) demonstrated the defective formation of pollen wall and pollen tube in *BcMF8*-silenced lines of *Brassica campestris*. *BcMF8* and *BcMF18* belong to AGP protein gene family, and the diverse roles that the AGP gene family plays in plant reproduction are well elucidated (Su and Higashiyama 2018). A more recent study in *Brassica rapa* also found lower expression of AGP and their homologs in CMS line as compared to its fertile counterpart (Lin et al. 2019).

Transcription factors (TFs) play important role in regulation of gene expression and a variety of TFs have been implicated in microspore development and function in plants (Yan et al. 2013). As shown in Table 3, present study reports differential expression for many TFs including bHLH (bHLH31, bHLH68, bHLH79, bHLH113, bHLH137, bHLH146), ethylene-responsive factor ERF (ERF 9, ERF 72, WIN1), MYB (MYB 44, MYB 59, MYB113), WRKY (WRKY 15, WRKY 33), GATA TF 11, GATA TF 9, TCP (TCP1, TCP 3), homeobox-leucine zipper protein HAT5, and NAC 29. Previous studies also established a connection between CMS formation and differential expression of many of these TFs including MYB, NAC, bHLH, WRKY, and AP2/ERF (Li et al. 2015; Yang et al. 2018; Xing et al. 2018; Lv et al. 2020).

Consistent with the finding of Wu et al. (2017) in cotton, we also observed change in expression level of the genes involved in circadian rhythm and plant-pathogen interaction. A growing body of evidence supports role of circadian clock in regulating various aspects of plant life cycle including floral development (Guadagno et al. 2018).

Conclusion

We described genome-wide transcriptional changes between CMS line and maintainer line during two different stages of floral bud development. We employed high-throughput RNA sequencing to elucidate candidate genes and multiple regulatory pathways associated with the aberrations caused in the normal process of flower and pollen development. Consistent with the visual examination, we observed critical changes towards advanced stage of floral bud development. We hypothesize insufficient ATP formation, premature PCD, and

Table 3 Details of transcription factors showing differential expression in current study

S. no.	TF	Family	Expression
DEGs between stage S1 and stage S2 of ICPA 2043			
1	*C.cajan_41441*	bZIP	Upregulated
2	*C.cajan_11771*	bHLH	Downregulated
3	*C.cajan_24723*	GATA	Downregulated
4	*C.cajan_28593*	GATA	Downregulated
5	*C.cajan_39178*	Trihelix	Downregulated
6	*C.cajan_00203*	ERF	Downregulated
7	*C.cajan_40811*	G2-like	Downregulated
8	*C.cajan_25511*	MYB	Downregulated
9	*C.cajan_47507*	MYB	Downregulated
DEGs between stage S1 and stage S2 of ICPB 2043			
1	*C.cajan_07243*	LBD	Upregulated
2	*C.cajan_16799*	WRKY	Downregulated
3	*C.cajan_19098*	NAC	Downregulated
4	*C.cajan_26764*	NAC	Downregulated
5	*C.cajan_28596*	bHLH	Downregulated
6	*C.cajan_43030*	NAC	Downregulated
7	*C.cajan_43786*	NAC	Downregulated
8	*C.cajan_22559*	ERF	Downregulated
9	*C.cajan_37494*	ERF	Downregulated
10	*C.cajan_05842*	MIKC_MADS	Downregulated
11	*C.cajan_14151*	HD-ZIP	Downregulated
12	*C.cajan_02318*	MYB	Downregulated
DEGs between stage S1 of both ICPA 2043 and ICPB 2043			
1	*C.cajan_13674*	ERF	Downregulated
DEGs between stage S2 of both ICPA 2043 and ICPB 2043			
1	*C.cajan_11771*	bHLH	Upregulated
2	*C.cajan_24723*	GATA	Upregulated
3	*C.cajan_26410*	bHLH	Upregulated
4	*C.cajan_28593*	GATA	Upregulated
5	*C.cajan_28642*	GATA	Upregulated
6	*C.cajan_29039*	bHLH	Upregulated
7	*C.cajan_33002*	SBP	Upregulated
8	*C.cajan_34129*	bHLH	Upregulated
9	*C.cajan_39178*	Trihelix	Upregulated
10	*C.cajan_40294*	TCP	Upregulated
11	*C.cajan_41268*	TCP	Upregulated
12	*C.cajan_00203*	ERF	Upregulated
13	*C.cajan_26507*	MYB	Upregulated
14	*C.cajan_47507*	MYB	Upregulated
15	*C.cajan_05407*	WRKY	Downregulated
16	*C.cajan_07789*	LBD	Downregulated
17	*C.cajan_10762*	bZIP	Downregulated
18	*C.cajan_13689*	WRKY	Downregulated
19	*C.cajan_14327*	WRKY	Downregulated
20	*C.cajan_15679*	bHLH	Downregulated
21	*C.cajan_17609*	bHLH	Downregulated
22	*C.cajan_17807*	NAC	Downregulated
23	*C.cajan_19098*	NAC	Downregulated
24	*C.cajan_20225*	NAC	Downregulated
25	*C.cajan_21058*	GRAS	Downregulated
26	*C.cajan_24537*	WRKY	Downregulated
27	*C.cajan_24620*	NF-YB	Downregulated
28	*C.cajan_26275*	WRKY	Downregulated
29	*C.cajan_26764*	NAC	Downregulated
30	*C.cajan_26864*	bHLH	Downregulated
31	*C.cajan_28596*	bHLH	Downregulated
32	*C.cajan_30472*	NAC	Downregulated
33	*C.cajan_39766*	GRAS	Downregulated
34	*C.cajan_41441*	bZIP	Downregulated
35	*C.cajan_43030*	NAC	Downregulated
36	*C.cajan_43785*	NAC	Downregulated
37	*C.cajan_43786*	NAC	Downregulated
38	*C.cajan_43923*	WRKY	Downregulated
39	*C.cajan_08320*	ERF	Downregulated
40	*C.cajan_13804*	ERF	Downregulated

Table 3 (continued)

S. no.	TF	Family	Expression
41	*C.cajan_22559*	ERF	Downregulated
42	*C.cajan_23477*	ERF	Downregulated
43	*C.cajan_25482*	ERF	Downregulated
44	*C.cajan_27281*	ERF	Downregulated
45	*C.cajan_31784*	ERF	Downregulated
46	*C.cajan_35277*	ERF	Downregulated
47	*C.cajan_01670*	HD-ZIP	Downregulated
48	*C.cajan_24870*	HD-ZIP	Downregulated
49	*C.cajan_14151*	HD-ZIP	Downregulated
50	*C.cajan_07787*	C2H2	Downregulated
51	*C.cajan_25566*	C2H2	Downregulated
52	*C.cajan_15613*	MYB	Downregulated
53	*C.cajan_21152*	MYB	Downregulated
54	*C.cajan_24526*	MYB	Downregulated
55	*C.cajan_08081*	MYB_related	Downregulated
DEGs between ICPA 2043 and ICPB 2043 irrespective of floral bud development stage			
1	*C.cajan_04102*	GATA	Upregulated
2	*C.cajan_19191*	B3	Upregulated
3	*C.cajan_42796*	C2H2	Upregulated
4	*C.cajan_02496*	MYB	Upregulated
5	*C.cajan_04036*	MYB	Upregulated
6	*C.cajan_26507*	MYB	Upregulated
7	*C.cajan_04665*	GRAS	Downregulated
8	*C.cajan_05407*	WRKY	Downregulated
9	*C.cajan_17609*	bHLH	Downregulated
10	*C.cajan_19098*	NAC	Downregulated
11	*C.cajan_20225*	NAC	Downregulated
12	*C.cajan_30472*	NAC	Downregulated
13	*C.cajan_39766*	GRAS	Downregulated
14	*C.cajan_41441*	bZIP	Downregulated
15	*C.cajan_43923*	WRKY	Downregulated
16	*C.cajan_07410*	ERF	Downregulated
17	*C.cajan_13804*	ERF	Downregulated
18	*C.cajan_35277*	ERF	Downregulated
19	*C.cajan_46958*	ERF	Downregulated
20	*C.cajan_24526*	MYB	Downregulated

defective pollen wall formation as the possible reason to explain occurrence of male sterility in A4-CMS pigeonpea. Besides leveraging the functional genomics of CMS trait, our current dataset provides resource for future transcriptomic and genomic studies in pigeonpea.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-021-00775-y.

Acknowledgements AB acknowledges support from Centre for Agricultural Bioinformatics (CABin: AGENIASRICOP201501000047) scheme of Indian Council of Agricultural Research (ICAR), New Delhi.

Author contribution AB conceptualized the idea and planned the experiments. AB, RKS, SNSJ, DD, and MR carried out the experiments. AR, PG, and AB performed the analysis and interpretation. AB prepared the original draft with support from PG, AT, MR, RKS, RKV, and NPS. FS and IPS contributed to field experimentation and sampling. RKV and NPS reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

Alexa A, Rahnenfuhrer J (2019) topGO: enrichment analysis for gene ontology. R package version 2.36.0.

An H, Yang ZH, Yi B, Wen J, Shen JX, Tu JX, Ma C, Fu T (2014) Comparative transcript profiling of the fertile and sterile flower buds of pol CMS in *B. napus*. BMC Genomics 15:258

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Blighe K, Rana S, Lewis M (2019) EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. R package version 1.4.0.

Bohra A, Jha UC, Adhimoolam P, Bisht D, Singh NP (2016) Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. Plant Cell Rep 35:967–993

Bohra A, Saxena KB, Varshney RK, Saxena RK (2020) Genomics assisted breeding for pigeonpea improvement. Theor Appl Genet. https://doi.org/10.1007/s00122-020-03563-7

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65:579–606

Ding B, Hao M, Mei D, Zaman QU, Sang S, Wang H, Wang W, Fu L, Cheng H, Hu Q (2018) Transcriptome and hormone comparison of three cytoplasmic male sterile systems in *Brassica napus*. Int J Mol Sci 19:4022

Guadagno CR, Ewers BE, Weinig C (2018) Circadian rhythms and redox state in plants: till stress do us part. Front Plant Sci 9:247

Hamid R, Marashi H, Tomar RS, Malekzadeh Shafaroudi S, Sabara PH (2019) Transcriptome analysis identified aberrant gene expression in pollen developmental pathways leading to CGMS in cotton (*Gossypium hirsutum L.*). PLoS One 14:e0218381

Horn R, Gupta KJ, Colombo N (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion 19:198–205

Huang MD, Hsing YI, Huang AHC (2011) Transcriptome of the anther sporophyte: availability and uses. Plant Cell Physiol 9:1459–1466

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12:357–360

Li J, Dai X, Liu T, Zhao PX (2012) LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes. Nucleic Acids Res 40(Database issue):D1221–D1229

Li J, Han S, Ding X, He T, Dai J, Yang S, Gai J (2015) Comparative transcriptome analysis between the cytoplasmic male sterile line NJCMS1A and its maintainer NJCMS1B in soybean (*Glycine max* (L.) Merr.). PLoS One 10:e0126771

Li Y, Qin T, Wei C, Sun J, Dong T, Zhou R, Chen Q, Wang Q (2019) Using transcriptome analysis to screen for key genes and pathways related to cytoplasmic male sterility in cotton (*Gossypium hirsutum* L.). Int J Mol Sci 20:5120

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930

Lin S, Dong H, Zhang F, Qiu L, Wang F, Cao J, Huang L (2014) *BcMF8*, a putative arabinogalactan protein-encoding gene, contributes to pollen wall development, aperture formation and pollen tube growth in *Brassica campestris*. Ann Bot 113:777–788

Lin S, Miao Y, Su S, Xu J, Jin L, Sun D et al (2019) Comprehensive analysis of Ogura cytoplasmic male sterility-related genes in turnip (*Brassica rapa* ssp. *rapifera*) using RNA sequencing analysis and bioinformatics. PLoS One 14:e0218029

Liu C, Ma N, Wang PY, Fu N, Shen HL (2013) Transcriptome sequencing and de novo analysis of a cytoplasmic male sterile line and its near-isogenic restorer line in chili pepper (*Capsicum annuum* L.). PLoS One 8:e65209

Liu Q, Lan Y, Wen C, Zhao H, Wang J, Wang Y (2016) Transcriptome sequencing analyses between the cytoplasmic male sterile line and its maintainer line in welsh onion (*Allium fistulosum* L.). Int J Mol Sci 17:1058

Liu Y, Wei G, Xia Y, Liu X, Tang J, Lu Y, Lan H, Zhang S, Li C, Cao M (2018) Comparative transcriptome analysis reveals that tricarboxylic acid cycle-related genes are associated with maize CMS-C fertility restoration. BMC Plant Biol 18:190

Liu Z, Liu Y, Sun Y, Yang A, Li F (2020) Comparative transcriptome analysis reveals the potential mechanism of abortion in tobacco *sua*-cytoplasmic male sterility. Int J Mol Sci 21:2445

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−ΔΔCT) method. Methods 25:402–408

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Lv J, Liu Z, Liu Y, Ou L, Deng M, Wang J, Song J, Ma Y, Chen W, Zhang Z, Dai X, Zou X (2020) Comparative transcriptome analysis between cytoplasmic male-sterile line and its maintainer during the floral bud development of pepper. Horticultural Plant J 6:289–298

Ma J, Wei H, Song M, Pang C, Liu J, Wang L, Zhang J, Fan S, Yu S (2012) Transcriptome profiling analysis reveals that flavonoid and ascorbate-glutathione cycle are important during anther development in upland cotton. PLoS One 7:e49244

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics 21:3448–3449

Masuko H, Endo M, Saito H, Hakozaki H, Park JI, Kawagishi-Kobayashi M, Takada Y, Okabe T, Kamada M, Takahashi H, Higashitani A, Watanabe M (2006) Anther-specific genes, which expressed through microsporogenesis, are temporally and spatially regulated in model legume, *Lotus japonicus*. Genes Genet Syst 81:57–62

Pazhamala LT, Shilp S, Saxena RK, Garg V, Krishnamurthy L, Verdier J, Varshney RK (2017) Gene expression atlas of pigeonpea and its application to gain insights into genes associated with pollen fertility implicated in seed formation. J Exp Bot 68:2037–2054

Saxena KB, Sultana R, Mallikarjuna N, Saxena RK, Kumar RV, Sawargaonkar KL (2010) Male-sterility systems in pigeonpea and their role in enhancing yield. Plant Breed 129:125–134

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Shukla P, Singh NK, Kumar D, Vijayan S, Ahmed I, Kirti PB (2014) Expression of a pathogen-induced cysteine protease (*AdCP*) in tapetum results in male sterility in transgenic tobacco. Funct Integr Genomics 14:307–317

Sinha P, Saxena KB, Saxena RK, Singh VK, Suryanarayana V, Sameer Kumar V, Katta MAVS, Khan AW, Varshney RK (2015) Association of nad7a gene with cytoplasmic male sterility in pigeonpea. Plant Genome 8:1–12

Su S, Higashiyama T (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. Plant Reprod 31:67–75

Tuteja R, Saxena RK, Davila J, Shah T, Chen W, Xiao YL, Fan G, Saxena KB, Alverson AJ, Spillane C, Town C, Varshney RK (2013) Cytoplasmic male sterility-associated chimeric open reading frames identified by mitochondrial genome sequencing of four *Cajanus* genotypes. DNA Res 20:485–495

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotech 30:83–89

Wu J, Zhang M, Zhang B, Zhang X, Guo L, Qi T, Wang H, Zhang J, Xing C (2017) Genome-wide comparative transcriptome analysis of CMS-D2 and its maintainer and restorer lines in upland cotton. BMC Genomics 18:454–465

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 39: W316–W322

Xing M, Sun C, Li H, Hu S, Lei L, Kang J (2018) Integrated analysis of transcriptome and proteome changes related to the Ogura cytoplasmic male sterility in cabbage. PLoS One 13:e0193462

Yan X, Dong C, Yu J, Liu W, Jiang C, Liu J, Hu Q, Fang X, Wei W (2013) Transcriptome profile analysis of young floral buds of fertile and sterile plants from the self pollinated offspring of the hybrid between novel restorer line NR1 and Nsa CMS line in *Brassica napus*. BMC Genomics 14:26

Yang Y, Dong C, Yu J, Shi L, Tong C, Li Z, Huang J, Liu S (2014) Cysteine protease 51 (CP51), an anther-specific cysteine protease gene, is essential for pollen exine formation in *Arabidopsis*. Plant Cell Tissue Organ Cult 119:383–397

Yang Y, Bao S, Zhou X, Liu J, Zhuang Y (2018) The key genes and pathways related to male sterility of eggplant revealed by comparative transcriptome analysis. BMC Plant Biol 18:209–212

Zhang D, Liu D, Lv X, Wang Y, Xun Z, Liu Z, Li F, Lu H (2014) The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in *Arabidopsis*. Plant Cell 26:2939–2961