

# SCIENTIFIC REPORTS

OPEN

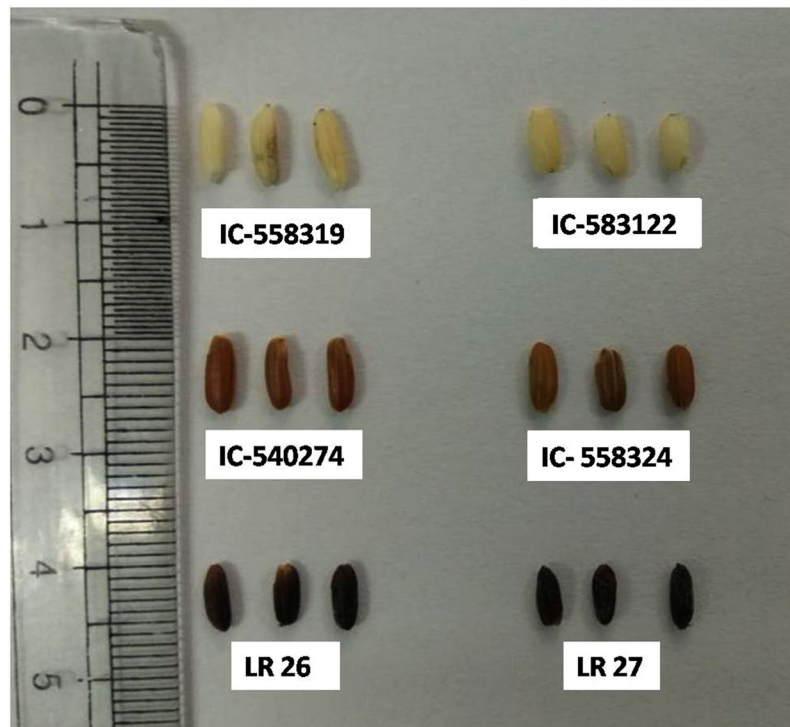
## Comparative transcriptomics approach in elucidation of carotenoid biosynthesis regulation in grains of rice (*Oryza sativa* L.)

Upasna Chetty<sup>1</sup>, Nikhil K. Chrungoo<sup>1</sup> & Kirti Kulkarni<sup>2</sup>

Estimation of phytoene, lycopene,  $\beta$ -carotene, lutein, and zeaxanthin in grains of white, brown and purple cultivars of rice revealed marked differences in the levels of these carotenoid intermediates amongst the cultivars. Grains of white rice did not show any significant accumulation of carotenoid intermediates at any stage of development. On the other hand, grains of the purple cultivar accumulated  $49.16 \pm 5 \mu\text{g}$  of  $\beta$ -carotene,  $28.89 \pm 3.2 \mu\text{g}$  of lutein and  $34.65 \pm 4.6 \mu\text{g}$  of zeaxanthin per gm of grain fresh weight. In addition to PSY1, higher expression of  $\beta$ LCY than  $\epsilon$ LCY appears to be an important factor in determining the flux of pathway towards synthesis of  $\beta$ - $\beta$  branch carotenoids in purple rice. This cultivar showed a higher fold change in carotenoid precursors during transition from milky to doughing stages and an enhanced flux of lycopene towards  $\beta$ -carotene during grain maturation. Our results indicate that higher level of carotenoids in purple rice is a consequence of higher expression of genes involved in pyruvate metabolism as well as those involved in carotenoid biosynthesis such as PSY1, PDS and  $\beta$ -LCY. Co-expression networking revealed a strong positive relationship between the expression profiles of genes involved in carotenoid biosynthesis and genes coding for geranylgeranyl transferase type II, glutathione S-transferase, DnaJ and SET domain containing proteins as well as MADS26 and R2R3MYB family of transcription factors.

As a staple food of bulk of the population of Southeast Asia, rice grains contribute significantly towards meeting nutritional requirements of the people of this region. However, low levels of calcium, iron, thiamine and riboflavin and near absence of  $\beta$ -carotene makes rice as one of the low nutrient staple foods. Thus, exclusive reliance on rice as a primary food leads to vitamin A deficiency, a serious public health problem. Strategies to boost the content of  $\beta$ -carotene, the precursor for vitamin A, in endosperm of rice grains have focused around expression of heterologous PSY and the bacterial CRTI, which replaced several consecutive reactions that are required for conversion of phytoene to lycopene<sup>1</sup>. Bai *et al.*<sup>2</sup> and Gayen *et al.*<sup>3</sup> have, however, identified pyruvate flux, isoprenoid precursor pool and metabolic sink as additional bottlenecks in carotenoid biosynthesis and accumulation in rice. With a limited knowledge of the regulatory controls involved in carotenoid biosynthesis in rice, availability of mutants with altered carotenoid biosynthetic pathway would have been an important milestone in understanding carotenoid biosynthesis in this crop. While such a collection of mapped and biochemically characterized mutants is available in other grain crops such as maize<sup>4,5</sup>, not many mutants of this type have been identified in rice. Wurtzel *et al.*<sup>6</sup> presented the first evidence of a vivipary mutation 84NMEMdr2 and 90KHEMdr1 in rice which also caused albino phenotype. Besides vivipary, the mutation also conferred a block in the conversion of phytoene to lycopene, thereby hindering biosynthesis of downstream intermediates of the carotenoid pathway in grains. Low expression of phytoene desaturase or impaired biosynthesis of plastoquinones was suggested as the main reason for block in the conversion of phytoene to lycopene. Fang *et al.*<sup>7</sup> have identified 4 *phs* mutants in rice which had mutations in PDS, ZDS CRITSO and  $\beta$ LCY genes. The mutants showed viviparous phenotype with altered carotenoid composition as well as impaired plastid assembly. While the mutations confirmed the role of PDS, ZDS CRITSO and  $\beta$ LCY in carotenoid biosynthesis, they did not provide any information on the regulatory controls involved in carotenoid biosynthesis in rice. Pigmented cultivars, which are rich in bioactive compounds including anthocyanins and carotenoids<sup>8-14</sup>, provide an alternative for analyzing the regulatory factors involved

<sup>1</sup>North Eastern Hill University, Department of Botany, Shillong, 793022, India. <sup>2</sup>Bionivid Technologies, Bangalore, India. Correspondence and requests for materials should be addressed to N.K.C. (email: [nchrungoo@nehu.ac.in](mailto:nchrungoo@nehu.ac.in))



**Figure 1.** Grains of different accessions of rice from North East India studied in the present investigation showing variations in the colour of their pericarp. White(IC-558319, IC-583122), Brown(IC-540274, IC-558324), Purple(LR 26, LR 27).

in carotenoid biosynthesis in rice. To address the challenge of understanding the regulators of carotenoid biosynthesis in rice, we used rice grains from Northeast India varying in the pericarp color (white, brown and purple) as contrasting phenotypes for deep transcriptome sequencing at different developmental stages.

## Material and Methods

**Plant Material.** Grains of white (IC-583122, IC-558319), brown (IC-540274, IC-558324) and purple (LR 26, LR 27) cultivars of rice (*Oryza sativa* L.) were procured from National Bureau of Plant Genetic Resources (NBPGR) and Central Agricultural University, Barapani, Shillong (Fig. 1). Seedlings of each accession were raised to maturity in an environment controlled green house under a 12 hour photoperiod and temperature of  $26 \pm 2^\circ\text{C}$  during day and  $22 \pm 2^\circ\text{C}$  during night. The grains were harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stage of development and dehulled. The dehulled grains were cleaned and stored at  $-80^\circ\text{C}$ .

**Targeted estimation of carotenoid intermediates.** Estimation of phytoene, lycopene,  $\beta$ -carotene, lutein, and zeaxanthin was carried out in whole grains of the six cultivars of rice harvested at milky (15DAP), doughing (25DAP) and mature (40DAP) stages of grain development according to Panfili *et al.*<sup>15</sup> with certain modifications. A suitable mass of the harvested tissue from each cultivar was crushed to fine powder under liquid nitrogen in a pre-chilled pestle and mortar. A defined mass of the powdered tissue was suspended in three volumes of 95% ethanol containing 0.1% ascorbic acid.  $1/3^{\text{rd}}$  volume of cold 1% sodium chloride was added to the mixture and the entire solution was vortexed and incubated in a water bath at  $85^\circ\text{C}$  for 10 minutes. Saponification was carried out by adding  $1/3$  volume of 80% potassium hydroxide and heating in a water bath at  $85^\circ\text{C}$  for 20 minutes. The suspension was mixed with equal volume of n-hexane-ethyl acetate (9:1) and the organic phase was transferred to a fresh tube. The aqueous phase was mixed again with equal volume of n-hexane-ethyl acetate (9:1) and the organic phase collected. The organic phases were pooled, dried under nitrogen gas at  $40^\circ\text{C}$  and re-dissolved in  $500\mu\text{l}$  of methanol:acetonitrile (40:60) for chromatography on Spherisorb ODS2 C18 reverse phase column ( $25 \times 3.2\text{ mm}$ ) at a flow rate of  $0.45\text{ ml min}^{-1}$  using methanol:acetonitrile:2-propanol (40:58:2) as the eluent. Each sample was filtered through a  $0.4\mu\text{m}$  nylon filter (SFNY25X) (Axiva) prior to injection into the column. The extinction values for each peak were monitored at 450 nm with phytoene, lycopene,  $\beta$ -carotene, lutein, and zeaxanthin as the reference standards. The concentration of a carotenoid in a fraction was calculated from the peak areas with reference to the peak area determined for the eluted reference standard. All analyses were carried out with eight biological and three technical replicates. The data was statistically tested for significance of variations by analysis of variance (ANOVA) at  $P 0.05$ .

**Transcriptome library preparation and sequencing.** Total RNA was extracted from grains of white (IC-583122), brown (IC-540274) and purple (LR 26) cultivars of rice harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stages of grain development, using Trizol reagent (Invitrogen) with certain

modification. The quality of RNA was assessed with Agilent 2100 Bioanalyzer. RNA samples with RIN value  $\geq 7.0$  were subjected to deep transcriptome sequencing using Illumina HiSeq2000 sequencing platform as per manufacturer's protocol.

***In-silico analysis of transcriptome data.*** *Quality Control Analysis and Alignment.* The raw reads were subjected to quality check using “NGS-QC tool kit v2.3”<sup>16</sup>. Reads having 70% of the bases with a “Phred” score of  $>20$  were considered as high quality and used for further analysis. Sample-wise high-quality reads were aligned to the rice reference genome (<http://rice.plantbiology.msu.edu/>) using “TopHat v2.0.3”. The aligned reads were subsequently used to create a reference annotation based transcript (RABT) assembly to identify uniquely aligned reads in each sample.

*Quantification and differential expression of transcripts.* We used “Cufflinks v2.2.0” to identify the differentially expressed transcripts across the cultivars at each stage of grain development and within cultivars for different stages of grain development. The assembled samples were combined using “Cuffmerge v2.2.0” and subjected to “Cuffcompare v2.2.0” for comparison with transcripts available in msu 7 database. The results obtained after data analysis by Cuffcompare were subjected to “featureCounts v1.5.0” to obtain the read count of all samples. Differences in gene expression across different stages of grain development within cultivars and at each stage of grain development across the cultivars were identified using “DESeq2 1.2.5R”. The thresholds for P-value and FDR (adjusted P-value or Q-value) were set to 0.001 for identifying significant differences in expression. Gene Ontology (GO) was derived from the msu 7 rice genome database and the pathways were identified using Kyoto Encyclopedia of Genes and Genomes (KEGG). The GO analysis of novel transcripts and highly expressed genes was plotted using the WEGO (Web Gene Ontology Annotation Plot) program (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>)<sup>17</sup>. Transcription factors were identified from the data sets using the annotation from Database of Rice Transcription Factors and Plant Transcription Factor Database.

*Pathway analysis and BAN (Biological Analysis Network) modelling.* The data obtained on Gene Ontologies, Pathways and DEGs was used as input for pathway analysis and biological analysis network modeling using “BridgeIsland Software” (Bionivid Technology, India). The output file from Bridge Island was fed to “CytoScape V 2.8” to derive molecular interaction networks.

*Co-expression networking.* For a comparative analysis of transcription factors co-expressing with the carotenoid biosynthesis pathway across the white, brown and purple cultivars of rice, a correlation analysis of genes involved in carotenoid biosynthesis and transcription factors was carried out using “CORREL” function in “EXCEL2013” following Zhang *et al.*<sup>18</sup> and Ye *et al.*<sup>19</sup>. In order to exclude false positives, structural genes and transcription factors with FPKM value of  $\geq 5.0$  in at least one of the three stages of grain filling were selected. Transcription factors with correlation coefficient values of  $\geq 0.8$  were considered to have an expression that was significantly correlated with the expression of carotenoid biosynthetic pathway. All data has been expressed as mean of independent replicates  $\pm$ SD at P0.05. The statistical significance of variations has been calculated using student's *t*-test.

**Validation of gene expression.** Validation of the expression profiles of PSY1, PDS,  $\beta$ -LCY,  $\epsilon$ -LCY, BCH, hexokinase, pyruvate kinase, fructose-1,6bisphosphatase, geranylgeranyl transferase, 9- cis-epoxycarotenoid dioxygenase, MADS26 and MYB genes was carried out by real-time PCR using gene-specific primers listed in Table s1. Each reaction was carried out with three independent biological replicates and three technical replicates. The relative expression levels for each gene were normalized to the expression level of actin (internal reference gene), which was calculated from cycle threshold values using the  $2^{-Ct}$  method according to Livak and Schmittgen<sup>20</sup>.

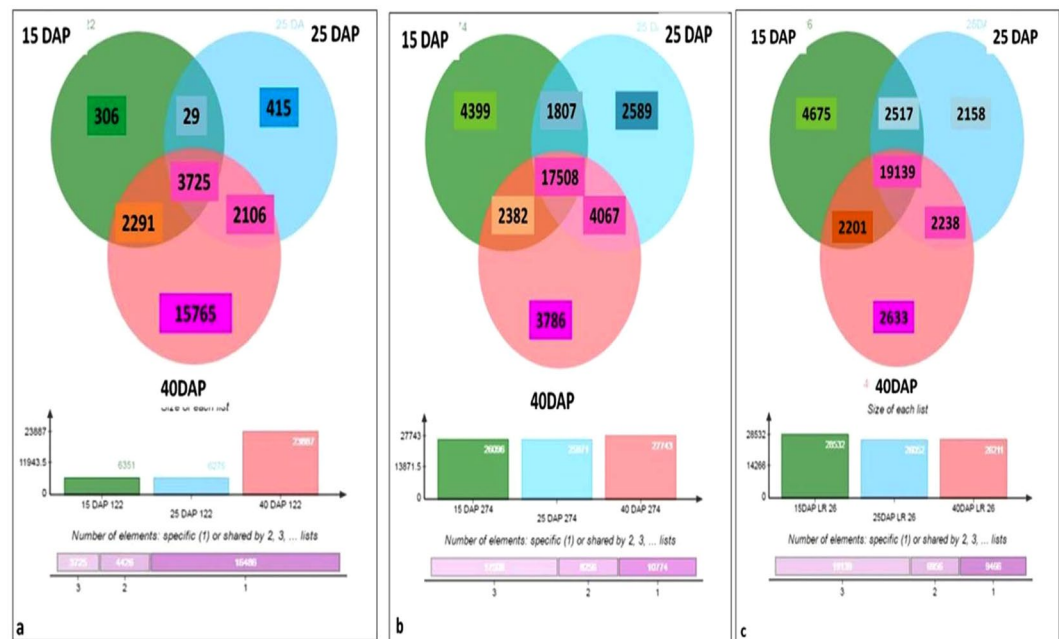
## Results

**Targeted estimation of carotenoid intermediates.** Estimation of phytoene, lycopene,  $\beta$ -carotene, lutein and zeaxanthin in whole grains of white (IC-583122, IC-558319), brown (IC-540274, IC-558324) and purple (LR 26, LR 27) cultivars of rice harvested at milky (15DAP), doughing (25DAP) and mature (40DAP) stages of grain development revealed significant differences in the levels of these carotenoid intermediates amongst the cultivars. Except for marginal levels of lutein in mature grains of IC-583122, no other carotenoid intermediate was detected in grains of white cultivars at any stage of development. While grains of brown cultivars showed the presence of phytoene, lycopene, lutein,  $\beta$ -carotene and zeaxanthin at milky stage, the content of each of these metabolites showed marked decrease with progressing grain development. Grains of the brown cultivars accumulated more lutein than  $\beta$ -carotene and zeaxanthin at all the three stages of development. In comparison to white and brown cultivars, grains of purple cultivars accumulated much higher levels of phytoene, lycopene,  $\beta$ -carotene, lutein, and zeaxanthin. The content of each of these carotenoid intermediates showed a progressive increase with grain development till maturity (Table 1, Fig. s1). In contrast to brown cultivars, grains of purple cultivars showed higher level of  $\beta$ -carotene and zeaxanthin than lutein.

**Whole transcriptome sequencing.** After removal of low quality and primer/adaptor contaminated reads, we obtained an average of 103,676,662 million high-quality reads from the raw RNAseq data. Mapping of the fragments to msu7 genome with “Tophat2 version 2.0.11” revealed an overall alignment of 83.54% (Table s2). Out of the total number of transcripts identified, 42,427 transcripts were found to be expressed with a sensitivity index of ReadCount  $\geq 0$ . Sequence analysis revealed a total of 31,640 known protein-coding genes apart from

Accessions	Pericarp color	Days After Pollination	Phytoene	Lycopene	Lutein	$\beta$ -carotene	Zeaxanthine
IC-583122	White	15DAP	nd	nd	nd	nd	nd
		25DAP	nd	nd	nd	nd	nd
		40 DAP	nd	nd	0.06 $\pm$ 0.02	nd	nd
IC-558319	White	15DAP	nd	nd	nd	nd	nd
		25DAP	nd	nd	nd	nd	nd
		40 DAP	nd	nd	nd	nd	nd
IC-540274	Brown	15DAP	5.43 $\pm$ 0.97	9.32 $\pm$ 2.40	14.00 $\pm$ 1.20	10.00 $\pm$ 1.00	14.00 $\pm$ 0.50
		25DAP	4.20 $\pm$ 0.12	6.63 $\pm$ 2.20	9.44 $\pm$ 2.00	5.60 $\pm$ 1.40	4.66 $\pm$ 0.30
		40 DAP	3.50 $\pm$ 0.10	3.4 $\pm$ 0.40	6.36 $\pm$ 0.80	2.01 $\pm$ 1.60	2.10 $\pm$ 0.90
IC-558324	Brown	15DAP	2.43 $\pm$ 0.60	7.66 $\pm$ 0.30	8.00 $\pm$ 1.30	7.00 $\pm$ 1.40	9.00 $\pm$ 2.00
		25DAP	1.26 $\pm$ 0.40	4.11 $\pm$ 0.40	4.4 $\pm$ 2.00	4.60 $\pm$ 1.20	7.43 $\pm$ 3.20
		40 DAP	0.66 $\pm$ 0.20	2.01 $\pm$ 0.20	3.10 $\pm$ 0.60	0.98 $\pm$ 0.30	1.64 $\pm$ 0.30
LR 26	Purple	15DAP	10.43 $\pm$ 0.06	14.66 $\pm$ 0.30	16.00 $\pm$ 3.00	29.97 $\pm$ 3.00	18.22 $\pm$ 2.00
		25DAP	15.43 $\pm$ 0.50	16.061 $\pm$ 1.00	23.42 $\pm$ 4.00	34.12 $\pm$ 6.00	20.43 $\pm$ 0.43
		40 DAP	18.36 $\pm$ 3.00	17.32 $\pm$ 4.00	28.89 $\pm$ 3.20	49.16 $\pm$ 5.00	34.65 $\pm$ 4.60
LR 27	Purple	15DAP	10.83 $\pm$ 1.00	8.43 $\pm$ 0.50	16.43 $\pm$ 9.00	10.00 $\pm$ 0.80	14.22 $\pm$ 0.80
		25DAP	13.43 $\pm$ 1.08	9.66 $\pm$ 3.20	24.32 $\pm$ 4.00	13.42 $\pm$ 0.50	15.43 $\pm$ 1.00
		40 DAP	14.86 $\pm$ 2.20	10.32 $\pm$ 2.30	27.16 $\pm$ 3.00	28.69 $\pm$ 4.00	15.65 $\pm$ 2.00

**Table 1.** Content of phytoene, lycopene, lutein,  $\beta$  carotene and zeaxanthin ( $\mu\text{g gm}^{-1}$  fresh weight) in the whole grain of six different cultivars/accessions representing white, brown and purple accessions/cultivars of rice from North-East India.



**Figure 2.** Venn diagram showing the genes expressed in whole grains of white (IC-583122), brown (IC-540274) and purple (LR 26) cultivars of rice from North-East India harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stages of development. (a) genes expressed in IC-583122 at 15 DAP, 25 DAP and 40 DAP. (b) genes expressed in IC-540274 at 15 DAP, 25 DAP and 40 DAP. (c) genes expressed in LR 26 at 15 DAP, 25 DAP and 40 DAP.

lincRNAs, pseudogenes and other types of long RNAs. Transcript type distribution revealed 74.0% sequences as already known transcripts, 22.0% as modified known transcripts and 2.7% as novel intergenic transcripts.

**Transcript distribution at different developmental stages.** Transcript distribution at different stages of grain development in each of the three cultivars of rice is given in Fig. 2. Grains of IC-583122 showed co-expression of 3,725 transcripts at all the three stages of grain development. Transcript distribution analysis revealed 29 co-expressed transcripts at milky and doughing stages. On the other hand, 2,106 transcripts showed

co-expression at doughing and mature stages and 2,291 transcripts showed co-expression at milky and mature stages of grain filling. The number of developmental stage-specific genes expressed at milky, doughing and mature stages was 306, 415 and 1,5765, respectively (Fig. 2a). On the other hand, grains of IC-540274 showed co-expression of 17,508 genes at all the three stages of grain development. Data analysis revealed co-expression of 1,807 genes at milky and doughing, 4,067 genes at doughing and mature and 2,382 genes at milky and mature stages of grain development. The number of developmental stage-specific genes expressed at milky, doughing and mature stages was 4,399, 2,599 and 3,786, respectively (Fig. 2b). Transcript distribution analysis revealed co-expression of 19,139 genes at milky, doughing and mature stages in the purple cultivar, LR 26. Whereas 2,517 genes showed co-expression at milky and doughing stages, 2201 genes showed co-expression at doughing and mature stages. 2,238 genes showed co-expression at milky and mature stages in this cultivar. While 4,675 transcripts were specific to milky stage, 2,158 and 2,633 genes were found to be expressed specifically at doughing and mature stages, respectively (Fig. 2c).

Out of the total of 1,12,495 genes designated under different core categories, 55,659 genes were functionally annotated to biological process, 27,490 genes were annotated to cellular processes and 29,346 genes were annotated to molecular functions. 169 genes showed unique ontology.

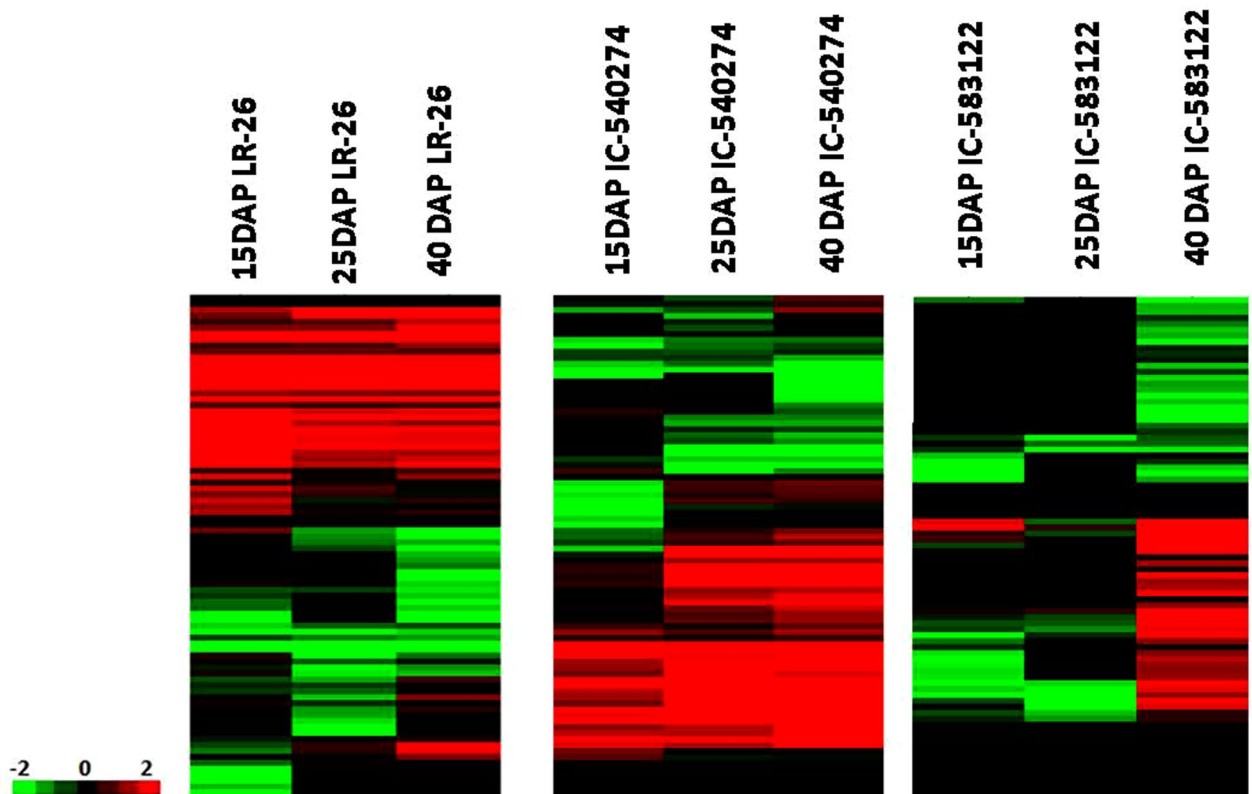
Within the biological process, the top 5 GOs were cellular process (13.4803%), metabolic process (13.06%), biosynthetic process (8.73%), nucleic acid metabolism (6.76%) and stress responses (6.09%). In the cellular processes category, the top 5 GOs included membrane (14.06%), plastid (12.55%), plasma membrane (11.61%), nucleus (9.92%) and cytosol (8.65%).

The top 5 GOs identified under the category of molecular functions included binding (13.01%), catalytic activity (12.52%), protein binding (12.26%), hydrolase activity (9.83%) and nucleotide binding (8.83%) (Table s3). Mapping of genes to Kyoto Encyclopedia of Genes and Genome database (<http://www.genome.ad.jp/kegg/>) lead to assignment of 1,180 genes to 121 KEGG pathways. Amongst these, the top 5 pathways with greatest representation of genes included metabolic pathways (18.13%), biosynthesis of secondary metabolites (10%), plant hormone signal transduction (2.9%), carbon metabolism (2.6%) and ribosomes (2.6%) (Table s3). We also carried out GO analysis of novel transcripts and highly expressed genes using the WEGO (Table s4). Gene annotation with Database of Rice Transcription Factors as well as Plant Transcription Factors identified 1,284 TFs belonging to 56 families in the transcriptome. These included NAC (8.48%), ERF (8.41%), bHLH (7.94%), MYB (8.9%), WRKY (6.15%), C2H2 (6.07%) and bZIP (5.37%) families (Table s3).

**Differentially expressed genes.** We used DESeq to estimate differential gene expression as a function of grain development in white (IC-583122), brown (IC-540274) and purple (LR 26) cultivars of rice harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stages of grain development. Grains of white rice showed 531 DEGs between milky and doughing stages. Out of these 284 genes showed upregulation and 247 genes showed down regulation. On the other hand, the number of DEGs between doughing and mature stages was 14,350. Out of these 6,864 genes were upregulated and 7,502 were down regulated. Grains of the brown cultivar showed 2,236 DEGs between milky and doughing stages. Out of these 1,180 genes were upregulated and 1,056 were down regulated. On the other hand, the number of DEGs between doughing and mature stages was 9,427, out of which 4,607 were upregulated and 4,820 were down regulated. The purple cultivar showed differential expression of 2,096 genes between milky and doughing stages of grain development. Out of these, 1,097 genes were upregulated and 999 genes were down regulated. Grains of the purple cultivar showed 8,690 DEGs between doughing and mature stages. Out of these 4,519 genes showed upregulation and 4,171 showed down regulation (Fig. 3). We could observe significant differences in the expression profiles of genes across the cultivars for each developmental stage. While 1,194 transcripts showed differential expression at milky stage between white and purple cultivar, 846 transcripts showed differential expression between the white and brown cultivar for the same stage of development. While the number of differentially expressed transcripts at doughing stage between the white and purple cultivars was 681, the number of differentially expressed transcripts at doughing between white and brown cultivars was 793. The number of differentially expressed transcripts detected at maturity between the white and purple cultivar and between the white and brown cultivar was 147 and 1,437, respectively (Fig. 4).

**Expression profiling of the genes involved in carotenoid biosynthesis.** RNASeq derived data revealed marked differences in the expression profiles of phytoene synthase1 (LOC\_Os12g43130), phytoene desaturase (LOC\_Os02g09780), lycopene  $\beta$ -cyclase (LOC\_Os02g09750) and lycopene  $\epsilon$ -cyclase (LOC\_Os01g39960) between the white, brown and purple cultivars at each stage of grain development (Table s5). In comparison to the white and brown cultivars, grains of the purple cultivar (LR 26) showed higher expression of PSY1, PDS and  $\beta$ -LCY at all the stages of development. Grains of white cultivar (IC-583122) showed no significant changes in the expression of PSY1 and PDS during transition from milky to doughing stage, however, both the genes showed a nearly 2 fold increase in expression between doughing and mature stages (Table s5). While the grains of the white cultivar did not show any expression of  $\beta$ -LCY at any stage of development, expression of  $\epsilon$ -LCY was detected only at maturity in this accession. The expression of  $\epsilon$ -LCY showed a near 2.5 fold increase during the transition from doughing to mature stage of grain development. The grains of brown cultivar (IC-540274) showed a progressive decrease in expression of PSY1 and PDS during transition from milky stage to maturity. On the other hand, grains of the purple cultivar (LR 26) showed a significant increase in expression of both these genes with progressing grain development. While the grains of brown cultivar showed higher expression of  $\epsilon$ -LCY than  $\beta$ -LCY at all the stages of development, those of purple cultivar showed higher expression of  $\beta$ -LCY than  $\epsilon$ -LCY at all stages of development (Table s5, Fig. 5).

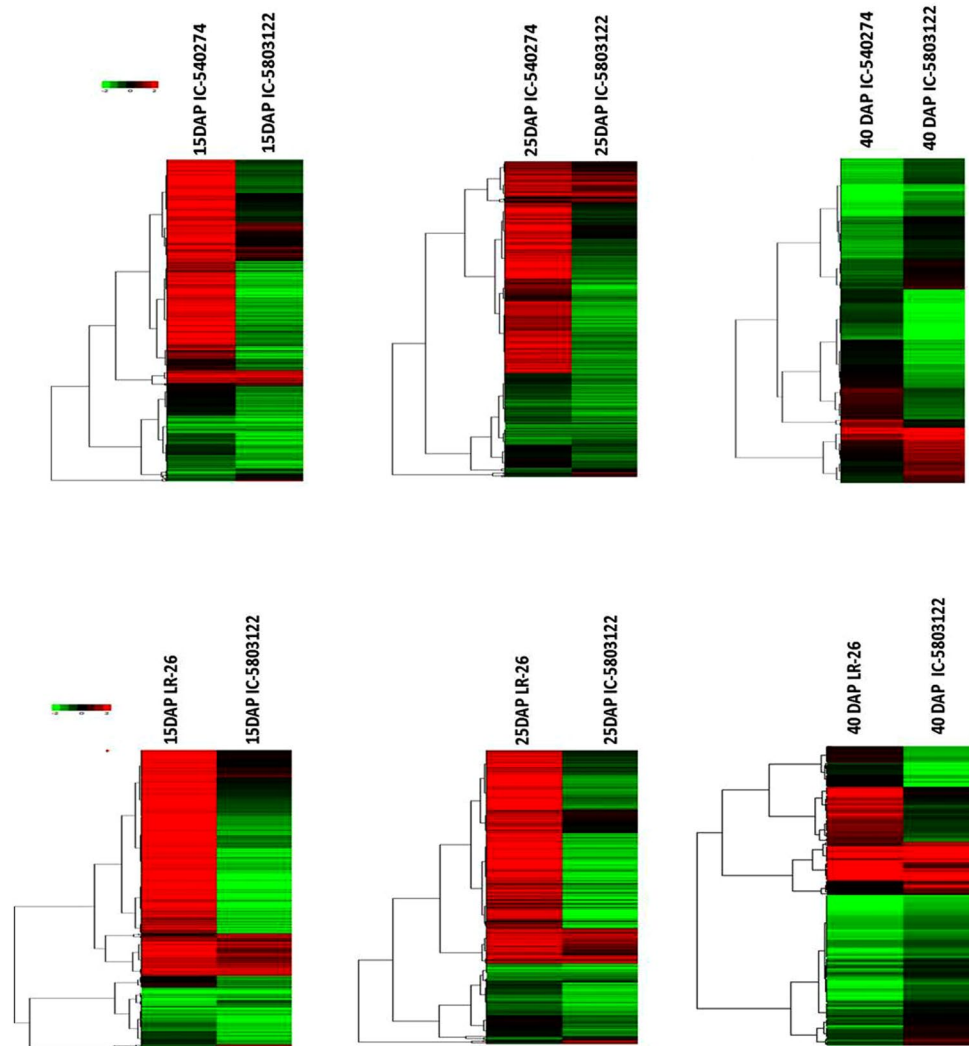
**Co-expression networking amongst pathways/processes.** We developed a nested network between the differentially expressed genes involved in phenyl propanoid biosynthesis, pyruvate metabolism, fatty acid



**Figure 3.** Comparative hierarchical cluster analysis of the differentially expressed genes from whole grains of white (IC-583122), brown (IC-540274) and purple (LR 26) cultivars of rice from North-East India harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stages of grain filling. Color bars show the corresponding scale for log<sub>2</sub> fold change in expression.

metabolism, carotenoid biosynthesis and plant hormone signal transduction using “CytoScape tool”. The network revealed a high degree of interaction between pyruvate metabolism, lipid metabolism, secondary metabolism, cell differentiation, inositol phosphate metabolism, embryo development, phenyl propanoid biosynthesis, isoprenoid biosynthesis, abiotic stress and carotenoid biosynthesis (Fig. 6a). Coexpression networking revealed a strong interaction between pyruvate metabolism, lipid metabolism, phenyl propanoid biosynthesis, isoprenoid biosynthesis, abiotic stress and carotenoid biosynthesis (Fig. 6b). While the genes coding for enzymes involved in pyruvate biosynthesis such as hexokinase (LOC\_Os05g45590.1), fructose1,6 -bisphosphatase aldolase (LOC\_Os01g67860.1), 6-phosphofructokinase2 (LOC\_Os04g39420.1), glyceraldehyde-3-phosphate dehydrogenase (LOC\_Os06g45590.1) and pyruvate kinase (LOC\_Os03g46910.1) did not show any significant expression in the grains of white cultivar at the milky and doughing stages, there was an increase in the expression of these genes in grains of this cultivar during transition from doughing stage to maturity (Table S6a). On the other hand grains of the brown cultivar showed a relatively higher expression of these genes at milky stage of development. The expression of each of these gene, however, showed a progressive decrease during transition of the grains from milky stage to maturity (Table s6a). Compared to the white and brown cultivars, grains of purple cultivar (LR 26) showed marked increases in the expression of hexokinase (LOC\_Os05g45590.1), fructose1,6 -bisphosphatase aldolase (LOC\_Os01g67860.1), 6-phosphofructokinase2 (LOC\_Os04g39420.1), glyceraldehyde-3-phosphate dehydrogenase (LOC\_Os06g45590.1) and pyruvate kinase (LOC\_Os03g46910.1) with progressing grain development from milky stage to maturity (Table s6a). Grains of the purple cultivar also showed higher expression of 9-cis-epoxycarotenoid dioxygenase (LOC\_Os02g47510, LOC\_Os12g42280) (Table s6a). Co-expression networking also revealed a strong positive relationship between the expression profiles of genes involved in carotenoid biosynthesis and genes coding for geranylgeranyl transferase type II, glutathione S-transferase, DnaJ and SET domain containing proteins (Table s6b). Compared to the brown and white cultivars, grains of the purple cultivar showed significantly higher expression of genes coding for geranylgeranyl transferase type II (LOC\_Os11g32090), glutathione S-transferase (LOC\_Os10g38489), chaperone DnaJ10 (LOC\_Os08g41110), DnaJ domain-containing protein (LOC\_Os05g48810) and SET domain-containing protein (LOC\_Os08g14660). The fold change in expression of these genes correlated strongly with changes in the level of carotenoid intermediates, especially during grain maturation (Table s6b).

Transcript abundance covariance, carried out between selected genes of the carotenoid biosynthesis pathway and transcription factors revealed the involvement of nineteen families of transcription factors with genes involved in carotenoid biosynthetic pathway (Table s7). Whereas light specific TFs belonging to WRKY, MYB, bHLH and zinc finger families showed higher expression at all stages of grain development in the purple and



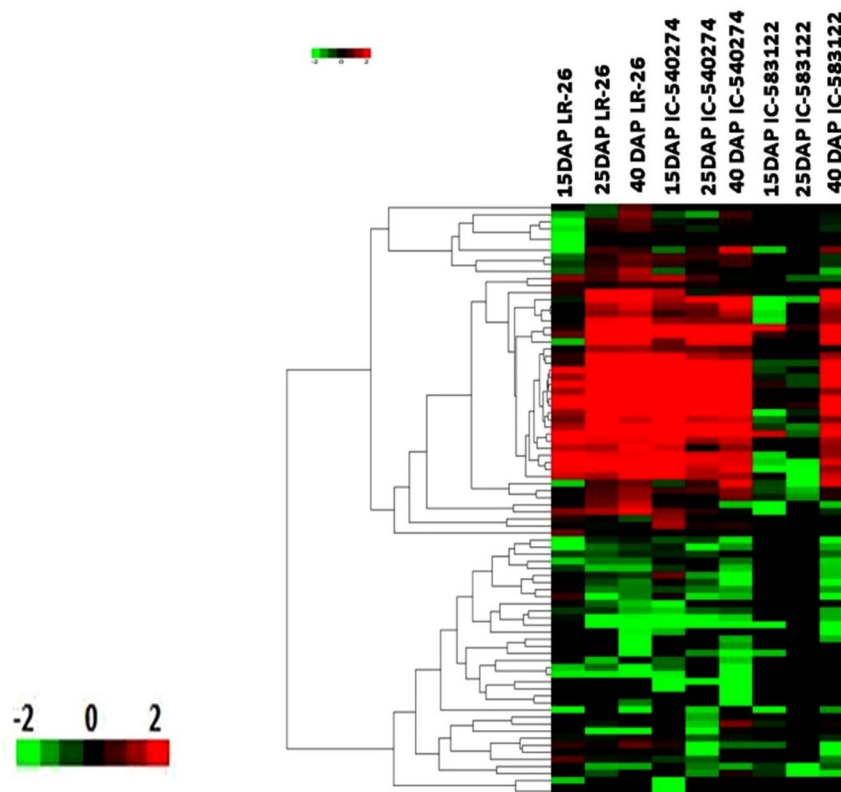
**Figure 4.** Comparative hierarchical cluster analysis between the white and pigmented cultivars at three developmental stages. (a) Pairwise comparison at 15DAP, 25DAP and 40DAP between IC-583122 and LR 26 (b) Pairwise comparison at 15DAP, 25DAP and 40DAP between IC-583122 and IC-540274. Color bars show the corresponding scale for log<sub>2</sub> fold change in expression.

brown cultivars, their expression was registered only at maturity in the grains of white cultivar (Fig. 7). On the other hand, AP-2 and AP2/ERF, which are negative regulators of carotenoid metabolism, showed either no expression or down regulation of expression in grains of purple and brown cultivar at all stages of development. The expression of these TFs however, showed a nearly 3 fold increase during grain maturation in white rice (Fig. 7). Data analysis also revealed a strong correlation between MADS26 (LOC\_Os08g02070), and R2R3MYB (LOC\_Os06g10350) transcription factors with expression of carotenoid biosynthetic genes (Table s6c). Whereas OsMADS26 and R2R3MYB showed no expression at milky and doughing stages in grains of white cultivar, the purple cultivar showed significantly high expression of both these genes at all stages of grain development; the fold changes in their expression correlated strongly with the expression profiles of genes involved in carotenoid biosynthesis (Table s6c).

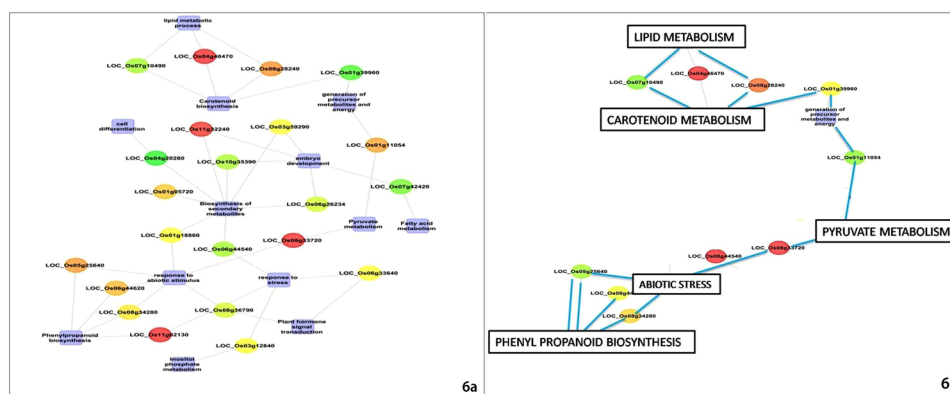
**Validation of differential gene expression.** qRT-PCR of selected genes listed in Table S1 was performed to validate RNASeq based transcriptome profile. A comparative analysis of the selected genes revealed a high degree of consistency between transcript abundance determined by qRT-PCR and RNAseq (Table S8).

## Discussion

Secondary metabolites, such as anthocyanins, flavonols, proanthocyanidins and carotenoids, play important roles in plant development. Even though the basic biosynthetic pathways leading to synthesis of the core secondary metabolites have been well-characterized, the mechanisms underlying various regulatory controls in carotenoid biosynthesis continue to be a matter of debate. Except for a small amount of lutein in the pericarp of mature grains of white rice, we could not detect any other carotenoid intermediate in the grains of white cultivars at any stage of development. On the other hand, mature grains of purple cultivars accumulated significant levels of



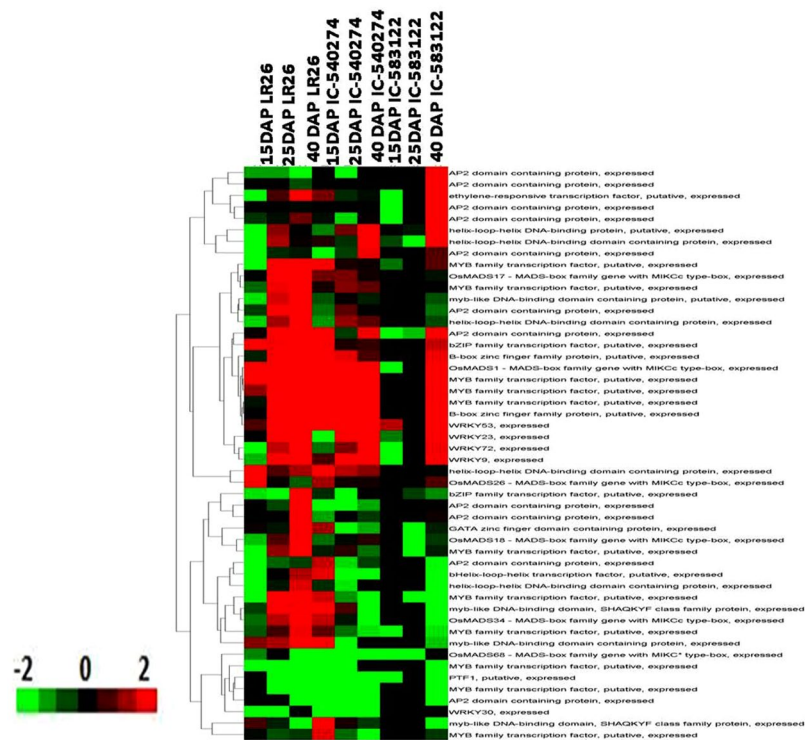
**Figure 5.** Hierarchical cluster analysis revealing gene expression of carotenoid biosynthetic genes from whole grains of white (IC-583122), brown (IC-540274) and purple (LR 26) cultivars of rice from North -East India harvested at milky (15 DAP), doughing (25 DAP) and mature (40 DAP) stages of grain development. Color bars at the bottom of each of the heat maps show the corresponding scale for log<sub>2</sub> fold change in expression. Heat maps were generated for which weighted average linkage method and Pearson correlation distance metric were used.



**Figure 6.** Functional gene network developed from the transcriptome data of grains studied in present investigation. **(a)** Pathways co-expressing with the carotenoid biosynthetic pathway: pyruvate metabolism, lipid metabolism, secondary metabolism, cell differentiation, inositol phosphate metabolism, embryo development, phenyl propanoid pathway, MEP pathway and abiotic stress. **(b)** Pathways/processes directly linked with the carotenoid metabolic pathway: pyruvate metabolism, lipid metabolism, phenyl propanoid pathway, MEP pathway and abiotic stress.

phytoene, lycopene,  $\beta$ -carotene, lutein, and zeaxanthin with LR 26 accumulating upto  $49.16 \pm 5 \mu\text{g}$  of  $\beta$ -carotene,  $28.89 \pm 3.2 \mu\text{g}$  lutein and  $34.65 \pm 0.6 \mu\text{g}$  of zeaxanthin  $\text{gm}^{-1}$  grain fresh weight. Similar reports on high level of carotenoid intermediates in the grains of black rice has been reported by Frei and Becker<sup>13</sup> and Kim *et al.*<sup>14</sup>. Schaub *et al.*<sup>21</sup> have attributed the absence of carotenoids in white rice to the absence of PSY1 transcripts and limited availability of precursor substrates necessary for PDS/ZDS activity in the endosperm. While the grains



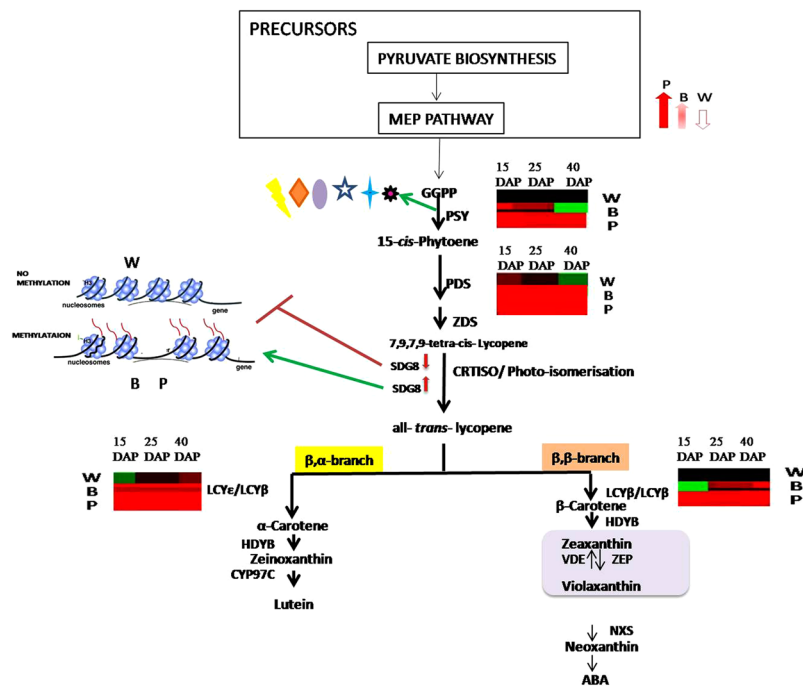


**Figure 7.** Heat map of the correlation analysis of differentially expressed transcription factors from whole grains of IC-583122, IC-540274 and LR26 with the carotenoid biosynthetic genes. The rice grains were harvested from North-East India at 15 DAP, 25 DAP and 40 DAP. Color bars at the bottom of each of the heat maps show the corresponding scale for log<sub>2</sub> fold change in expression.

of white cultivar did not show any expression of PSY1 at milky and doughing stages, expression of this gene was recorded in mature grains of this cultivar. Accumulation of marginal levels of lutein in mature grains of white cultivar can be ascribed to expression of PSY1 gene in grains of this cultivar at maturity.

Irrespective of the cultivar, the number of DEGs recorded in grains at doughing stage and maturity was much higher than that recorded between milky and doughing stages. Enrichment analysis at FDR < 0.05 revealed higher expression of genes involved in signal transduction, amino acid, lipid and energy metabolism, nucleic acid replication/processing and transcriptional regulation at milky stage. On the other hand, genes involved in protein biosynthesis showed higher expression during doughing stage and those involved in storage protein accumulation, starch metabolism and secondary metabolism showed higher expression during grain maturation. These results clearly indicate differential patterning of metabolic pathways during transition of grains from milky stage to maturity. Similar expression kinetics of DEGs has been reported in other cultivars of rice<sup>22</sup> and *Arabidopsis*<sup>23</sup>. Interestingly, our results revealed a concomitant upregulation of the genes involved in ABA [NCEDs (LOC\_Os12g42280, LOC\_Os02g47510), aldehyde oxidase2 (LOC\_Os03g57720)] and GA [ent-kaurene synthase A (LOC\_Os04g36760), KA01 (LOC\_Os10g23160), OsGA20ox1 (LOC\_Os03g42130.1), GA2OX7 (LOC\_Os04g33360, LOC\_Os04g44150)] biosynthesis in the grains during their development. Such a simultaneous upregulation of genes involved in ABA and GA<sub>3</sub> biosynthetic pathways during grain maturation has also been reported in seeds of *Hordeum vulgare*<sup>24</sup>. Sreenivasulu *et al.*<sup>24</sup> suggested that GA not only played an important role in triggering major proteinases during early seed germination but also acted in mobilization of distinct reserve polysaccharides during grain maturation. GA<sub>3</sub> is known to cause endosperm tissue weakening at the micropylar end by modulating expression of genes encoding cell wall remodeling enzymes<sup>25</sup>. In addition, the scutellum of the embryo secretes GA to the aleurone layer of rice grains which is devoid of starch molecules. GA induces the secretion of amylases and proteases that break down the stored starch and proteins in the starchy endosperm, making free sugars and amino acids available for other metabolic processes<sup>25</sup>. In barley, differentiation of nucellar projection, which forms a link to filial organs was shown to be modulated by a shift from lower to higher GA/ABA ratios during grain maturation<sup>26</sup>. It was suggested that the change in GA/ABA ratios was a prerequisite for the formation of a differentiation gradient which is essential for ordered assimilate transfer through the nucellar projection.

Our results showed a positive correlation in expression profile of PSY1 with differences in the level of phytoene and other carotenoid intermediates between the white, brown and purple cultivar, thereby indicating the significance of PSY1 in conditioning carotenoid content and composition in different cultivars of rice. Miller *et al.*<sup>27</sup> have made similar observations on the correlation between PSY1 expression and total carotenoid content in other cultivars of rice. Whereas the ratio of lutein to  $\beta$ -carotene in milky and mature grains of brown cultivar was 1.4 and 3.16, respectively, it was 0.53 and 0.58 at corresponding stages of in the grains of purple cultivar. The



**Figure 8.** Pictorial representation of the carotenogenic regulators modulating in different varieties of rice. Red upward pointing arrows indicates gene expression positively correlated with carotenoid biosynthesis; Red downward pointing arrows indicate gene expression negatively correlated with carotenoid biosynthesis. Green arrows indicates factors that positively regulates gene expression; Red arrow with block, indicates factors that negatively regulate gene expression. Transcription factors and light are represented in different symbols. PSY: phytoene synthase, PDS: phytoene desaturase,  $\beta$ lcyl:  $\beta$ -lycopene cyclase,  $\epsilon$ -LCY:  $\epsilon$ -lycopene cyclase, BCH:  $\beta$ -carotene hydroxylases. W: White, B: brown, P: Purple (Expression of genes has been represented in form of heat map (red-upregulation, green -downregulation and black-no regulation)).

difference in the ratio of lutein to  $\beta$ -carotene in grains of brown and purple cultivars is most likely due to the differences in the expression of  $\beta$ -LCY and  $\epsilon$ -LCY in these cultivars. Differences in the expression of  $\beta$ -LCY and  $\epsilon$ -LCY in grains of brown and purple cultivars determines a preferential flux of lycopene towards the synthesis of either the  $\beta$ -branch carotenoids or  $\epsilon$ -branch carotenoids. The higher expression of  $\beta$ -LCY in grains of purple cultivar would determine the flux of lycopene towards  $\beta$ -branch carotenoids in this cultivar. Such a flux of lycopene towards synthesis of  $\beta$ , $\epsilon$ -carotenoids or  $\beta$ , $\beta$ -carotenoids by activities of  $\beta$ -LCY and  $\epsilon$ -LCY has been well illustrated in crops belonging to Poaceae by Bai *et al.*<sup>28</sup> and Kandianis *et al.*<sup>29</sup>. Bai *et al.*<sup>28</sup> has ascribed the differential pattern of carotenoid expression in maize embryo and endosperm to differential expression of genes encoding lycopene  $\beta$ -cyclase and lycopene  $\epsilon$ -cyclase in these two tissues.

Coexpression networking clearly indicated a strong relationship between pyruvate metabolism and carotenoid biosynthesis in the grains. Cultivars with higher carotenoid levels in their grains also showed higher expression of genes involved in pyruvate metabolism. Such relationships between enhanced pyruvate flux and carotenoid accumulation in the grains have also been reported in rice<sup>3</sup>, maize<sup>30</sup> and *Arabidopsis*<sup>31</sup>. Additionally, quantitative proteome analysis carried by Chen *et al.*<sup>32</sup> in grains of black rice (*O. sativa* L. *indica* var. SSP) also revealed high expression of enzymes belonging to carbohydrate metabolism as well as high sugar levels in the grains of black rice. It is thus possible that higher expression of genes involved in pathways lying upstream of carotenoid metabolism in the purple cultivar (LR 26) makes enhanced pool of metabolites available to the isoprenoid biosynthesis pathway for synthesis of carotenoids. Grains of purple cultivar also had higher expression of 9-cis-epoxycarotenoid dioxygenases that provide the precursors for ABA biosynthesis. This correlates with the observations of Meier *et al.*<sup>31</sup>, Lindgren *et al.*<sup>33</sup> and Barickman *et al.*<sup>34</sup> who have argued for a strong relationship between ABA biosynthesis and carotenoid levels in plants. We opine that both upstream pathways supplying carotenoid precursors as well as downstream pathways that metabolize or sequester carotenoids to deplete the carotenoid pool have an important bearing on the rate of carotenoid biosynthesis in rice. Purple rice also showed a consistently higher expression of transcripts coding for DnaJ10 and DnaJ domain-containing proteins. OR encodes a Cys-rich DnaJ-like zinc finger domain-containing protein<sup>35</sup>. Li *et al.*<sup>35</sup> and Lu *et al.*<sup>36</sup> have observed that DnaJ-like zinc finger domain-containing protein induced chromoplast differentiation thereby enhancing the storage sink for carotenoids. Overexpression of Or gene from cauliflower has been reported to enhance accumulation of carotenoids in potato tubers by inducing differentiation of chromoplasts with enhanced sink strength<sup>37</sup>. The high expression of transcripts for DnaJ10 and DnaJ domain-containing proteins in grains of purple rice could be inducing enhanced sink strength for carotenoid sequestration in this cultivar. Our results also showed higher expression of transcripts coding for SET domain-containing protein in grains of purple cultivar. The fold change in expression of SET domain-containing protein correlated with changes in the level of carotenoid intermediates,

especially so during grain filling, thereby indicating a role for these proteins in regulating carotenogenesis in plants. Cazzonelli *et al.*<sup>38</sup> had earlier demonstrated the essential requirement of histone methyltransferase (SET DOMAIN GROUP 8, *SDG8*) that methylates histone H3 on Lys 4 and/or 36 (H3K4 and H3K36) for maintenance of permissive expression of carotenoid isomerase (CRTISO) in *A. thaliana*. CRTISO is required for converting *cis* form of lycopene to *trans* form which is essential for cyclization to beta-carotene. Mutations in *SDG8* (*ccr1*) were shown to result in a significant reduction in transcript levels of CRTISO, leading to the accumulation of *cis*-isomers of carotenoid intermediates<sup>38</sup>.

Correlation analysis between the genes involved in carotenoid biosynthesis and various families of transcription factors revealed higher expression of light specific transcription factors in the grains of purple and brown cultivar than that in white rice. Similar observations on the involvement of WRKY, MYBs, bZIPs and bHLHs (PIFs) in transcriptional regulation of genes involved in carotenoid biosynthesis have been made by Mohanty *et al.*<sup>39</sup> and Liang and Jiang<sup>40</sup>. Whereas information on specific transcription factors controlling the expression of carotenoid biosynthetic genes is available for horticultural crops, information on such proteins in controlling carotenogenesis in cereals is sketchy. Our results show a strong correlation between the carotenoid content in the different rice cultivars and the expression of *OsMADS26* and *R2R3MYB* genes. Although *OsMADS26* has been reported to manifest phenotypes related to pigment accumulation in rice<sup>41</sup> and *R2R3MYB* has been regarded as candidate gene for hull pigmentation in rice<sup>42</sup>, evidences of involvement of MADS and *R2R3MYB* have reported in controlling carotenoid accumulation in tomato<sup>43</sup> and *Mimulus lewisii*<sup>44</sup> flowers respectively.

Our investigation points towards a multifaceted regulation of carotenoid biosynthesis in rice. Besides the transcriptional regulation of genes such as *PSY1*, *PDS*,  $\beta$ -*LCY* and  $\epsilon$ -*LCY*, additional regulatory controls leading to higher carotenoid content in purple rice include pathway crosstalk between the upstream channels (pyruvate metabolism, MEP pathway), as well as the downstream pathways including carotenoid sequestration and ABA biosynthesis. Co-expression analysis indicates a positive relationship between expression of genes involved in carotenoid biosynthesis and light specific transcription factors such as WRKY, MYBs, bZIPs and bHLHs, glutathione S-transferase, Or protein and SET DOMAIN containing protein genes, thereby suggesting a role of these factors in regulating carotenogenesis in rice (Fig. 8). High expression of Or protein encoding gene and SET DOMAIN in the purple cultivar indicates a role of sequestration site and epigenetic factors such as *SDG8*-mediated histone methylation of regulatory domain of CRTISO in regulating rice carotenogenesis.

## References

- Ye, X. *et al.* Engineering the provitamin A  $\beta$ -carotene biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**(5451), 303–305 (2000).
- Bai, C. *et al.* Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor-product balance. *Plant Biotechnol J.* **14**(1), 195–205 (2016).
- Gayen, D. *et al.* Metabolic Regulation of Carotenoid-Enriched Golden Rice Line. *Front. Plant Sci.* **7**, 1622 (2016).
- Robertson, D. S. Phenotypic and genetic studies of a new mutant of yellow endosperm in maize. *J Hered.* **66**(3), 127–130 (1975).
- Li, Z. H., Matthews, P. D., Burr, B. & Wurtzel, E. T. Cloning and characterization of a maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway. *Plant Mol Biol.* **30**, 269–279 (1996).
- Wurtzel, E. T., Luo, R. & Yatou, O. A simple approach to identify the first rice mutants blocked in carotenoid biosynthesis. *J Exp Biol.* **52**(354), 161–166 (2001).
- Fang, J. *et al.* Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *The Plant J.* **54**(2), 177–189 (2008).
- Seechamnaturakit, V., Karrila, T. T., Sontimuang, C. & Sukhoom, A. The natural pigments in pigmented bran and their relation to human health: A literature review. *KMUTNB. Int J ApplSci Technol.* **11**(1), 3–13 (2018).
- Yodmanee, S., Karrila, T. T. & Pakdeechanuan, P. Physical, chemical and antioxidant properties of pigmented rice grown in Southern Thailand. *IFRJ.* **18**(3), 901–906 (2011).
- Pengkumsri, N. *et al.* Physicochemical and antioxidative properties of black, brown and red rice varieties of northern Thailand. *Food Sci Technol(Campinas).* **35**(2), 331–338 (2015).
- Pereira-Caro, G., Cros, G., Yokota, T. & Crozier, A. Phytochemical profiles of black, red, brown, and white rice from the Camargue region of France. *J Agric Food Chem.* **61**(33), 7976–7986 (2013).
- Melini, V. & Acquistucci, R. Health-Promoting Compounds in Pigmented Thai and Wild Rice. *Foods.* **6**(1), 9 (2017).
- Frei, M. & Becker, K. Agro-biodiversity in subsistence-oriented farming systems in a Philippine upland region: nutritional considerations. *Int J BiodiversConserv.* **13**(8), 1591–1610 (2004).
- Kim, J. *et al.* Variation and correlation analysis of flavonoids and carotenoids in Korean pigmented rice (*Oryzasativa* L.) cultivars. *J Agric Food Chem.* **58**(24), 12804–12809 (2010).
- Panfili, G., Alessandra, F. & Mario, I. Improved normal-phase high-performance liquid chromatography procedure for the determination of carotenoids in cereals. *J Agric Food Chem.* **52**(21), 6373–6377 (2004).
- Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS one.* **7**(2), e30619 (2012).
- Ye, J. *et al.* WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* **34**, W293–W297 (2006).
- Zhang, J. *et al.* Molecular characterization and transcriptome analysis of orange head Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Planta.* **241**(6), 1381–1394 (2015).
- Ye, J. *et al.* Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *PLoS One.* **10**(7), e0130885 (2015).
- Livak, K. J. & Thomas, D. S. Analysis of relative gene expression data using real-time quantitative PCR and the 2–CT method. *Methods.* **25**(4), 402–408 (2001).
- Schaub, P., Al-Babili, S., Drake, R. & Beyer, P. Why is golden rice golden (yellow) instead of red? *Plant Physiol.* **138**(1), 441–450 (2005).
- Xu, H., Gao, Y. & Wang, J. Transcriptomic analysis of rice (*Oryza sativa*) developing embryos using the RNA-Seq technique. *PLoS One.* **7**(2), e30646 (2012).
- Fait, A. *et al.* *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* **142**, 839–854 (2006).
- Sreenivasulu, N. *et al.* Barley grain maturation and germination: metabolic pathway and regulatory network commonalities and differences highlighted by new MapMan/PageMan profiling tools. *Plant Physiol.* **146**(4), 1738–1758 (2008).

25. Yan, D., Duermeyer, L., Leoveanu, C. & Nambara, E. The functions of the endosperm during seed germination. *Plant Cell Physiol.* **55**(9), 1521–1533 (2014).
26. Weier, D. *et al.* Gibberellin-to-abscisic acid balances govern development and differentiation of the nucellar projection of barley grains. *J Exp Bot.* **65**(18), 5291–5304 (2014).
27. Miller, B., Grunden, H. & Wiley, E. J. Carotenoid metabolism is induced in rice bran during very high temperature stress. *J Sci Food Agric.* **94**(9), 1808–1815 (2014).
28. Bai, L., Kim, E. H., DellaPenna, D. & Brutnell, T. P. Novel lycopene epsilon cyclase activities in maize revealed through perturbation of carotenoid biosynthesis. *Plant J.* **59**(4), 588–599 (2009).
29. Kandianis, C. B. *et al.* Genetic architecture controlling variation in grain carotenoid composition and concentrations in two maize populations. *Theor Appl Genet.* **126**(11), 2879–2895 (2013).
30. Decourcelle, M. *et al.* Combined transcript, proteome, and metabolite analysis of transgenic maize seeds engineered for enhanced carotenoid synthesis reveals pleiotropic effects in core metabolism. *J Exp Bot.* **66**(11), 3141–3150 (2015).
31. Meier, S., Tzfadia, O., Vallabhaneni, R., Gehring, C. & Wurtzel, E. T. A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. *BMC Sys Biol.* **5**(1), 77 (2011).
32. Chen, L. *et al.* iTRAQ-based quantitative proteomics analysis of black rice grain development reveals metabolic pathways associated with anthocyanin biosynthesis. *PLoS one* **11**(7), e0159238 (2016).
33. Lindgren, L. O., Stålberg, K. G. & Höglund, A. S. Seed-specific overexpression of an endogenous *Arabidopsis* phytoene synthase gene results in delayed germination and increased levels of carotenoids, chlorophyll, and abscisic acid. *Plant Physiol.* **132**(2), 779–785 (2003).
34. Barickman, T. S., Kopsell, D. A. & Sams, C. E. Abscisic acid increases carotenoid and chlorophyll concentrations in leaves and fruit of two tomato genotypes. *J Amer Soc Hort Sci.* **139**, 261–266 (2014).
35. Li, L. *et al.* The *Or* gene enhances carotenoid accumulation and stability during post-harvest storage of potato tubers. *Mol Plant.* **5**(2), 339–352 (2012).
36. Lu, S. *et al.* The cauliflower *Or* gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. *Plant Cell.* **18**(12), 3594–3605 (2006).
37. Lopez, A. B. *et al.* Effect of the cauliflower *Or* transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. *J Exp. Biol.* **59**(2), 213–23 (2008).
38. Cazzonelli, C. I., Miller, T., Finnegan, E. J. & Pogson, B. J. Promoting gene expression in plants by permissive histone lysine methylation. *Plant Signal Behav.* **4**(6), 484–488 (2009).
39. Mohanty, B. *et al.* Light-specific transcriptional regulation of the accumulation of carotenoids and phenolic compounds in rice leaves. *Plant Signal Behav.* **11**(6), e1184808 (2016).
40. Liang, M. H. & Jiang, J. G. Analysis of carotenogenic genes promoters and WRKY transcription factors in response to salt stress in *Dunaliellabardawil*. *Sci Rep.* **7**, 37025 (2017).
41. Lee, S. *et al.* Further characterization of a rice AGL12 group MADS-box gene, OsMADS26. *Plant Physiol.* **147**(1), 156–168 (2008).
42. Sun, X. *et al.* The CSA gene system regulates hull pigmentation and reveals evolution of anthocyanin biosynthesis pathway in rice. *J. Exp. Biol.* **69**(7), 1485–1498 (2018).
43. Martel, C., Vrebalov, J., Tafelmeyer, P. & Giovannoni, J. J. The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. *Plant Physiol.* **157**(3), 1568–1579 (2011).
44. Sagawa, J. M. *et al.* An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers. *New Phytol.* **209**(3), 1049–1057 (2016).

## Acknowledgements

Financial support received from Department of Biotechnology, Govt. of India, New Delhi for undertaking the work under DBT Biotech Hub project vide grant no. BT/04/NE/2009 to NKC and INSPIRE fellowship from Department of Science & Technology, Govt. of India to UC is gratefully acknowledged.

## Author Contributions

N.K.C. conceived and designed the study. U.C. performed the bench work as well as computational analysis. K.K. carried out the raw data analysis and pathway networking. N.K.C. and U.C. interpreted the results and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-38233-8>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019