Expression profiling of immature florets of IR58025A, a wild-abortive cytoplasmic male sterile line of rice and its cognate, isonuclear maintainer line, IR58025B

K. Pranathi¹ · M. B. Kalyani¹ · B. C. Viraktamath¹ · S. M. Balachandran¹ · S. K. Hajira¹ · P. Koteswar Rao¹ · S. R. Kulakarni¹ · G. Rekha¹ · M. Anila¹ · M. B. V. N. Koushik¹ · P. Senguttuvel¹ · A. S. Hariprasad¹ · S. K. Mangrautia¹ · M. S. Madhav¹ · R. M. Sundaram¹

Received: 2 January 2019 / Accepted: 10 June 2019 / Published online: 21 June 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

Interaction between gene products encoded by the cytoplasm and nucleus form the core of wild abortive cytoplasmic male sterile (WA-CMS) system of hybrid breeding in rice. Gaining insights into such interactions can be helpful in the development of better three-line rice hybrids and also identify novel male sterility systems. In the present study, the whole transcriptome profiles of immature florets of IR58025A, a WA-CMS line and its isonuclear maintainer line, IR58025B, collected at pre-anthesis stage were compared to delineate the pathways involved in pollen abortion and male sterility. Among the 774 differentially expressed transcripts (DETs), 496 were down regulated and 278 were up regulated in IR58025A compared to IR58025B. The genes associated with oxidative stress response, defense response, etc. were significantly up-regulated, while those associated with respiration, cell wall modifications, pectinesterase activity, etc. were significantly down-regulated in the WA-CMS line. Gene ontology and pathway enrichment analyses revealed the down-regulation of both nuclear and organellar genes involved in key metabolic processes of cell respiration, photosynthesis and other energy yielding metabolites in IR58025A, relative to IR58025B, indicating a general shift toward conservation of energy and other key resources in the florets of WA-CMS line. The data derived from RNA-Seq analysis were validated through qRT-PCR analysis. Based on the results obtained, it can be hypothesized that pollen abortion principally occurs due to up-regulation of pathways leading to oxidative stress leading to energy starvation conditions in consonance with reduced expression of genes associated with the cell wall formation, respiration, and other key metabolic processes.

Keywords WA-CMS · Maintainer · IR58025A · IR58025B · RNA-Seq · Hybrid rice

Introduction

Cytoplasmic male sterility (CMS) is a common trait among many higher plants and it is maternally inherited. CMS was discovered in wild germplasm or is obtained by inter- or intra-subspecies backcrossing and is generally known to arise due to imbalance in the interaction between nuclear and mitochondrial genes. The trait of CMS circumvents the need for removal of anthers and helps in successful and large-scale cultivation of heterotic hybrids of crop plants like rice. In India, since past two decades high yielding rice hybrids are being developed based on wild abortive cytoplasmic male sterility system [WA-CMS] (Khera et al. 2012) and till today at least 102 rice hybrids have been released (<http://drdp.at.bih.nic.in>). Despite the widespread commercialization of WA-CMS-based rice hybrids in India and elsewhere, efforts on understanding the

K. Pranathi, M. B. Kalyani and R. M. Sundaram contributed equally.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1806-5>) contains supplementary material, which is available to authorized users.

✉ R. M. Sundaram
rms_28@rediffmail.com

¹ ICAR-Indian Institute of Rice Research (IIRR),
Hyderabad 500030, India

molecular basis of the traits of CMS and restoration of fertility in the hybrids are very limited.

Earlier studies have revealed that CMS trait can be attributed to genes/proteins encoded by novel chimeric open reading frames [orfs] (Fujii and Toriyama 2008; Kubo et al. 2011; Horn et al. 2014) and their interaction with proteins encoded by the nuclear genome. The molecular pathways involving production of cytotoxic CMS protein, ORF79, and ORFH79 coupled with energy deficiency leading to pollen abortion in BT-CMS and HL-CMS of rice, respectively, have been implicated with the trait of cytoplasmic male sterility (Wang et al. 2006, 2013). To identify the putative causal factors of WA-CMS trait in rice, complete sequencing of the mitochondrial genomes of the male fertile and male sterile was analyzed and a WA-CMS-specific ORF, *orf126*, was identified as a plausible candidate (Bentolila and Stefanov 2012). Later, rearrangements were identified in mitochondrial genomic region downstream of *rpl5* gene and a WA-CMS causal chimeric orf named WA352 (comprising of three rice mitochondrial genomic segments, viz., *orf284*, *orf224*, and *orf288*) and one segment of unknown origin was identified to be responsible for the trait of male sterility associated with WA-CMS. The novel orf was observed to encode a 352 amino acid putative protein with three trans-membrane segments (Luo et al. 2013). Later, it was established that *orf126* was indeed a part of WA352 (Huang et al. 2014; Pranathi et al. 2016). Even though these studies have concluded that WA352 is the causal gene for WA-CMS trait, functional studies that elucidate pathways leading to pollen abortion in WA-CMS lines remain unclear. A comparative, whole genome expression study of the WA-CMS line and its cognate, isonuclear maintainer line can give vital insights into molecular mechanisms underlying WA352-induced male sterility and also on mitochondrial–nuclear incompatibility. In view of these points, through the present study, we carried out global RNA sequencing of transcript profiles of florets (at pre-anthesis stage) of two isonuclear lines, (i) a WA-CMS line, IR58025A which is the maternal parent for the Indian rice hybrids, KRH2, DRRH1, Sahyadri, CoRH2, ADTRH1, Narendra Sankar Dhan-2, Pant Sankar Dhan-1 (Sundaram et al. 2008) and (ii) its cognate maintainer, IR58025B. The focus of the study was to understand and unravel the gene expression patterns and molecular events encoded by both mitochondrial and nuclear genomes, leading to male sterility in the WA-CMS line and their comparison with the its cognate, isonuclear maintainer line.

Materials and methods

Plant materials

The plant materials utilized in the study include two isonuclear lines, IR58025A (wild-abortion cytoplasmic

male-sterile line) and its cognate fertile maintainer line, IR58025B. About 40–50 seedlings of each genotype were planted in IIRR research farm during both successive rice-growing seasons in crop grown in the experiment farm of ICAR-Indian Institute of Rice Research, Hyderabad, India in 2015. Based on the days to flowering, the approximate days to booting or emergence of panicle from flag leaf were estimated for collecting the unopened florets of the whole panicle in triplicates. Under highly sterile conditions, in RNase Away reagent (Ambion, USA) treated surfaces, the entire panicles of both lines were collected at sporophytic stage (containing immature anthers) and stored immediately in liquid nitrogen (−196 °C). The remaining panicles of the plants from which the immature florets were collected, were bagged to assess the seed setting in those plants.

RNA extraction and library preparation

The total RNA was isolated from the florets tissue of IR58025A and IR58025B using Trizol reagent (Invitrogen) according to the manufacturer's instructions in two technical replicates for each sample. The quantity and quality of RNA were measured using Nanodrop-1000 (Thermo Fisher Scientific), and their quality was further assessed in the Agilent 2100 Bio analyzer (Agilent Technologies). Samples with RNA integrity number (RIN) > 8 were used for library preparation. An aliquot of total RNA was kept aside for single strand cDNA synthesis for later use to analyze relative gene expression through real-time quantitative PCR. The TruSeq mRNA stranded kit (Illumina, California, US) was used to prepare Illumina RNA-Seq libraries. After checking for appropriate library concentration and insert size using Agilent 2100 bio analyzer (Agilent technologies), the libraries were pooled according to their effective concentration with expected data volume and fed to HiSeq2500 sequencer (Illumina) at the sequencing facility at Nucleome informatics Pvt. Ltd. (Hyderabad, India).

RNA-Seq and data analysis

The raw reads obtained, were filtered using FASTQC tool kit (Andrews 2010). The high quality paired end clean reads were mapped against the reference rice genome (Nipponbare) IRGSP 1.0 using TopHat (Trapnell et al. 2010), allowing up to a mismatch of two base pairs. The mapped reads were visualized using IGV and quantified using HTSeq v0.6.1 (<http://www-huber.embl.de/HTSeq>) in union mode. The Fragments per Kilo base of transcript sequence per Millions base pairs-sequenced (FPKM) method was used for estimating gene expression levels (Trapnell et al. 2010) with threshold parameter set as 0.1 or 1. Cufflinks 2.1.1 assembler, which uses all the mapping information as input for prediction of novel genes or exons was deployed and the

assembled transfrags were then compared to the reference transcripts to determine if they are sufficiently different to be considered novel. All the read counts were normalized using DEGseq1.12.0 (<http://bioinfo.au.tsinghua.edu.cn/software/degseq>) with a threshold setup as $|\log_2(\text{Fold Change})| > 1$ and $q\text{-value} < 0.005$, where q value represented the normalized p value that was based on Poisson distribution. The significant differentially expressed transcripts (DETs) were retrieved through cutting off based on p and FDR (false discovery rate) values < 0.05 . The DETs, which are possibly associated with WA-CMS trait, were identified and functionally significant pathways were clustered to generate a heat map using Multiple experiment viewer, MeV v4.9.0 (Saeed et al. 2003). The transcription factors (TF) identified in differentially expressed transcripts were analyzed using iTAK v1.2 (Zheng et al. 2016). Gene ontology (GO) enrichment was performed using Goseq (Young et al. 2010) to describe cellular component, molecular function, and biological process of each gene.

Alternative splicing and variant identification

The reads that span splicing junctions were considered for analysis of alternative splicing events using rMATS 3.0.8 with default parameters. The alternative splicing (AS) events are classified into five broad categories including skipped exon, mutually exclusive exon, alternative 5' splice site, alternative 3' splice site, retained intron (Barbazuk et al. 2008). Similarly, the sequence variants were also identified using GATK3 v3.4 with parameters including mapping quality (MQ) < 40.0 and quality depth (QD) < 2.0 .

Pathway analysis

The Kyoto Encyclopedia of Genes and Genomes Pathway (KEGG) enrichment analysis (KOBAS v2.0) was used to identify the significantly enriched metabolic pathways or signal transduction pathways associated with differentially expressed transcripts compared with the whole genome background. Also, to investigate the involvement of specific pathways associated with CMS, the RNA-Seq data were supplemented with data in public domain, where the known mitochondrial genes, mostly associated with respiration pathway as listed in Oryzabase (<http://shigen.nig.ac.jp/organelle/organelleListAction>) were considered for back mapping with transcripts.

Validation of RNA-Seq gene expression through real-time quantitative PCR

Real-time PCR was used to validate 18 selected transcripts, which are associated with respiration and other key metabolic pathways. All the primers (details given in Supplementary

Table 1) were designed using, Quant Prime (<http://quantprime.mpimp-golm.mpg.de>) and synthesized at Integrated DNA Technologies Inc. (Iowa City, IA). The first strand cDNA was synthesized using AMV reverse transcriptase kit (New England Biolabs). Reaction was performed in PCR LC 96-well plate (Roche Light Cycler® 96, Roche, USA) using three biological replicates and two technical replicates for each sample. Each PCR reaction mix volume was 20 μl and contained 10 μl of 2 $\times$ SYBR FAST qPCR Universal Mix (Kapa bio systems), 10 pmol each of forward and reverse primers, 0.4 μl of high rox dye, and 1–2 μl (~50 ng) of the cDNA samples. The reaction was initiated in the system (Roche light cycler) with a pre-incubation at 55 °C for 2 min and initial denaturation at 95 °C for 3 min followed by 40 cycles of denaturation at 95 °C for 30 s and annealing at 55 °C for 30 s, and extension at 72 °C for 1 min. *Actin* was used as internal control. The expression ratios were calculated using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001).

Results

Global gene expression of RNA-Seq data

A total of 49 million and 45 million raw reads, respectively, were obtained from the IR58025A and its isonuclear maintainer, IR58025B at pre-anthesis stage. After filtering low-quality sequences and adaptor sequences, 47 million and 43 million clean reads of IR58025A and IR58025B were acquired, respectively. Reference alignment of IR58025A and IR58025B high-quality reads with Nipponbare genome gave mapping percentage of 78 and 80, respectively. In addition, 9678 novel transcripts were identified in IR58025A, while a total of 9685 such transcripts were identified in IR58025B (Table 1).

Of the total mapped transcripts, 27,222 transcripts were commonly expressed between the two isonuclear lines (i.e., IR58025A and IR58025B). After filtering for differentially expressed transcripts (DETs) based on P and FDR values (i.e., < 0.05) as well as $\log_2$ fold change (i.e., > 1.5), 774 transcripts were obtained with 278 up-regulated and 496 down-regulated in IR58025A as compared to IR58025B (Fig. 1). Among the up-regulated differentially expressed transcripts (DETs), chitinases showed the highest expression in IR58025A and among down-regulated DETs, pectinesterases have shown lowest expression in IR58025A compared to IR58025B.

Expression of mitochondrial genes associated with respiration

The mitochondrial genes of reference Nipponbare genome, associated with respiration showed differential expression at pre-anthesis stage. Out of these 54 mitochondrial genes

Table 1 Statistics of mapping transcriptome data to reference Nipponbare genome

Sample name	IR58025A	IR58025B
Raw reads	49147620	45206446
High-quality reads with $Q \geq 20$	47285794 (96.21%)	43185622 (95.91%)
Total mapped reads	38120318 (80.62%)	33999221 (78.73%)
GC content	50.14%	51.89%
Multiple mapped reads	429242 (0.91%)	442863 (1.03%)
Uniquely mapped reads	37691076 (79.71%)	33556358 (77.7%)
Non-splice reads	29630198 (62.66%)	26039561 (60.3%)
Splice reads	8060878 (17.05%)	7516797 (17.41%)
Novel transcripts	9678 (0.020%)	9685 (0.022%)

(<http://shigen.nig.ac.jp/rice/oryzabase>), 21 expressed at pre-anthesis stage in both the lines. Interestingly, 20 out of 21 mitochondrial genes were down-regulated in IR58025A at pre-anthesis stage, except *orf165* that was slightly up-regulated in IR58025A. Eight mitochondrial genes (*cob*, *cox2*, *nad7*, *orf162*, *orf183*, *orf224*, *rps2*, and *rps4*) were expressed exclusively in maintainer line, IR58025B with no expression in IR58025A. The mitochondrial gene, *atp9* (GI: 23495404) showed the highest transcript abundance of 22.9471 FPKM in IR58025B and FPKM of 8.3228 in IR58025A. The next two genes with increased transcript abundance in IR58025B were, *atp1* (GI: 60498752) with FPKM of 16.945 and *rps12* (GI: 23495391) with FPKM of 16.121. As compared to the expression of nuclear genes, mitochondrial gene expression was found to be relatively low. Also, the abundance of differentially expressed transcripts associated with GO terms of cellular organelles i.e.,

mitochondria and chloroplast were comparatively lesser in number than the GO terms associated with nuclear expression (Supplementary Fig. 1).

A global transcriptomic analysis of IR58025A and IR58025B revealed that three key genes associated with F₁ subunit of Complex V (ATP synthase) of oxidative phosphorylation chain were either down-regulated or did not express in IR58025A. One of the key genes, ATP synthase (OS02G0750100) with 94.40 FPKM in IR58025A and 188.93 FPKM in IR58025B showed onefold down regulation while no expression of two other genes belonging to P-type ATPases family, ATPase 8 (OS02G0797300) and ATPase 6 (OS06G0181500) was recorded in IR58025A. Normal expression of these genes was observed in fertile line, IR58025B. Interestingly, putative inorganic pyrophosphatase (OS06G0178900) of Complex V was the only up-regulated gene in IR58025A (Fig. 2a). Cytochrome C1 (OS02G0216300) associated with Complex III subunit 4 of mitochondrial respiratory chain was also significantly down-regulated in IR58025A and expressed at normal levels in IR58025B (Table 2—placed at the end of manuscript).

DETs involved in photosynthesis

Photosynthesis is one of the important anabolic pathway required for plants. Numerous genes associated with not only photosynthesis but also its accessory components like photosynthesis antenna proteins, chloroplast, and carbon fixation were significantly down-regulated in the WA-CMS line with a fold change range of 1–1.8 (Fig. 2b and Table 2). The genes associated with photosystem-I reaction center, Psal-PS1 and photosystem II subunit viz., *Psbs*, *Psb27*, and oxygen-evolving complex of photosystem II, *PsbO*, chlorophyll

Fig. 1 Representation of expressed transcripts and DETs

Table 2 List of DETs associated with specific pathways

Gene	Gene name	IR58025A	IR58025B	Fold change	P value	Q value
<i>Photosynthesis</i>						
OS01G0869800	PsbS	453.4117528	926.2860856	− 1.0306	1.64E−25	3.21E−23
OS01G0501800	PsbO	2868.397641	5792.856219	− 1.014	1.04E−145	2.58E−142
OS03G0333400	Psb27	283.8722631	661.8079807	− 1.2212	1.03E−25	2.03E−23
OS09G0481200	PsaG-PS1	1507.714458	3105.256965	− 1.0423	1.79E−83	2.12E−80
<i>Photosynthesis antenna proteins</i>						
OS06G0320500	Cab26-Lcha1	2621.199211	6194.727419	− 1.2408	1.32E−233	6.03E−230
OS01G0600900	CAB2R-Lchb1	9055.805147	18560.97104	− 1.0354	0	0
OS09G0346500	Lchb-1-1	4806.007807	10628.71835	− 1.1451	0	0
OS09G0439500	Lcha2	43.84062762	102.6483569	− 1.2274	3.35E−05	0.0010107
<i>Carbon fixation in photosynthetic organisms</i>						
OS12G0169600	Fructose-bisphosphate aldolase	26.25958411	76.4095915	− 1.5409	9.89E−06	0.0003386
OS03G0267300	FBPase	177.0982186	449.3028149	− 1.3431	3.03E−21	4.87E−19
OS12G0274700	RBCS	198.0946878	659.4291915	− 1.735	8.35E−47	3.51E−44
OS05G0405000	PPDK1	99.7192309	199.7462057	− 1.0022	2.54E−06	9.83E−05
<i>Oxidative phosphorylation</i>						
OS02G0750100	ATP synthase	94.40012538	188.9335277	− 1.001	4.89E−06	0.00018054
OS06G0181500	ATPase 6	0	30.203414	− 6.7534	4.08E−08	2.20E−06
OS02G0797300	ATPase 8	0	32.00552701	− 6.574	1.45E−08	8.23E−07
OS06G0178900	Putative inorganic pyrophosphatase	67.80459776	28.18504743	1.2665	2.11E−06	8.29E−05
OS02G0216300	Cytochrome C1	2006.198631	948.920625	1.0801	1.1E−119	2.15E−116
<i>Starch and sucrose metabolism</i>						
OS03G0757900	UGD3	107.9498468	29.77090688	1.8584	3.27E−14	3.12E−12
OS12G0443500	UGD4	78.77875232	33.23096385	1.2453	4.28E−07	1.93E−05
OS09G0369400	TPP7	0	55.72133418	− 8.3739	2.23E−12	1.83E−10
OS10G0465700	Beta-amylase	55.76662211	1132.015307	− 4.3433	1.97E−228	7.70E−225
OS03G0141200	Beta-amylase	131.0179676	1164.381256	− 3.1517	7.47E−181	2.27E−177
OS05G0366600	BGLU22	30.17892502	75.47249273	− 1.3224	0.00013286	0.0034038
OS07G0656200	BGLU26	39.19340911	10.23600188	1.937	2.84E−06	0.00010784
<i>Phenylalanine metabolism and phenylpropanoid biosynthesis and amino acid metabolism</i>						
OS02G0627100	Phenylalanine ammonia-lyase	160.8049585	487.2192726	− 1.5993	9.53E−31	2.46E−28
OS06G0165800	Caffeoyl-CoA O-methyltransferase	118.1401332	244.2944392	− 1.0481	4.89E−08	2.60E−06
OS03G0152300	Class III peroxidase 36	5.151133769	41.9531908	− 3.0258	1.01E−07	5.07E−06
OS01G0962700	Putative peroxidase	58.28619841	27.39211771	1.0894	6.76E−05	0.0018799
OS01G0327400	Class III peroxidase 16	175.5304822	30.49175208	2.5252	3.39E−30	8.50E−28
OS02G0741200	Class III peroxidase 31	2.127642209	35.8980911	− 4.0766	1.80E−08	1.01E−06
OS05G0135500	Putative peroxidase	32.36255781	8.073466272	2.0031	1.45E−05	0.00047818
OS05G0499300	PRX74	170.9952449	63.43437785	1.4306	3.89E−16	4.40E−14
OS09G0481400	Putative Caffeoyl-CoA O-methyl-transferase	37.96161625	7.929297231	2.2593	5.26E−07	2.31E−05
OS05G0366600	BGLU22	30.17892502	75.47249273	− 1.3224	0.00013286	0.0034038
OS07G0656200	BGLU26	39.19340911	10.23600188	1.937	2.84E−06	0.00010784
OS04G0688100	Cationic peroxidase	80.34648868	3.027549852	4.73	1.37E−21	2.24E−19
OS05G0399400	Chitinase 9	183.2011923	11.10101612	4.0447	2.33E−44	9.21E−42
OS05G0399700	Chitinase 7	12532.87643	1686.849859	2.8933	0	0
OS01G0660200	OsChib3a	138.3527342	18.74197527	2.884	4.17E−27	8.98E−25
<i>ROS and PCD</i>						
OS01G0974200	Metallothionein-like protein 2B	525.9195597	1059.858702	− 1.011	7.33E−28	1.63E−25
OS12G0542000	Rapid alkalization factor	10.47023929	45.34116326	− 2.1145	1.10E−05	0.00037389
OS12G0541700	Rapid alkalization factor	34.37821885	11.10101612	1.6308	7.49E−05	0.0020629

Table 2 (continued)

Gene	Gene name	IR58025A	IR58025B	Fold change	P value	Q value
OS01G0357900	Rapid alkalization factor	114.8926793	264.4060204	− 1.2025	6.77E−11	4.89E−09
OS01G0201400	Protein RALF-like 33 rapid alkalini- zation factor (RALF)	0	33.59138645	− 7.6437	1.77E−08	9.93E−07
OS01G0294700	Class III peroxidase 11	49.04775197	22.56245485	1.1203	0.00019682	0.0048427
OS04G0689000	Class III peroxidase 64	34.60218119	5.11800094	2.7572	1.26E−07	6.19E−06
OS10G0529800	Probable glutathione S-transferase GSTU6	0	106.7571745	− 8.8969	3.67E−21	5.82E−19
OS09G0241700	Glutathione S-transferase T3	0	43.17862765	− 6.4322	4.20E−11	3.09E−09
OS10G0378100	Cytochrome P450 family protein	24.91581008	99.62080703	− 1.9994	4.13E−10	2.77E−08
OS02G0503900	Putative cytochrome P450	12.09396624	56.58634842	− 2.2262	3.45E−07	1.59E−05
OS03G0345700	Heavy metal-associated domain- containing protein	48.37586496	177.760427	− 1.8776	1.87E−15	2.00E−13
OS03G0152000	Heavy metal-associated domain- containing protein	27.93930164	150.9449855	− 2.4337	8.62E−19	1.17E−16
OS03G0222100	MA3 domain-containing protein	6.382926627	92.26818596	− 3.8535	9.30E−19	1.26E−16
OS03G0100100	Pectinesterase inhibitor domain containing protein	2.911510391	30.34758304	− 3.3817	1.64E−06	6.51E−05
OS03G0309400	Pectinesterase family protein	0	23.21121553	− 6.6954	1.46E−06	5.87E−05
OS04G0458900	Pectinesterase	0	59.75806731	− 8.4748	4.78E−13	4.17E−11
OS08G0450200	Pectinesterase	2.071651625	71.72409768	− 5.1136	3.57E−17	4.38E−15
OS09G0433700	Pectinesterase	0	18.09321459	− 5.336	2.08E−05	0.00065502
OS11G0683700	Pectinesterase	0	56.00967226	− 6.8076	8.39E−14	7.77E−12
OS11G0683800	Pectinesterase	0	73.74246425	− 8.7781	2.77E−15	2.88E−13
OS05G0360900	Pectin methylesterase inhibitor	0	56.08175678	− 6.3832	5.36E−14	5.05E−12
OS01G0311700	Similar to pectin methylesterase inhibitor	0	29.69882236	− 5.5916	4.35E−08	2.33E−06
OS01G0135700	Calcium-binding protein CML16	12.42990975	136.6001659	− 3.4581	8.89E−25	1.66E−22
OS01G0949500	CML10	119.4839072	341.6085416	− 1.5155	3.28E−20	4.92E−18
OS01G0955100	CML31	8.958493511	40.65566944	− 2.1821	2.05E−05	0.00064774
OS03G0331700	CML27	37.0657669	86.78976242	− 1.2274	0.00013639	0.0034877
OS05G0380900	CML15	15.17344838	389.7610012	− 4.683	4.27E−83	4.86E−80
OS06G0691600	CML29	64.4451627	157.2163387	− 1.2866	7.74E−08	3.94E−06
OS08G0144100	CML32	1.567736364	19.67907404	− 3.6499	6.46E−05	0.0018142

with phenylalanine and phenylpropanoid biosynthesis, phenylalanine ammonia-lyase (ZB8; OS02G0627100), caffeoyl-CoA *O*-methyltransferase (Os06g0165800), class III peroxidases 36 (OS03G0152300), and class III peroxidases 31 (OS02G0741200).

DETs involved in ROS

Numerous transcripts associated with direct and indirect elevation of reactive oxygen species (ROS) levels [leading to programmed cell death (PCD) at pre-anthesis stage] were observed to be up-regulated in WA-CMS line, IR58025A. The comparatively low expression of five pectinesterases (OS03G0309400, OS04G0458900, OS09G0433700, OS11G0683700, OS11G0683800) and very low expression of two COBRA-like proteins along with simultaneous

depletion of energy and metabolite pool were also noticed in IR58025A, which could also be responsible for oxidative stress (Table 2 and Fig. 2d). The expression of seven peroxidases with increased transcript abundance, viz., cationic peroxidase (SPC4; OS04G0688100) with 80.34 FPKM in IR58025A and 3.02 FPKM in IR58025B was also observed, indicating up-regulation of various components associated with increased production of ROS in the WA-CMS, possibly leading to programmed cell death and abortion of pollen grain formation (Table 2). Various DETs involved in detoxification and scavenging of ROS, viz., glutathione S-transferases, cytochrome P450, metallothionein-like proteins, rapid alkalization factors, heavy metal-associated domain-containing proteins (Table 2) were all significantly down-regulated in IR58025A as compared to IR58025B indicating a selective shut down of the ROS scavenging pathways in the

WA-CMS line. More specifically, the expression of RALF-like 33-rapid alkalization factor (RALF) (OS01G0201400) and glutathione S-transferase T3 (OS09G0241700) was completely shut off in IR58025A.

Differential expression of transcription factors (TFs)

Gene regulation is usually by binding of transcription factors (TFs) at transcriptional level. About 71 differentially expressed TFs grouped into 26 different groups (comprising both up- and down-regulated transcripts) were found to be expressed at pre-anthesis stage (Fig. 3). Among them, AP2-EREBP 2-ethylene responsive binding protein domain (AP2-EREBP) containing transcription factors was the most abundant family of TFs with 14 AP2-EREBP TFs being down-regulated and 2 other were up-regulated in IR58025A. Homeo domain protein (HB) TFs, which bind to specific DNA sequence, homeobox, and mainly involved in morphogenesis were the next family of TFs found in large number. Interesting, eight TFs groups were observed to be significantly up-regulated transcripts in IR58025A and among them, MADS box group TFs (MADS 8, MADS 4, MADS 21, MADS 13, MADS 7) were the predominant one. A total of 13 TFs groups were down-regulated in IR58025A compared to IR58025B with NAC domain containing TFs noticed to be the predominant class.

Alternative splicing and SNP

About 1356 alternative splicing (AS) events with 772 events in IR58025A and 651 events in IR58025B were identified

through read span splicing junctions only in rMATS. The events were categorized as altered 3' splice site (A3SS-7.7%), altered 5' splice site (A5SS-2.8%), mixed exon (MXE-0.6%), intron retention (RI-15.4%), and skipped exon (SE-73%). Among AS event categories, SE was most predominant followed by RI (Supplementary Fig. 2). Eighteen of the significantly differentially expressed transcripts showed alternative splicing. Among these 18 DETs, 14 DETs showed SE type, 2 DETs showed RI type, and 1 DET showed A3SS type of AS events (Supplementary Table 2). Interestingly, one of the DETs with SE-type AS event, metallothionein-like protein 2B was involved in scavenging of ROS. Interestingly, the total number of sequence variants in the form of SNP included 84,080 with higher number, 42,780 SNPs in IR58025A and 41,300 SNPs in IR58025B (Supplementary Table 3) while *in-dels* were also found higher in IR58025A with 5762 and 5501 *in-dels* in IR58025B. Compared to reference genome, the transcripts with sequence variant of A/T type were high with 3445 SNPs, while G/C-type SNPs were only 2344. The abundance of SNPs in transcripts was highest on the chromosome 1 (13.4%) and the lowest on chromosome 10 (5.4%).

GO classification and enrichment analysis of DETs

Gene ontology (GO) classification analysis was performed to assign function to the DETs. The transcripts were categorized into three major groups *viz.*, biological processes (BP), molecular function (MF), and cellular component (CC) gene ontologies and were assigned appropriate GO numbers. The 774 DETs were classified into 40 different classes based on

Fig. 3 Histogram showing transcription factor groups

gene function (Fig. 4). Among the biological process GO, “metabolic process” (GO: GO:0008152) was the dominant class, followed by “cellular process” (GO:0009987) and “biological regulation” (GO:0065007). In case of molecular function group, “catalytic activity” (GO:0003824) was the largest class followed by “binding activity” (GO:0005488) and “transporter activity” (GO: 0005215). In cellular component group, the dominant classes were “cell” (GO:0005623), “cell part” (GO:0044464), and “organelle”(GO:0043226). Interestingly, two unique categories, “translation regulator” (GO:0006413) and “biological adhesion” (GO:0022610) were found specific only to down-regulated GO terms.

GO enrichment analysis of functional significance was performed to identify significantly enriched GO terms among DETs in IR58025A with GO term with $P \leq 0.05$ defined as the significant DETs-enriched GO. A total of 14 down-regulated enriched GO terms were recorded as compared to only 6 up-regulated enriched GO terms. The up-regulated DETs within the biological process-enriched GO term were largely associated with “defense response” (GO:0006952) and “stress response” to biotic stresses (GO:0009620) along with the DETs within the molecular function term associated with “sequence-specific DNA binding” (GO:0043565) activity. The other enriched molecular function GO terms viz., “oxidoreductase activity” acting on peroxide as acceptor (GO:0016684) and “phosphatidylglycerol phosphatase activity” (GO:0008962), which play a key role in mitochondrial

metabolism. Interestingly, highest number of DETs was associated with hydrolase activity (GO:0016787) GO term. No up-regulated DETs were found enriched within the cellular component GO term (Supplementary Fig. 3a).

The down-regulated DETs within biological process-enriched GO term were associated with “cell wall organization” (GO: 0071555) and “cell wall modifications” (GO:0042545) along with the DETs within molecular function GO terms associated with “pectinesterase” activity (GO:0030599) and “carboxylic ester hydrolase” activity (GO:0052689) (Supplementary Fig. 3b). The other enriched dominant GO terms of molecular function group were enzyme inhibitor (GO: 0004857), enzyme regulator (GO:0030234), molecular function regulator (GO:0098772), pectatelyase (GO:0030570) and carbon–oxygen lyase activity acting on polysaccharides (GO:0016837), solute: cation antiporter (GO:0015298) and solute: proton antiporter activities (GO:0015299). In case of cellular component group of down-regulated DETs, extracellular matrix (GO:0031012) was the only enriched GO term.

Analysis of DETs specific for metabolic pathways, possibly associated with CMS

Significantly enriched down-regulated pathways identified through KEGG analysis were “metabolic pathways” with large complexes comprising various metabolic pathways

Fig. 4 Histogram showing gene ontology functional analysis of DETs

like energy metabolism, carbohydrate, amino acid, nucleotide and lipid metabolism, and metabolism of terpenoids and polyketides, xenobiotic degradation (Fig. 5a). The other enriched biological pathways of down-regulated DETs include “photosynthesis”, “oxidative phosphorylation”, and “plant–pathogen interactions”. The results of both the GO and pathway enrichment analysis revealed that the DETs in functional class of biotic stress including proteolytic activity, pectinesterase activity, chitinases activity, and calcium-binding proteins activity were highly altered during pre-anthesis stage in IR58025A. The significantly enriched up-regulated pathways included “biosynthesis of secondary metabolites” followed by phenylpropanoid biosynthesis and phenylalanine metabolism (Fig. 5b). The other enriched up-regulated pathways were found to be associated with starch and sucrose metabolism, protein processing in endoplasmic reticulum and phagosome.

Correlation of RNA-Seq gene expression with qRT-PCR

A total of 18 genes (11 nuclear DETs and 7 mitochondrial transcripts) with functional significance were selected for validation with qRT-PCR and *Actin* was used as internal control (Fig. 6). The expression patterns of the selected genes as determined by qRT-PCR were observed to be consistent with those obtained in RNA-Seq analysis, with minor variation noticed with respect to two genes, viz., cytochrome

P450 (OS10G0378100) and glutathione S-transferase T3 (OS09G0241700). This confirms the accuracy of the RNA-Seq results obtained (Fig. 6). In addition, relative gene expression of reported candidate for WA-CMS trait, WA352 also showed the expected higher expression in IR58025A and no expression in its fertile counterpart, i.e., IR58025B (Fig. 6). All the seven mitochondrial gene associated with respiration (*atp6*, *cox2*, *cox3*, *nad1*, *nad4*, *nad5*, *nad7*) were also observed to show expression levels in qRT-PCR experiments, similar to those observed in RNA-Seq expression analysis (Supplementary Fig. 4).

Discussion

CMS technology has emerged as a boon to harness the hybrid vigor in many crop plants including rice, owing to its distinctive ability to produce sterile male gametophyte, facilitating selective hybridization between the male and female parents in a cost-effective manner. Among the CMS types deployed in rice, wild-abortive cytoplasmic male sterility (WA-CMS) is the most popular and widely deployed one in many countries including India (Hari Prasad et al. 2014). In WA-CMS lines, the male sterility phenotype is not manifested until flowering stage; hence it is possible to differentiate a sterile (i.e., WA-CMS) line from fertile one (i.e., the maintainer line) only during flowering stage (Bentolila et al. 2002). The WA-CMS line (i.e., A line) is maintained by crossing it with its cognate, isonuclear fertile maintainer

Fig. 5 Scatter plots for representation of enriched KEGG pathways

Fig. 6 Validation of expression of selected transcripts

line (i.e., B line) (Whitford et al. 2013). The manifestation of CMS trait involves layers of interaction between mitochondrial and nuclear genes (Chen and Liu 2014). To gain a better understanding of these molecular events and to study the specific gene expression patterns of mitochondrial and nuclear genomes of A and B lines, we carried out global gene expression analysis of immature florets of panicles of a popular WA-CMS line, IR58025A and its isonuclear cognate maintainer, IR58025B using Illumina HiSeq2500 platform. Owing to the isonuclear nature of the A and B lines, the differential expression arising at pre-anthesis stage in the male fertile and sterile line could be captured.

The results of Illumina HiSeq2500 analysis revealed that the transcripts from the WA-CMS line and its maintainer line aligned very well with the rice reference genome (> 78%). A high P value (0.005) and FDR value (< 0.005) were considered, while analyzing the expressed genes, so as to maintain stringent filtration to identify the significant DETs between IR58025A and IR58025B. When the expression of set of mitochondrial genes (from Nipponbare) was analyzed, as expected, the absolute expression of mitochondrial genes

was relatively low as compared to the high level of expression observed for most of the nuclear genes (Takakura et al. 2000; Chen and Liu 2014). It has been reported that the plant mitochondrial genomes have only approximately 60 known genes for the electron transfer chain, ribosomal proteins, transfer RNAs, and ribosomal RNAs (Kubo and Newton 2000). However, the proteomic analysis of plant mitochondria shows > 1000 proteins, with most of them encoded by nuclear genome (Kruft et al. 2001). This highlights the need for analyzing the dynamics of expression of mitochondrial and nuclear genes in consonance among the A and B line analyzed in the present study.

GO enrichment analysis is a commonly deployed, powerful approach used for gene expression analysis (Subramanian et al. 2005). By categorizing the global transcript changes into standardized GO terminology, metabolic shifts are unraveled, allowing detection and interpretation of possible changes that may occur at the molecular and cellular level at given conditions (Du et al. 2010). In the present study, most of the enriched GO terms were associated with down-regulated DETs in IR58025A (Fig. 4 and Supplementary

Fig. 3), indicating a shift from normal metabolism during pre-anthesis in the WA-CMS line (IR58025A) as compared to the maintainer line (IR58025B), which is the cognate isonuclear counterpart of the male-sterile line.

Down-regulated DETs in IR58025A were enriched with GO terms related to cell wall modifications (GO:0042545), cell wall organization (GO:0071555), pectin esterase activity (GO:0030599), and carboxylic ester hydrolase activity (GO:0052689). The change in cell wall structure by decreased activity of cell wall modification genes may cause cell wall integrity damage. Such cell wall integrity damage can induce stress response in plants, which is similar to pathogen attack (Bellincampi et al. 2014; Pogorelko et al. 2013). The cell wall integrity damage can also eventually lead to programmed cell death (PCD; Paparella et al. 2015). A response similar to the attack from a typical plant pathogen was observed in IR58025A, wherein the up-regulated DETs were enriched with GO terms related to defense response (GO:0006952) and stress response (GO:0009620) in specific to biotic stress response that may eventually lead to PCD. Furthermore, high number of up-regulated DETs were associated with hydrolase activity (GO:0016787), indicating a progressive increase in cell degradation during pre-anthesis stage in IR58025A. Even though there was no significant enrichment of GO terms related to cellular components in the WA-CMS line, the DETs of molecular function associated with mitochondrial metabolism viz., oxidoreductase (GO:0016684) and phosphatidylglycerol phosphatase (GO:0008962) were significantly enriched in IR58025A. Mitochondria are the sites for ROS production, and a well-controlled mechanism exists for limiting the quantity of ROS produced. However, if the threshold increases, it leads to PCD. During various stresses, the concentration of ROS in the cells can rise significantly reaching the threshold that can trigger PCD (Lamb and Dixon 1997; Tiwari et al. 2002).

The key function of mitochondria is cellular respiration, oxidative phosphorylation along with electron transport, and impairment of genes associated with these activities can lead to drop in membrane potential. It is reported in plants that functional disorders resulting in reduction of mitochondrial membrane potential leads to the depletion of ATP pool. This in turn will elicit the increase of ROS and ultimately result in PCD (Bergman et al. 2000; Tiwari et al. 2002). In HL-CMS system of rice, during microgenesis, PCD correlated with change in membrane potential of mitochondria, oxidative stress, and ROS production (Li et al. 2004). Interestingly, in the present study also, we observed decreased transcripts of most of the mitochondrial genes involved in respiration in the male-sterile line, IR58025A (Supplementary Fig. 4). As mitochondrial transcripts generally represent only a fraction of total RNA in global transcriptome analysis and also there was no additional enrichment of RNA extracted with mitochondrial fraction in our study, the expression levels

of mitochondria remained relatively lower as compared to the higher expression noticed in the genes encoded by the nuclear genome. Due to this, we could not detect the transcript reads of key genes associated with WA-CMS trait like *WA352*, the candidate gene for WA-CMS trait (Luo et al. 2013) in our study. However, the *WA352* transcripts were detected in IR58025A in qRT-PCR analysis (Fig. 6), supporting earlier reports on *WA352* as the candidate gene for WA-CMS trait (Luo et al. 2013; Pranathi et al. 2016). Surprisingly, the transcript abundance of earlier reported candidate for WA-CMS trait, *orfB* (Das et al. 2010) was found to be more in IR58025B (FPKM of 4.76) as compared to IR58025A (FPKM of 2.80). Thus, *orfB* may not be associated with the trait of WA-CMS in rice as reported earlier (Pranathi et al. 2016).

Apart from mitochondrial genes, a few key nuclear genes, especially ATP synthase (OS02G0750100) and P-type ATPases, ATPase6 (OS06G0181500), ATPase8 (OS02G0797300), which are associated with Complex V of oxidative phosphorylation and respiration also showed decreased transcript abundance in IR58025A (Table 2 and Fig. 6). Similar to our study, decreased transcript abundance of ATP synthase in male sterile line was reported in comparative transcriptomic analysis studies of *Brassica napus* (Yan et al. 2013), cotton (Yang et al. 2014), chilli pepper (Liu et al. 2013), welsh onion (Liu et al. 2016). The key genes involved in oxidative phosphorylation and encoded by mitochondrial and nuclear genomes were down-regulated in IR58025A leading to decrease in ATP pool, limiting high energy production required for male gamete production (Li et al. 2013).

The pathways associated with “metabolic process” and “biosynthesis of secondary metabolites” were significantly overrepresented in KEGG analysis of both down-regulated and up-regulated DETs in IR58025A (Fig. 5). KEGG is a popular public pathway-related database that integrates genomic, chemical, and systemic functional information and provides classifications valuable for studies on biologically and genetically complex processes (Kanehisa and Goto 2000). At pre-anthesis stage, the pathways associated with carbon and energy metabolisms were mostly enriched with down-regulated DETs in IR58025A. The main source of starch and sugars in plants is through process of photosynthesis. Not only the genes involved in photosynthesis viz., those encoding the photosystem-I and photosystem-II including *PsaG*, *Psbs*, *PsbO*, *Psb27*, but also those associated with photosynthesis antenna proteins and carbon fixation were down-regulated in IR58025A (Table 2) indicating a general shift toward conservation of resources and energy in the WA-CMS line. Reduction in photosynthesis efficiency is one of the key reasons responsible for limiting the starch accumulation in pollen grains. The insufficiency in starch synthesis in pollen grains is believed to be responsible for male sterility (Wen and Chase 1999; Datta et al. 2002; Wu

et al. 2016). Interestingly, the reduced photosynthetic efficiency was associated with ROS production and oxidative signaling in plants (Foyer 2018). However, their implication with the trait of CMS requires more in depth studies.

Interestingly, the pathway significantly enriched with up-regulated DETs in IR58025A as compared to its isogenic maintainer line was “biosynthesis of secondary metabolites” (Fig. 5). In plants, various stress responses trigger an increase in biosynthesis of secondary metabolites including flavonoids, terpenes, and phenylpropanoids (Dixon and Paiva 1995; Winkelshirley 2002; Tholl 2006). Similarly, in IR58025A, up-regulation of most of the genes involved in biosynthesis of secondary metabolites and phenylpropanoid biosynthesis was noticed. Phenylalanine is precursor for synthesis of flavonoids and lignin in plants. Also, five peroxidases involved in catalysis activity and synthesis of secondary metabolites (Os01g0962700, Os01g0327400-Class III peroxidase 16; Os05g0135500-Putative peroxidase; Os05g0499300-Peroxidase 1; Os04g0688100-Peroxidase; Os03g0152300-Class III peroxidase 36; Os02g0741200-Class III peroxidase 31) were up-regulated in IR58025A. Class III peroxidases have specific roles in cell wall lignification, scavenging of reactive oxygen species and their activity is remarkably increased under various stresses leading to ROS (Cosio and Dunand 2009). Various reported genes involved in scavenging of ROS-like rapid alkalization factors (RALFs), metallothionein (Zhang et al. 2010; Cherian and Kang 2006), and those involved in detoxification mechanism, heavy metal-associated domain-containing proteins (de Abreu-Neto et al. 2013), etc. showed decreased transcript abundance in IR58025A and this may be lead to production of uncontrolled ROS at pre-anthesis (Table 2). Even though, cytochrome P450 and glutathione S-transferases showed increased transcript abundance in IR58025A in RNA-Seq analysis (Table 2), the results of qRT-PCR analysis showed a lesser degree of their up-regulation in IR58025A (Fig. 6). Such slight deviations from RNA-Seq results have been reported in earlier studies (Wakasa et al. 2014; Wu et al. 2017). Interestingly, the results obtained in the qRT-PCR analysis of cytochrome P450 and glutathione S-transferases are similar to the results obtained in microarray analysis of set of rice alloplasmic sporophytic CMS lines as reported by Hu et al. (2016). Two reported stress-associated TFs, AP2-EREBP (Licausi et al. 2013), and NAC domain containing TFs, which mediate stress-associated response (Nuru-zaman et al. 2013) were observed to be up-regulated in IR58025A (Fig. 3). Interestingly, an abiotic stress response gene, OS01G0978500, known to encode a mutator-like transposon (Park et al. 2012), harbored a SNP (C/T) and *indel* (T/TA) in IR58025A compared to the reference *Japonica* genome (Supplementary Table 2) and the maintainer line IR58025B. Thus, based on the above findings, it may be inferred that at pre-anthesis in IR58025A, various stresses,

particularly oxidative stress and simultaneous depletion of energy metabolites may be responsible for WA352-induced male sterility of the WA-CMS line.

Genes associated with the development of fertile and mature anthers in rice plants, e.g., pectinesterases, calcium-binding proteins, etc. showed decreased transcript abundance in IR58025A, while normal expression of these genes was observed in IR58025B (Table 2). In rice, Ling et al. (2015) reported that activity of pectinesterases and calcium-binding proteins is essential for fertile pollen development. The same study also identified a mature anther-specific gene, *OsSTA99* that was regulated by activity of pectinesterases. The degree of pectin esterification in different pollen tube locations was controlled by competitive effect of pectin methylesterase (PME) and pectin methylesterase inhibitor (PMEI) (Bosch et al. 2005; Zhang et al. 2010). In the present study also, pectin methylesterase (OS05G0360900) showed significantly low expression level in IR58025A (Fig. 6) compared to IR58025B. Calcium ions and calcium-binding proteins are required for pollen grains germination and tube growth (Steinhorst and Kudla 2013). In rice, *OsSTA99RNAi* plants showed reduced activity of calcium-binding proteins (Ling et al. 2015). Interestingly, reduced activity of seven calcium-binding proteins, observed in the present study is also associated with the down-regulated pathway, “plant–pathogen interactions”, indicating that these events make the plant more vulnerable for different stresses. Thus, down regulation of DETs associated with pectinesterase activity and calcium-binding proteins activity at pre-anthesis may be a cause of pollen sterility and vulnerability of the WA-CMS to biotic stresses.

One of the predominant features, observed among most of the WA-CMS lines is the incomplete exertion of panicle, preventing access to about 20% of the spikelets and the failure of about 20% of spikelets to open at all (Yan and Li 1987; Tian 1991), thus limiting hybrid and parental line seed production. To make all the spikelets available for pollination and seed production, external application of gibberellin (GA₃) is generally adopted (Yuan 1981; Yuan and Fu 1995), increasing the cost and efforts involved in seed production. In view of possible association of genes responsible with gibberellin biosynthesis and/or response with the trait of incomplete panicle exertion in WA-CMS lines, we attempted to analyze up-regulation/down-regulation of such genes interestingly, three DETs (*OsGASR1*, *OsGSR1*, *CIGR2*) belonging to GA-stimulated transcripts gene family were up-regulated only in IR58025A. Ling et al. (2015) reported that *OsSTA* genes involved in rice anther wall development and anther dehiscence. and they are known to be regulated by the phytohormone gibberellin. In barley, the transgenic plants developed from aleurone cells showed higher expression of gibberellin responsive transcription factor, *HvGAMYB*, which resulted in male sterility due to a failure in dehiscence with the stomium remaining intact

(Murray et al. 2003). Further investigations are required to clarify the involvement of the above mentioned differentially expressed transcripts with the trait of male sterility and also with respect to panicle exertion.

Alternative splicing (AS) is a key regulatory mechanism that increases transcriptional complexity and protein diversity (Stamm et al. 2005). At pre-anthesis stage, a total of 1356 AS events were noticed in the two isogenic rice lines analyzed, with the higher number of AS events in IR58025A ($n=772$) compared to IR58025B ($n=651$). Among the five categories of AS events observed, skipped exon (SE-73%) was the predominant type followed by intron retention (RI-15.4%) (Supplementary Fig. 2). In plants, usually intron retention (RI) is the most predominant type of AS event and exon skipping (SE) was found to be the predominant type in humans (Wang and Brendel 2006). Few studies have reported that exon skipping is correlated with mitochondrial circular RNA and is also involved in mitochondria-mediated cell membrane-associated repair in humans (Kelly et al. 2015; Vila et al. 2017). With the data available from the present study, we are tempted to hypothesize that selective exon skipping observed in the male-sterile line could be a possible cause or consequence of WA352-induced male sterility. Further studies are in progress to conclusively validate the hypothesis on involvement of AS events in WA-CMS trait in rice.

Data

The transcriptome sequencing data from this study have been deposited in the NCBI SRA database and are accessible through accession number SRP133630 (<http://www.ncbi.nlm.nih.gov/sra>).

Conclusions

In this study, the differences of global transcriptome at the pre-anthesis stage of the popular WA-CMS line, IR58025A and its cognate isonuclear maintainer line, IR58025B were analyzed using high throughput RNA-Seq technique to gain critical insights into the gene expression patterns and molecular events associated with CMS trait. GO and KEGG enrichment analysis indicated significant association of down-regulated DETs in IR58025A with various metabolisms pathways, especially respiration and also cell wall modifications. Decrease in the expression of genes associated with ATP synthesis, increase in the expression of various genes associated with biotic and abiotic stress response, cell wall damage, and increased expression of genes associated with production of reactive oxygen species at pre-anthesis stage in the WA-CMS line, IR58025A can be possibly be a consequence of the action of the candidate gene for CMS trait, WA352, which ultimately leads to pollen sterility. The

data obtained in this study have enhanced our understanding on the molecular mechanisms involved in WA-CMS trait.

Acknowledgements The first author would like to thank Department of Science and Technology (DST), Government of India, for the INSPIRE (Innovation in Science Pursuit for Inspired Research) fellowship awarded for her PhD. studies. The authors would also like to thank the Indian Council of Agricultural Research and Department of Biotechnology, Government of India, for the generous funding support for the research work presented in the study, and past and present Directors of ICAR-Indian Institute of Rice Research, Hyderabad, India for providing research support for the study.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res* 18:1381–1392
- Bellincampi D, Cervone F, Lionetti V (2014) Plant cell wall dynamics and wall-related susceptibility in plant–pathogen interactions. *Front Plant Sci* 5:228
- Bentolila S, Stefanov S (2012) A reevaluation of rice mitochondrial evolution based on the complete sequence of male-fertile and male-sterile mitochondrial genomes. *Plant Physiol* 158:996–1017
- Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci USA* 99:10887–10892
- Bergman P, Edqvist J, Farbos I, Glimelius K (2000) Male sterile tobacco displays abnormal mitochondrial atp1 transcript accumulation and reduced floral ATP/ADP ratio. *Plant Mol Biol* 42:531–544. <https://doi.org/10.1023/A:1006388814458>
- Bosch M, Cheung AY, Hepler PK (2005) Pectin methylesterase, a regulator of pollen tube growth. *Plant Physiol* 138:1334–1346
- Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579–606
- Cherian MG, Kang YJ (2006) Metallothionein and liver cell regeneration. *Exp Biol Med* (Maywood) 231:138–144
- Cosio C, Dunand C (2009) Specific functions of individual class III peroxidase genes. *J Exp Bot* 6:391–408
- Das S, Sen S, Chakrabort A, Chakraborti P, Maiti MK, Basu A, Basu D, Sen KS (2010) An unedited 1.1 kb mitochondrial orfB gene transcript in the wild abortive cytoplasmic male sterility (WA-CMS) system of *Oryza sativa* L. subsp. Indica. *BMC Plant Biol* 10:39
- Datta R, Chamusco KC, Chourey PS (2002) Starch biosynthesis during pollen maturation is associated with altered patterns of gene expression in maize. *Plant Physiol* 130:1645–1656
- de Abreu-Neto JB, Turchetto-Zolet AC, de Oliveira LFV, BodaneseZanettini MH, Margis-Pinheiro M (2013) Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS J* 280:1604–1616
- Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell* 7:1085–1097

- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:64–70
- Foyer CH (2018) Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environ Exper Bot* 154:134–142
- Fujii S, Toriyama K (2008) Genome barriers between nuclei and mitochondria exemplified by cytoplasmic male sterility. *Plant Cell Physiol* 49:1484–1494
- Hari Prasad AS, Viraktamath BC & Mohapatra T (2014) Hybrid rice development in Asia: assessment of limitations and potential. In: *Proceedings of the regional expert consultation on hybrid rice development in Asia: assessment of limitations and potential*. Bangkok, Thailand. pp 85–100
- Horn R, Gupta KJ, Colombo N (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* 19:198–205
- Hu J, Chen G, Zhang H, Qian Q, Ding Y (2016) Comparative transcript profiling of alloplasmic male-sterile lines revealed altered gene expression related to pollen development in rice (*Oryza sativa* L.). *BMC Plant Biol* 16:175
- Huang JZ, Zhi-Guo E, Zhang HL, Shu QY (2014) Workable male sterility systems for hybrid rice: genetics, biochemistry, molecular biology, and utilization. *Rice* 7:13
- Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:27–30
- Kelly S, Greenman C, Cook PR, Papantonis A (2015) Exon skipping is correlated with exon circularization. *J Mol Biol* 427:2414–2417
- Khera P, Priyadarshi R, Singh A, Mohan R, Gangashetti MG Singh BN, Kole C, Shenoy V (2012) Molecular characterization of different cytoplasmic male sterile lines using mitochondrial DNA specific markers in rice. *J Biol Sci* 12:154–160
- Kruft V, Eubel H, Jansch L, Werhahn W, Braun H (2001) Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol* 127:1694–1710
- Kubo T, Newton KJ (2000) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8:5–14
- Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T (2011) Male sterility-inducing mitochondrial genomes: how do they differ? *Crit Rev Plant Sci* 30:378–400
- Lamb CJ, Dixon RA (1997) The oxidative burst in plant disease resistance. *Annu Rev Plant Physiol Plant Mol Biol* 76:419–422
- Li S, Wan C, Kong J, Zhang Z, Li Y, Zhu Y (2004) Programmed cell death during microgenesis in a Honglian CMS line of rice is correlated with oxidative stress in mitochondria. *Funct Plant Biol* 31:369–376
- Li J, Pandeya D, Jo YD, Liu WY, Kang BC (2013) Reduced activity of ATP synthase in mitochondria causes cytoplasmic male sterility in chili pepper. *Planta* 237:1097–1109
- Licausi F, Ohme-Takagi M, Perata P (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol* 199:639–649
- Ling S, Chen C, Wang Y, Sun X, Lu Z, Ouyang Y et al (2015) The mature anther-preferentially expressed genes are associated with pollen fertility, pollen germination and anther dehiscence in rice. *BMC Genomics* 16:101
- Liu C, Ma N, Wang P-Y, Fu N, Shen H-L (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(6):e65209. <https://doi.org/10.1371/journal.pone.0065209>
- 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
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ $\Delta\Delta C_T$ Method. *Methods* 25:402–408
- Luo DP, Xu H, Liu ZL, Guo JX, Li HY, Chen LT, Fang C, Zhang QY, Bai M, Yao N, Wu H Ji CH, Zheng HQ, Chen YL, Ye S, Li XY, Zhao XC, Li RQ, Liu YG (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat Genet* 45:573–577
- Murray F, Kalla R, Jacobsen J, Gubler F (2003) A role for HvGAMYB in anther development. *Plant J* 33:481–491
- Nuruzzaman M, Sharoni AM, Kikuchi S (2013) Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Front Microbiol* 4:248
- Paparella S, Tava A, Avato P, Biazzi E, Macovei A, Biggiogera M, Carbonera D, Balestrazzi A (2015) Cell wall integrity, genotoxic injury and PCD dynamics in alfalfa saponin-treated white poplar cells highlight a complex link between molecule structure and activity. *Phytochemistry* 111:114–123
- Park SH, Chung PJ, Juntawong P, Bailey-Serres J, Kim YS et al (2012) Posttranscriptional control of photosynthetic mRNA decay under stress conditions requires 3' and 5' untranslated regions and correlates with differential polysome association in rice. *Plant Physiol* 159(3):1111–1124
- Pogorelko G, Lionetti V, Bellincampi D, Zabolina O (2013) Cell wall integrity: targeted post-synthetic modifications to reveal its role in plant growth and defense against pathogens. *Plant Signal Behav* 8:e25435
- Pranathi K, Viraktamath BC, Neeraja N, Balachandran SM, Hari Prasad AS, Koteswara Rao P et al (2016) Comparative analysis of sequences of mitochondrial genomes of wild abortive male sterile (WA-CMS) and male fertile lines of rice, development of functional markers for WA-CMS trait and their use in assessment of genetic purity of seeds of WA-CMS lines. *Mol Breeding* 36:21
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M et al (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378
- Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. *Gene* 344:1–20
- Steinhorst L, Kudla J (2013) Calcium—a central regulator of pollen germination and tube growth. *Biochim Biophys Acta* 1833:1573–1581
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102:15545–15550
- Sundaram RM, Naveenkumar B, Biradar SK, Balachandran SM, Mishra B, Ilyasahmed M, Viraktamath BC, Ramesha MS, Sarma SP (2008) Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment. *Euphytica* 163:215–224
- Takakura Y, Ito T, Saito H, Inoue T, Komari T, Kuwata S (2000) Flower-predominant expression of a gene encoding a novel class I chitinase in rice (*Oryza sativa* L.). *Plant Mol Biol* 42:883–897
- Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr Opin Plant Biol* 9:297–304
- Tian DC (1991) Cultivation for out-crossing in rice-principles and techniques to achieve high yield in hybrid seed production. Sichuan Press of Sci & Tec Chengdu (Chinese)
- Tiwari BS, Belenghi B, Levine A (2002) Oxidative stress increased respiration and generation of reactive oxygen species, resulting in ATP depletion, opening of mitochondrial permeability transition,

- and programmed cell death. *Plant Physiol* 128:1271–1281. <https://doi.org/10.1104/PP.010999>
- Trapnell C, Williams AB, Pertea G, Mortazavi A, Kwan G, Baren M, Salzberg S, Wold B, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515
- Vila MC, Rayavarapu S, Hogarth MW, Van der Meulen JH et al (2017) Mitochondria mediate cell membrane repair and contribute to Duchenne muscular dystrophy. *Cell Death Diff* 24(2):330–342
- Wakasa Y, Oono Y, Yazawa T, Hayashi S, Ozawa K, Handa H et al (2014) RNA sequencing-mediated transcriptome analysis of rice plants in endoplasmic reticulum stress conditions. *BMC Plant Biol* 14:101
- Wang BB, Brendel V (2006) Genome wide comparative analysis of alternative splicing in plants. *PNAS* 103:7175–7180
- Wang Z, Zou Y, Li X, Zhang Q, Chen L et al (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 18:676–687
- Wang K, Gao F, Ji Y, Liu Y, Dan Z et al (2013) ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. *New Phytol* 198:408–418
- Wen LY, Chase CD (1999) Mitochondrial gene expression in developing male gametophytes of male-fertile and S male-sterile maize. *Sex Plant Reprod* 11:323–330
- Whitford R, Fleury D, Reif JC, Garcia M, Okada T, Korzun V, Langridge P (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J Exp Bot* 64:5411–5428
- Winkelshirley B (2002) Biosynthesis of flavonoids and effect of stress. *Curr Opin Plant Biol* 5:218–223
- Wu Y, Fox TW, Trimmell MR, Wang L, Xu R, Cigan AM, Huffman GA, Garnaat CW, Hershey H, Albertsen MC (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. *Plant Biotechnol J* 14:1046–1054
- 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
- Yan WG, Li SF (1987) Study on out-crossing characteristics among male sterile lines containing same nucleus in rice. *Hybrid Rice* 4:8–11 (Chinese)
- 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 P, Han J, Huang J (2014) Transcriptome sequencing and de novo analysis of cytoplasmic male sterility and maintenance in JA-CMS cotton. *PLoS ONE* 9(11):e112320. <https://doi.org/10.1371/journal.pone.0112320>
- Young M, Wake eld M, Smyth G, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:14
- Yuan LP (1981) Hybrid rice. United States patent and trademark office. 4305225
- Yuan LP, Fu XQ (1995) Technology of hybrid rice production. Food and Agriculture Organization of the United Nations, Rome
- Zhang GY, Feng J, Wu J, Wang XW (2010) BoPMEI1, a pollen-specific pectin methylesterase inhibitor, has an essential role in pollen tube growth. *Planta* 231:1323–1334
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M et al (2016) iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant* 9:1667–1670