



OPEN Transcriptome analysis of the coexpression network of genes related to antioxidant characteristics after grain filling in purple rice

Changhui Sun^{1,4}, Junkai Zhu^{1,4}, Ying Zhu¹, Jinxia Cao¹, Jiao Zhang², Yanqiong Zhang¹, Huijuan Zhou¹, Yangang Zhu¹, Yiming Ji¹, Rui Ding¹, Qiangqiang Xiong^{2,3}✉ & Xiaobin Liu¹✉

Antioxidant capacity is an important indicator for evaluating the growth and developmental quality of rice. This study has guiding significance for the cultivation of high-nutrient-value varieties. To investigate the molecular mechanisms underlying the antioxidant characteristics of rice grains after the filling stage, Yangzinuo 1 (YZN1) was used as the experimental material, and grains collected at five different time points (7 days apart) after the filling stage were used for transcriptome sequencing. Through weighted gene coexpression network analysis (WGCNA), a coexpression network of gene weights related to antioxidant characteristics was constructed. *LOC_Os10g39140*, *LOC_Os10g38276*, and *LOC_Os05g45740* were identified from the 2 modules showing the highest correlations with the target traits. GO functional annotation showed that target modules were enriched in pathways related to phenylalanine, flavonoids, and other related pathways, such as GO:0006558, GO:0006559, GO:0009812, and GO:0009813. Correlation analysis with metabolites revealed that differentially expressed genes were significantly enriched in pathways related to antioxidant characteristics and energy metabolism processes, such as glycolysis/gluconeogenesis and flavonoid biosynthesis. The core genes identified in this study were found to be highly correlated with antioxidant characteristics and enriched in pathways related to metabolic and energy pathways and molecular activities. These results provide an effective dataset supporting breeding targeting functional rice characteristics.

Keywords Purple rice, Antioxidant characteristics, WGCNA, Transcriptome, Genes

Rice is the main food crop for more than half of the world's population. With the improvement of living standards and the ongoing development of the rice trade, the demand for high-quality rice is increasing. Therefore, breeding high-quality rice is crucial to meet market demand¹. The completion of rice genome sequencing has laid a solid foundation for extracting valuable information from vast biological datasets. In contrast to traditional molecular biology methods, bioinformatics tools not only expedite the identification of target genes but also enhance the evaluation of key genes associated with specific traits. WGCNA is a method in which a scale-free topology overlap matrix is constructed based on RNA-seq expression data, which describes the interactions between genes and divides genes with similar expression patterns into gene expression modules. WGCNA is predominantly used to investigate the biological relationships between coexpressed gene modules and target traits, while also identifying core genes within the coexpression network. As a representative systems biology method, WGCNA has found broad application in plant research². For example, 22 gene modules were identified by analysing 17 rice (*Oryza sativa*) transcriptome datasets obtained at different time points under cadmium treatment, and combined with differential expression analysis, a total of 164 cadmium stress response-related genes were revealed³. To further accurately annotate the rice genome, WGCNA can integrate expression datasets generated under different experimental conditions and biological systems, and different periods are

¹Jiangsu Kingearth Seed Co., Ltd., Yangzhou 225009, China. ²Jiangsu Key Laboratory of Crop Genetics and Physiology, Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College of Yangzhou University, Yangzhou 225009, China. ³Jiangxi Irrigation Experiment Central Station, Nanchang 330201, China. ⁴Changhui Sun and Junkai Zhu contributed equally to this work. ✉email: xqq_qiang@163.com; jtdzyzy@126.com

a more important variable³. WGCNA has also been employed to analyse transcriptome data from two cotton strains at various developmental stages, identifying five fibre development-specific modules. Additionally, core genes were identified within these modules⁴. By performing WGCNA on transcriptome data from 14 different developmental stages of maize (*Zea mays*), researchers identified 14 tissue-specific modules and further studied the gene interaction networks in two of the modules, discovering flowering-related core genes such as *ZCN8*, *ZCN7*, and *COL1*⁵.

Antioxidant capacity is an important indicator of rice growth and development quality. Rice varieties with a high antioxidant capacity can effectively resist nonbiological environmental stresses such as drought and waterlogging^{6,7}. Enhancing the antioxidant capacity can also effectively alleviate salt toxicity and achieve an effective response to salt stress during plant growth and development. To ensure high-yield and high-quality rice production goals, many methods have been applied to enhance antioxidant capacity. For example, the foliar application of methyl jasmonate (MeJA) can increase the content of 2-acetyl-1-pyrroline (2-AP) in aromatic rice, thereby regulating antioxidant properties and further promoting yield formation⁸. The foliar spraying of salicylic acid can also enhance the antioxidant capacity, which is beneficial for the growth and yield formation of cereal crops⁶. In addition, in the context of breeding, a strong antioxidant capacity has considerable guiding significance for the cultivation of high-nutrient-value varieties⁹. Ongoing research in this area, aimed at accelerating breeding strategies for these special rice varieties, will significantly contribute to improving dietary health, advancing agricultural development, and boosting farmers' incomes¹. Brown rice with a red, purple, or black pericarp is more beneficial to human health than traditional white-pericarp rice because of the accumulation of more antioxidant compounds^{10,11}. There is a strong correlation between flavonoid and phenolic metabolite contents and total antioxidant capacity. The contents of flavonoids and oligomeric proanthocyanidins in purple rice and the total antioxidant capacity are significantly higher than those in white rice¹². Related studies have also shown that the total antioxidant capacity and phenolic substance content of red rice varieties are higher. Furthermore, varieties with lower nutritional quality but a higher antioxidant capacity are considered better¹³. The elevated polyphenol content in red rice is a key factor contributing to its significantly higher antioxidant capacity compared to white rice. Moreover, red rice retains its nutritional properties more effectively during processing¹⁴. More phytochemicals are observed in black rice bran with high antioxidant activity, providing researchers with opportunities to cultivate new genotypes of rice with higher nutritional value¹⁵.

In this study, we used YZN1, a type of purple sticky rice, as the experimental material and collected transcriptome data at five time points after grain filling. Differential expression analysis was performed on the data. By constructing a weighted gene coexpression network, genes were classified into modules, and specific modules related to antioxidant enzymes were selected. Core genes related to antioxidant properties were identified within these modules. This study provides a theoretical basis for further elucidating the antioxidant characteristics of purple rice grains after filling, offering new gene resources for breeding special rice varieties.

Methods

Plant materials and growth conditions

The experimental material used in this study was the purple rice cultivar YZN1, which was independently bred by the College of Agriculture, Yangzhou University. The rice was grown at the Shatou Experimental Station of Yangzhou City, Jiangsu Province. The rice was sown on May 20, 2022, and transplanted at age 25 days, with 4 seedlings per hill and a spacing of 30 cm × 12 cm. Three replicates were set up for the cultivation of YZN1. Compound fertilizer with a nitrogen, phosphorus, and potassium ratio of 15% each was used for rice fertilization. A total of 300 kg of pure nitrogen was applied per hectare, with a ratio of 5:3:2 for the basal, tillering, and panicle fertilizers. The prevention and control of diseases, pests, and weeds were carried out in accordance with the conventional high-yield cultivation requirements for rice.

Sample collection

Three hundred panicles of rice with similar panicle types, growth and size were selected from each plot during the heading stage of rice. The selected panicles were marked and tagged. After 7 (Y1G1), 14 (Y1G2), 21 (Y1G3), 28 (Y1G4), and 35 (Y1G5) days of grain filling, marked rice panicles were quickly selected and taken back to the laboratory. Using tweezers, the rice shells were removed, and the grains were placed in cryovials before being frozen in liquid nitrogen. Three biological replicates, each comprising a mixed sample of 10 rice panicle grains, were prepared. The samples were stored in a -20 °C freezer until data measurement. In our previous work, we conducted an analysis of physiological and biochemical indicators in each sample, including catalase (CAT), polyphenol oxidase (PPO), phenylalanine ammonia-lyase (PAL), total phenols (TP), flavonoids (FD), oligomeric proanthocyanidin (OPC), and total antioxidant capacity (ABTS method, 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid; DPPH method, 2,2-diphenyl-1-picrylhydrazyl; FRAP method, ferric ion reducing antioxidant power method)¹⁶.

RNA extraction

Total RNA was extracted from the tissue using TRIzol[®] Reagent (Plant RNA Purification Reagent for plant tissue) according to the manufacturer's instructions (Invitrogen), and genomic DNA was removed using DNase I (TaKaRa). RNA degradation and contamination were monitored on 1% agarose gels. Then, RNA quality was determined with a 2100 Bioanalyzer (Agilent Technologies) and quantified using an ND-2000 instrument (NanoDrop Technologies). Only high-quality RNA samples (OD260/280 = 1.8 ~ 2.2, OD260/230 ≥ 2.0, RIN ≥ 8.0, 28 S:18 S ≥ 1.0, > 1 µg) were used to construct the sequencing library.

Library preparation and sequencing

RNA purification, reverse transcription, library construction and sequencing were performed at Shanghai Majorbio Biopharm Biotechnology Co., Ltd. (Shanghai, China) according to the manufacturer's instructions (Illumina, San Diego, CA). The transcriptome library was prepared following the instructions of the TruSeq™ RNA sample preparation kit from Illumina (San Diego, CA) using 1 µg of total RNA. Briefly, messenger RNA was isolated according to the polyA selection method by using oligo(dT) beads and then fragmented with fragmentation buffer. Second, double-stranded cDNA was synthesized using a SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) with random hexamer primers (Illumina). Then, the synthesized cDNA was subjected to end repair, phosphorylation and 'A' base addition according to Illumina's library construction protocol. Libraries were size selected for cDNA target fragments of 300 bp on 2% Low Range Ultra Agarose, followed by PCR amplification using Phusion DNA polymerase (NEB) for 15 PCR cycles. After quantification with a TBS380 system, a paired-end RNA-seq library was sequenced with an Illumina NovaSeq 6000 sequencer (2 × 150 bp read length).

Quality control and read mapping

The raw paired-end reads were trimmed and subjected to quality control with fastp (<https://github.com/OpenGene/fastp>)¹⁷ with the default parameters. Then, clean reads were separately aligned to the reference genome in orientation mode using HISAT2 (<http://ccb.jhu.edu/software/hisat2/index.shtml>)¹⁸ software. The mapped reads of each sample were assembled by StringTie (<https://ccb.jhu.edu/software/stringtie/>) according to a reference-based approach¹⁹. Reference genome version: IRGSP-1.0 (<http://rice.uga.edu/>).

Differential expression analysis and GO enrichment analysis

To identify differentially expressed genes (DEGs) between two different samples/groups, the expression level of each gene was calculated according to the transcripts per million reads (TPM) method. RSEM (<http://deweylab.biostat.wisc.edu/rsem/>)²⁰ was used to quantify gene abundances. Essentially, differential expression analysis was performed using DEGseq²¹, and DEGs with a $|\log_2(\text{fold change})| \geq 1$ and $P\text{-adjusted} \leq 0.001$ were considered significantly differentially expressed genes. In addition, GO enrichment analysis (Gene Ontology, <http://www.geneontology.org>) was performed to identify which DEGs were significantly enriched in GO terms and metabolic pathways according to a $P\text{-adjust} \leq 0.05$ compared with the whole-transcriptome background. GO functional enrichment was carried out with Goatools (<https://github.com/tanghaibao/Goatools>) and KOBAS²².

Weighted correlation network analysis (WGCNA)

We used the WGCNA package in R project to construct a weighted gene coexpression network. The input was a normalized gene expression matrix from 15 transcriptome samples (5 stages, each with three replicates, according to the results of DEG analysis). We integrated the differentially expressed genes associated with the antioxidant properties of fully filled grains across five stages with previously published physiological and biochemical indicator data. The mean expression values were set to 1, and the coefficient of variation was set to 0.1. After background correction and standardization of gene expression data, genes with outliers and low variability were filtered out. This process was conducted to ensure that the resulting correlation strengths between genes followed a scale-free distribution. Through WGCNA and gene set-phenotype correlation analysis, we aimed to explore the core genes regulating the network of antioxidant properties and the nutrient composition in fully filled purple rice grains during the grain-filling stage. We selected the top 50% of genes with the highest variability in expression levels across samples by calculating the variance of each gene. After threshold filtering, we used $\beta = 9$ to raise the original scale-free matrix to a power to obtain the unscaled adjacency matrix. To better evaluate the correlation of gene expression patterns, we further transformed the adjacency matrix into a TOM and used dissTOM (1-TOM) as the topological dissimilarity matrix. We then used the dynamic cut method to perform gene clustering and module partitioning. The minimum number of genes in a module was set to 30 ($\text{minModuleSize} = 30$), and the network type was "signed" ($\text{type} = \text{"signed"}$ or $\text{networkType} = \text{"signed"}$, depending on the function).

To construct the coexpression network of DEGs, it was necessary to calculate the correlation coefficient between each gene pair and then use the formula $\text{Smn} = \text{cor}(\text{xm}, \text{xn})$ and $S = [\text{Smn}]$ to obtain the similarity matrix between genes. Smn represents the Pearson correlation coefficient between gene m and gene n, and S represents the similarity matrix. The WGCNA package in R software was used to construct the weighted gene coexpression network. To make the network conform to the scale-free distribution, the pickSoftThreshold function in the WGCNA package was used to calculate the weight value. According to the results shown in Fig. 1A–B, a soft power threshold value of $\beta = 9$ was chosen to construct the coexpression network. After determining the soft power threshold value of $\beta = 9$, the similarity matrix was converted to an adjacency matrix according to the formula $\text{Amn} = [(1 + \text{Smn})/2]^\beta$, and the adjacency matrix was then converted to the topological overlap matrix (TOM). In addition, the dissTOM = 1/TOM function was used to obtain the dissimilarity matrix of the topological overlap matrix, which was aimed at eliminating the errors caused by background noise and false association. Finally, the hclust function was used to perform hierarchical clustering on the dissimilarity matrix. The dynamic tree cut method was used to cut the resulting clustering tree. This process can merge genes with similar expression patterns into the same branch, where each branch represents a coexpression module, with different colours representing different modules. Differentially expressed genes were analysed and clustered based on their fragments per kilobase million (FPKM) values. We built upon the foundation of various physiological and biochemical indicators from our earlier work¹⁶. For module division, parameters such as minModuleSize were set to 30, minKMEtoStay was set to 0.3, and mergeCutHeight was set to 0.25, and genes with high correlations were assigned to the same module (Fig. S3).

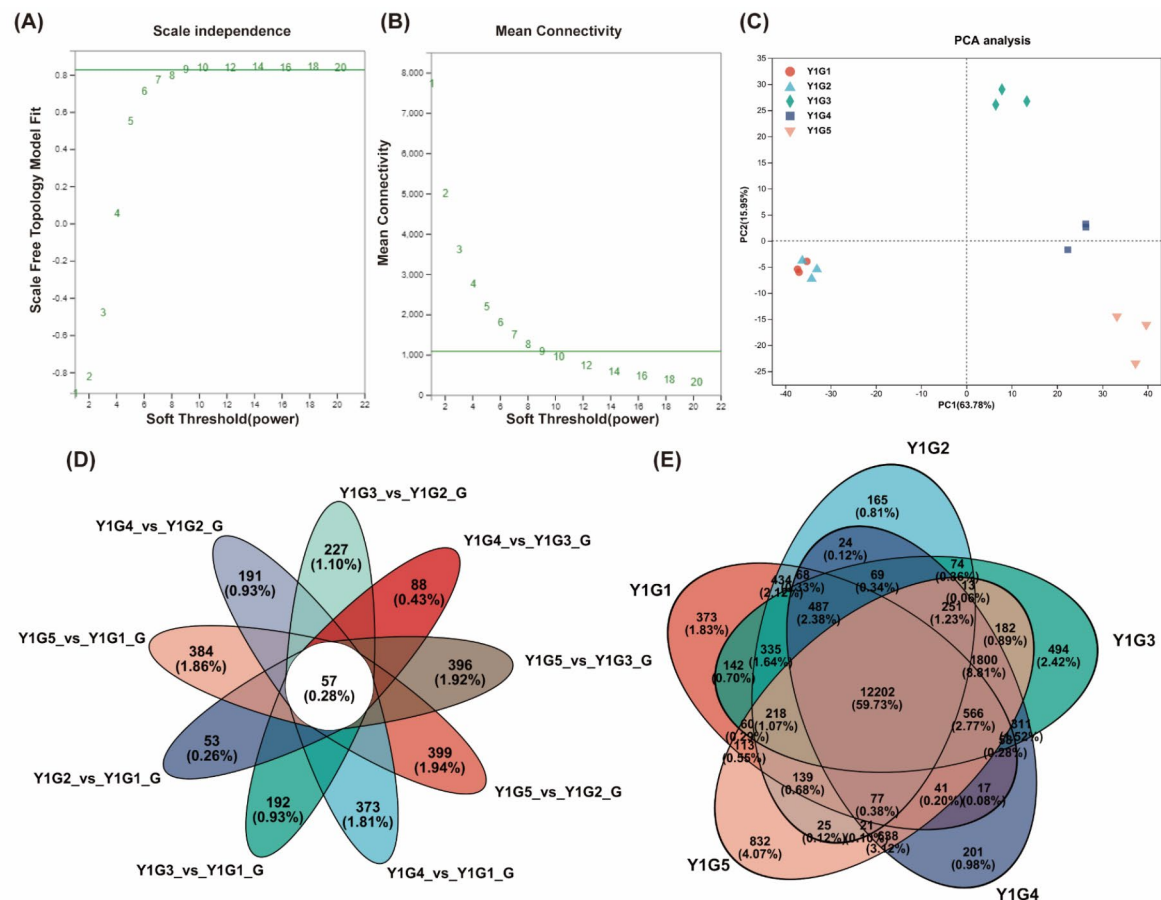


Fig. 1. Grain sample analysis. **(A)** Scale independence. The abscissa represents the power exponent-weighted beta value, and the ordinate represents the fitting degree (R^2) between the corresponding beta value-transformed adjacency matrix and the scale-free network assumption. **(B)** Mean connectivity. The horizontal axis represents the power exponent-weighted β value, and the vertical axis represents the degree of fit (R^2) between the corresponding β value-transformed adjacency matrix and the assumption of the scale-free network. **(C)** PCA plot. **(D)** Venn diagram showing the distribution of genes/transcripts among different samples/groups. **(E)** Venn diagram.

qRT-PCR (quantitative reverse transcription polymerase chain reaction)

Based on the enrichment results, we selected differentially expressed genes that were involved in activities such as peroxidase activity, metabolism, and ATP metabolism at different time points. We then selected eight genes for qRT-PCR validation using the Thermo Fisher Scientific QuantStudio 5 real-time fluorescent quantitative PCR system and the 2× SYBR Green qPCR Mix kit from Novogene. SYBR Green was used as the fluorescent dye, and β -actin was used as the reference gene. The reaction program included an initial denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 15 s and annealing at 60 °C for 15 s. The melting curve program was set to 95 °C for 15 s, 60 °C for 1 min, and 95 °C for 15 s. The reaction system consisted of 10 μ L of 2× SYBR qPCR Mix, 0.8 μ L of DNA template (diluted 10 times), 0.4 μ L of the forwards primer (10 μ mol·L⁻¹), 0.4 μ L of the reverse primer (10 μ mol·L⁻¹), and 8.4 μ L of ddH₂O. The relative expression levels of the genes were analysed using the $2^{-\Delta\Delta C_t}$ method with three biological replicates²³. The gene primers used are listed in Table S1.

Data analysis

The grain physiological and biochemical data were organized, average values were calculated, and graphs were drawn using Excel 2019 software. The variance analysis of the grain physiological and biochemical data¹⁶ was conducted using SPSS 18.0 software. Adobe Illustrator 2022 software was used to combine the various graphs.

Results

Sequencing quality statistics

Transcriptome sequencing was performed on 15 grain samples, generating a total of 108.37 Gb of clean data. Each sample showed over 6.5 Gb of clean data and a Q30 base percentage above 93.36%, meeting the sequencing requirements. The results are available for further analysis (Table S2).

The results of principal component analysis (PCA) showed that PC1 and PC2 accounted for 63.78% and 15.95%, respectively, of the gene expression variation among all samples. The combined contribution of these two principal components was 79.73%. This analysis provided a preliminary understanding of the overall transcriptional differences among the sample groups and the degree of variation within each group (Fig. 1C). The different transcriptional patterns of developing seeds at various stages after purple rice grain filling were also consistent with the PCA results (Fig. S1).

DEG analysis of different developmental stages

The Venn diagram of DEGs at different developmental stages after purple rice grain filling indicated that there were specific genes expressed in the grains at each stage. A total of 12,202 DEGs were identified, accounting for 59.73% of all DEGs (Fig. 1E). A total of 38,879 DEGs were identified in the 5 developmental stages of grains using the criteria of a $|\log_2FC| \geq 1$ and $p_{\text{adjust}} < 0.05$ (Fig. 2A). The number of differentially expressed genes

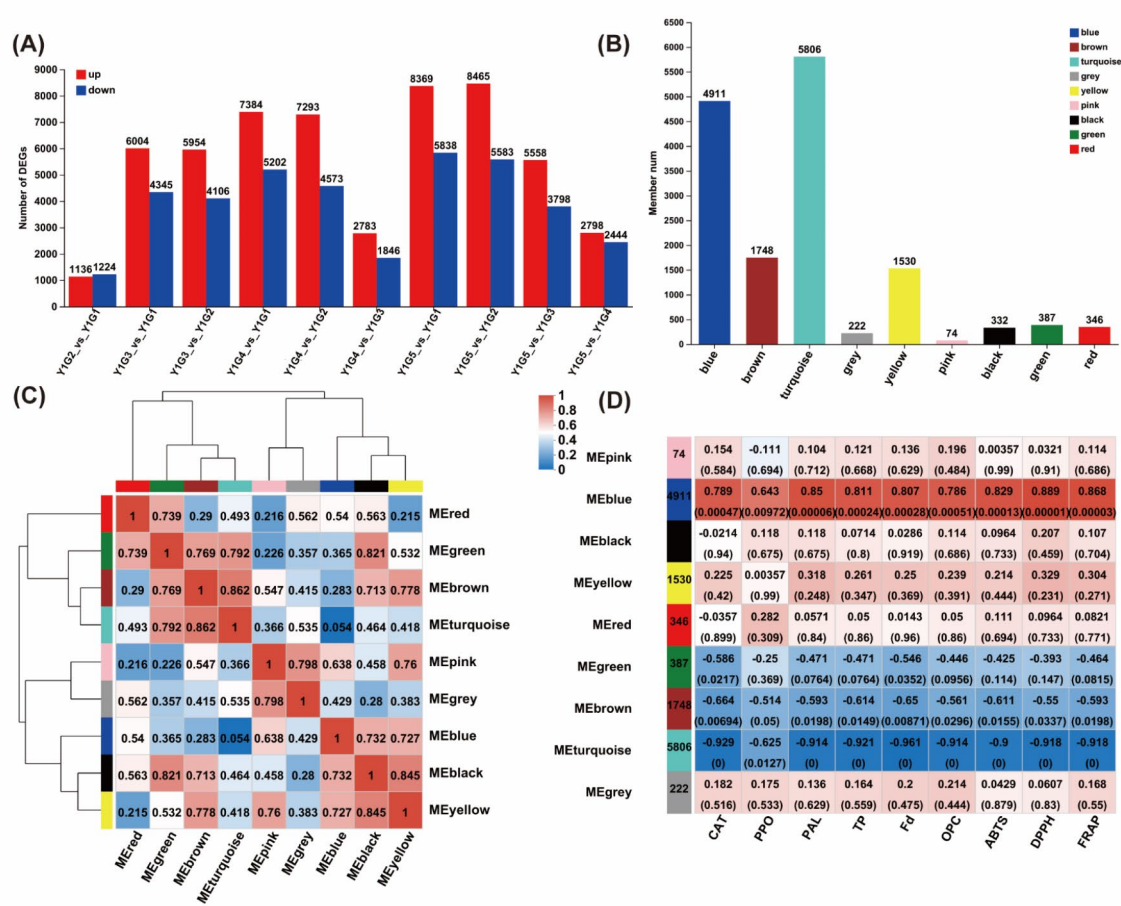


Fig. 2. DEGs and gene module information. **(A)** The statistical graph of differential expression shows different comparison groups on the x-axis and the corresponding number of upregulated and downregulated genes/transcripts on the y-axis, with red representing upregulation and blue representing downregulation. **(B)** The module membership statistical graph shows the number of members belonging to each module (represented by module colour) on the y-axis. **(C)** The module correlation graph shows the correlations between different modules. Red represents a higher correlation between modules, and green represents a lower correlation. **(D)** The module-phenotype correlation analysis graph shows different phenotypes/samples/groups on the x-axis and different modules on the y-axis. The number of genes/transcripts for each module is displayed in the left column, and each group of data on the right represents the correlation coefficient and significance P value (in parentheses) between the module and phenotype. A larger absolute value indicates a stronger correlation, with blue representing a negative correlation and red representing a positive correlation, as indicated by the colour scale in the lower right corner.

and their up- and down-regulation between the comparison groups at each developmental stage can be found in Fig. 1E and Tables S3–S12. The distribution of upregulated and downregulated DEGs is shown in a more intuitive manner in a scatter plot (Fig. S2).

The number and proportion of shared DEGs between different comparison groups are different. We can conclude that the number of DEGs and the up- and downregulation of different comparison groups are quite different according to Fig. 1D.

WGCNA reveals associations between modules and physiological and biochemical traits

Based on the 9 physiological and biochemical traits that we identified in the preliminary study, we obtained 9 modules that were associated with these traits in different treatment periods, each of which was represented by a different colour (Fig. 2D). Some modules were highly correlated with the physiological and biochemical traits, with the turquoise module including the largest number of genes (5806), followed by the blue module with 4911 genes, while the pink module comprised the fewest genes, with only 74 (Fig. 2B). Due to the large number of genes in each module, a dimensionality reduction approach was used to study gene modules by selecting representative gene-module eigengenes (ME). Using ME to represent the thousands of genes in a module, we conducted a correlation analysis, which allowed us to further clarify the relationships between gene modules and phenotypic traits at different treatment times and to screen for target gene modules. After performing clustering analysis on all MEs, we found that some MEs were highly correlated with each other (Fig. 2C). The higher the correlation is between MEs, the higher the correlation of the module in which they are located will be. Through correlation analysis between MEs, we found that the correlation between MEs in the brown and turquoise modules reached 0.86, and the correlation between MEs in the yellow and black modules reached 0.85 (Fig. 2C). Moreover, the gene expression patterns in the turquoise and blue modules gradually changed from a negative correlation with antioxidant-related physiological and biochemical traits to a highly significant positive correlation. While the correlation values of the 9 physiological and biochemical indicators in the blue module were all greater than 0.643, those in the turquoise module were all less than −0.625. Additionally, the gene expression patterns in the turquoise and blue modules in different treatment periods were highly correlated with the target traits (Fig. 2D), indicating that these two modules could be considered target gene modules.

GO enrichment analysis

The results showed that the turquoise module was enriched in pathways related to ADP metabolism (GO:0046031), ATP metabolism (GO:0046034), and oxidoreductase activity (GO:0016860, GO:0016864, GO:0016861). On the other hand, the blue module was enriched in pathways related to antioxidant activity genes (GO:0016209), hydrogen peroxide metabolism (GO:0042743, GO:0042744, GO:0042542, GO:0004096), synthesis and metabolism of phenolic compounds (GO:0018958, GO:0046189), synthesis and metabolism of flavonoids (GO:0009812, GO:0009813), and metabolism of L-phenylalanine (GO:0006558, GO:0006559) (Table 1). Overall, the results suggested that both gene modules were highly associated with antioxidant properties and showed some relationship with energy metabolism pathways. These findings indicated that the WGCNA method could effectively construct coexpression modules of genes related to antioxidant properties after grain filling, and these two modules were the focus of further investigation.

Module	GO ID	Description	Ratio_in_study	Ratio_in_pop	P value
Turquoise	GO:0046031	ADP metabolic process	26/4810	71/42,651	4.35E-07
Turquoise	GO:0046034	ATP metabolic process	39/4810	125/42,651	1.05E-06
Turquoise	GO:0016860	Intramolecular oxidoreductase activity	19/4810	57/42,651	8.53E-06
Turquoise	GO:0016864	Intramolecular oxidoreductase activity, transposing S-S bonds	7/4810	14/42,651	0.000385235
Turquoise	GO:0016861	Intramolecular oxidoreductase activity, interconverting aldoses and ketoses	6/4810	24/42,651	0.046208292
Blue	GO:0016209	Antioxidant activity	43/3960	301/42,651	0.004957872
Blue	GO:0042743	Hydrogen peroxide metabolic process	28/3960	202/42,651	0.028729882
Blue	GO:0042744	Hydrogen peroxide catabolic process	28/3960	201/42,651	0.02794752
Blue	GO:0018958	Phenol-containing compound metabolic process	11/3960	28/42,651	2.09186E-05
Blue	GO:0009812	Flavonoid metabolic process	10/3960	23/42,651	1.72741E-05
Blue	GO:0009813	Flavonoid biosynthetic process	10/3960	23/42,651	1.72741E-05
Blue	GO:0042542	Response to hydrogen peroxide	9/3960	27/42,651	0.000503918
Blue	GO:0006558	L-phenylalanine metabolic process	8/3960	36/42,651	0.015604145
Blue	GO:0006559	L-phenylalanine catabolic process	5/3960	16/42,651	0.012556104
Blue	GO:0046189	Phenol-containing compound biosynthetic process	4/3960	9/42,651	0.006376417
Blue	GO:0004096	Catalase activity	3/3960	5/42,651	0.006926473
Blue	GO:0046283	Anthocyanin-containing compound metabolic process	3/3960	6/42,651	0.012898615
Blue	GO:0009718	Anthocyanin-containing compound biosynthetic process	3/3960	6/42,651	0.012898615

Table 1. GO enrichment analysis of target modules.

Interactions between core gene networks in the target modules

To obtain the core genes in the two modules discussed above, the gene regulatory networks were visualized and processed using Cytoscape software to screen for highly connected genes in the modules, which were then identified as the core genes in each module (Fig. 3). In these networks, each node represents a gene, and nodes

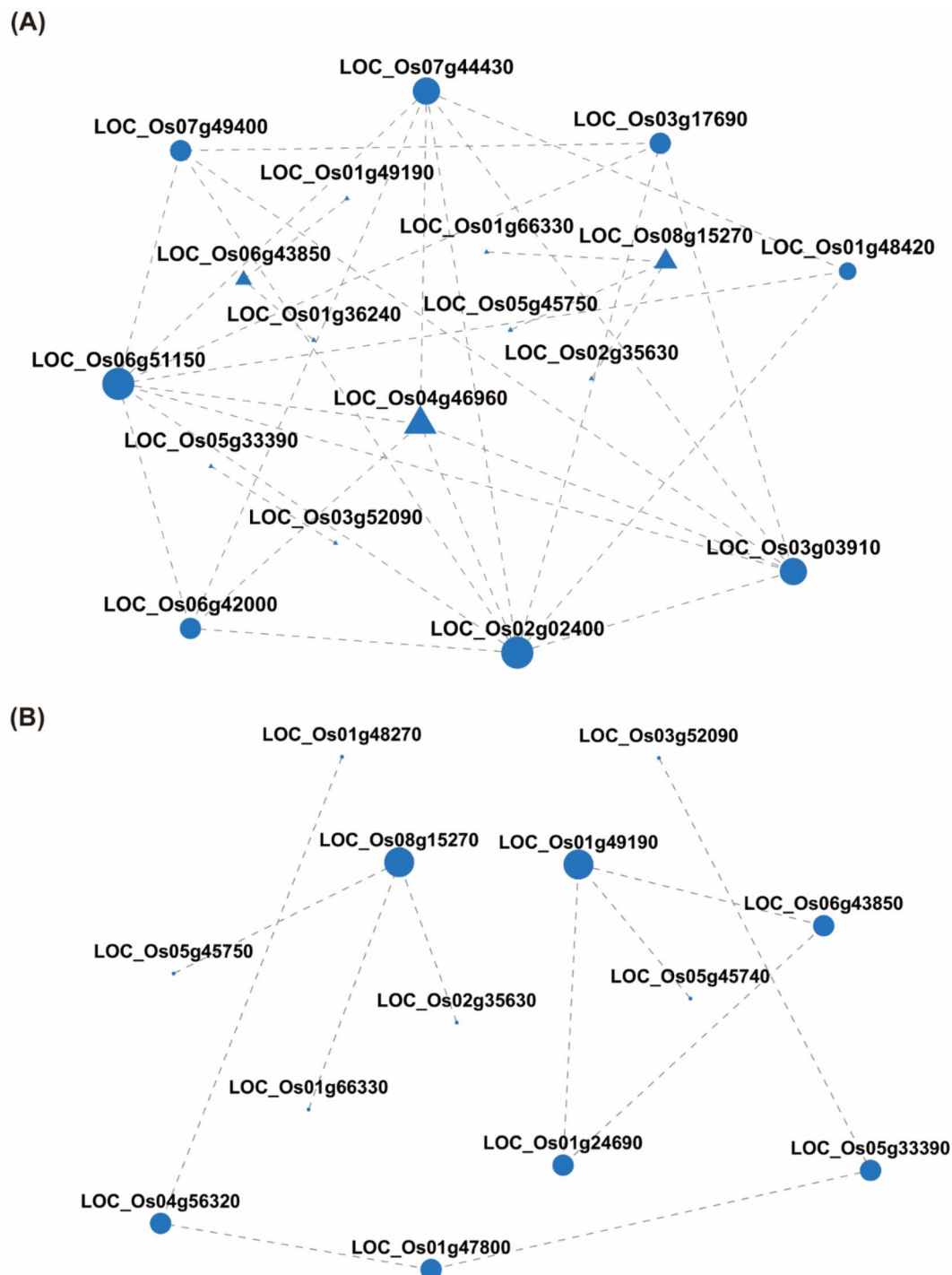


Fig. 3. Gene coexpression network and hub genes in blue and turquoise modules. (A) Blue module. (B) Turquoise module. Genes with the same shape (circle or triangle) represent genes belonging to the same functional class. The size of a node is proportional to its connectivity, meaning that nodes with more edges connected to them will be larger. This reflects the significance of a gene in the network – the more connections it has, the larger the node, indicating its greater importance within the network.

are connected by lines, where genes at either end of a line are typically considered to have similar biological functions. In the blue module, 31 core genes were identified (Fig. 3A and Table S13), among which 28 were related to catalase, 2 were related to phenylalanine, and 1 was related to flavonoids. In the turquoise module, 22 core genes were identified (Fig. 3B and Table S14), all of which were related to energy metabolism. The screened core genes were functionally involved in antioxidant properties and energy metabolism processes.

qRT-PCR analysis of 8 core genes

We selected 8 core genes from the turquoise and blue modules that were associated with antioxidant properties in purple rice grains for quantitative fluorescence PCR validation. The results of the validation showed that the changes in expression trends were consistent with the transcriptome sequencing results (Table S1 and Fig. 4), further confirming the reliability of the transcriptome sequencing results.

Association analysis of the transcriptome and metabolome

In our previous work, we conducted a comprehensive analysis of metabolites in each sample¹⁶. To further understand the relationship between differentially expressed genes and related metabolites in purple rice grains, the differentially expressed genes were annotated and analysed in the KEGG (Kyoto Encyclopedia of Genes and Genomes) Pathway database, as shown in Fig. 5A. We found that the differentially expressed genes were significantly enriched in glycolysis/gluconeogenesis, starch and sucrose metabolism, flavone and flavonol biosynthesis, and flavonoid biosynthesis pathways. These pathways could be broadly classified into flavonoid

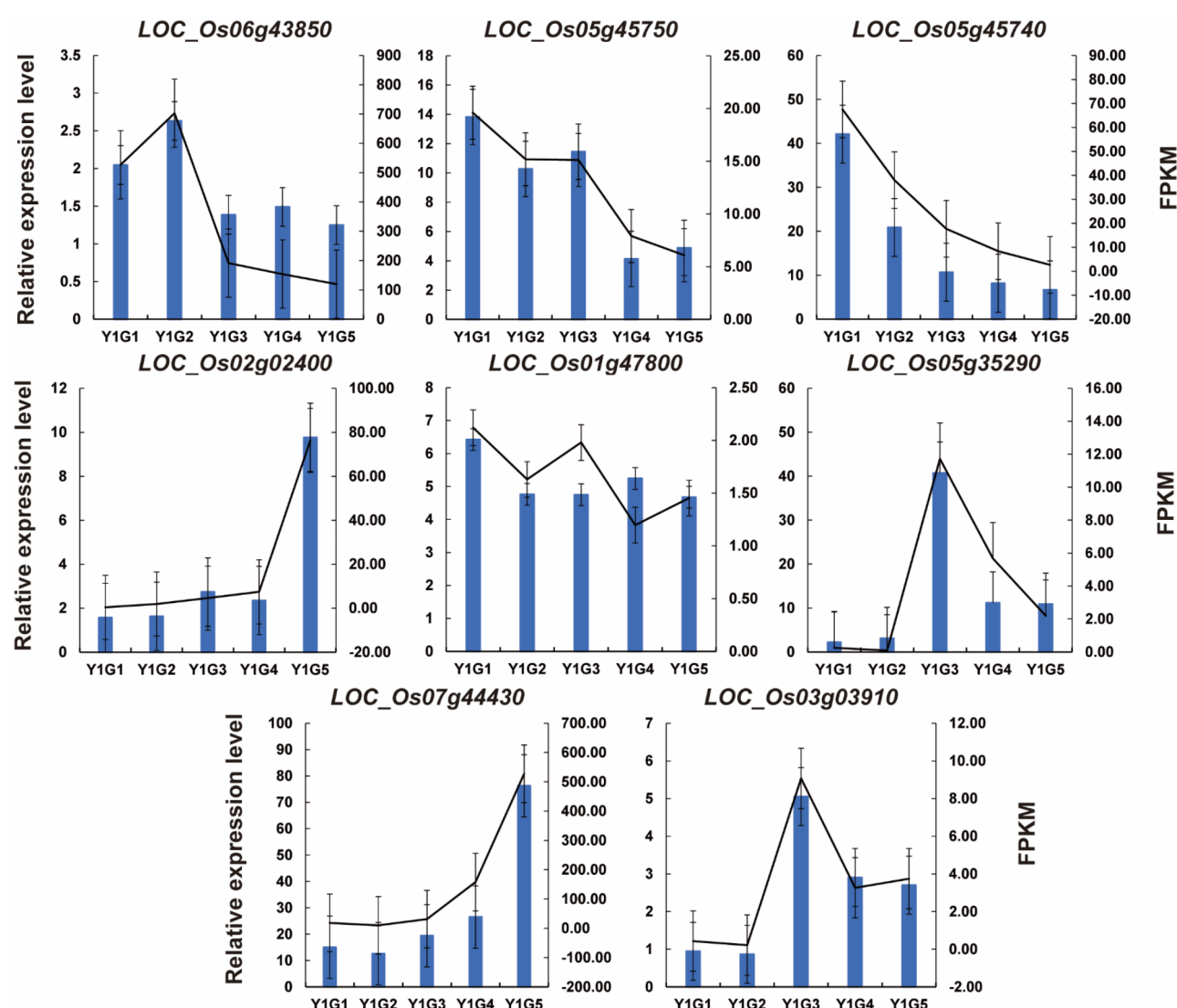


Fig. 4. Expression analysis was performed by qRT-PCR. The horizontal axis represents the different time points, the left vertical axis represents the relative expression levels of the genes, shown as bars, and the right vertical axis represents fragments per kilobase million values, shown as lines.

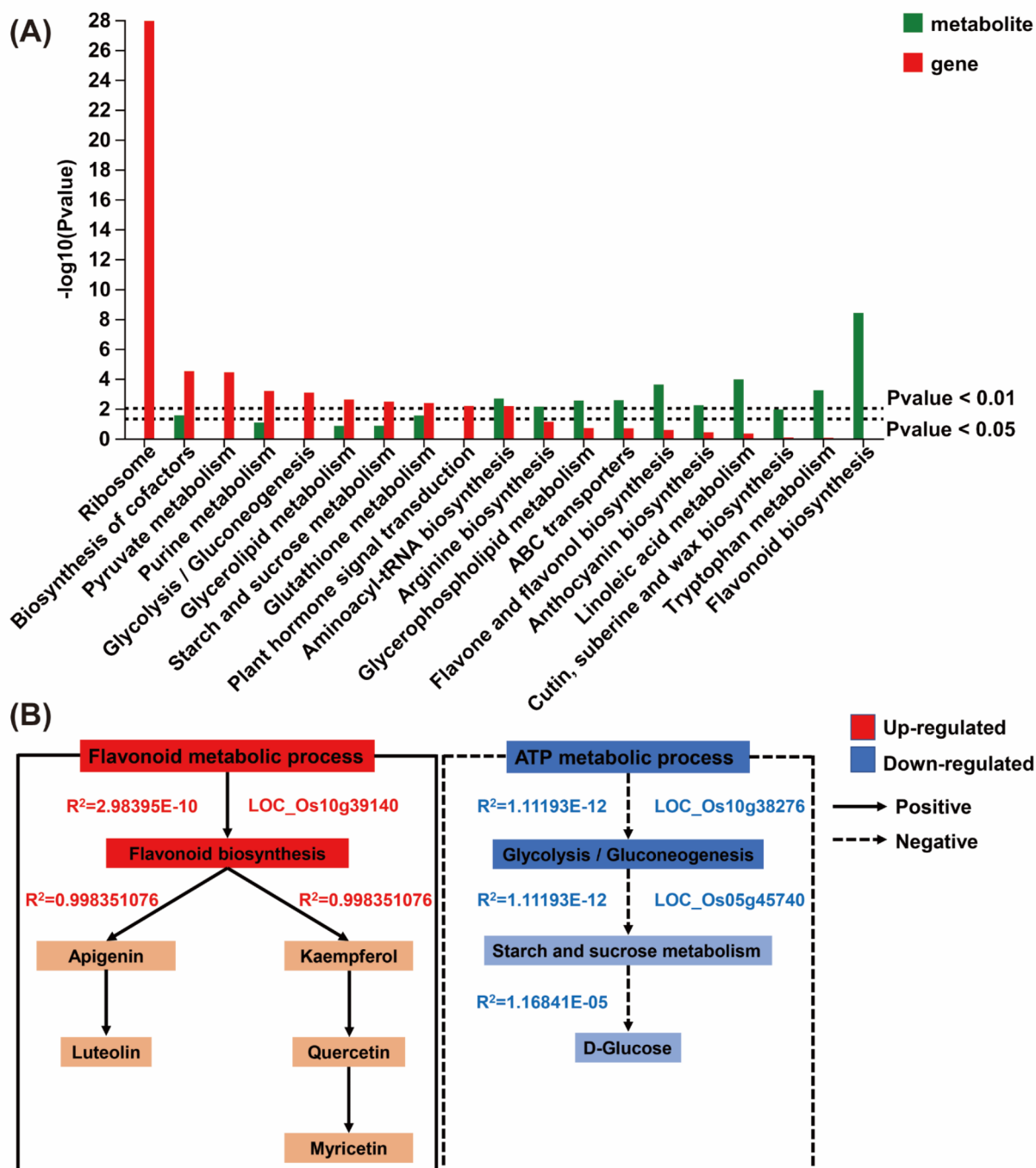


Fig. 5. Pathway analysis. **(A)** KEGG pathway enrichment statistical graph. **(B)** Metabolic pathways of flavonoid metabolic process and ATP metabolic process. The coefficient R^2 is a measure of the goodness of fit of a regression equation. It indicates how well the regression equation fits the data points. Generally, a higher R^2 value closer to 1 indicates a better fit of the regression equation to the data points.

metabolic processes and ATP metabolic processes. One core gene (*LOC_Os10g39140*) was annotated to flavonoid biosynthesis, 1 core gene (*LOC_Os10g38276*) was annotated to glycolysis/gluconeogenesis, and 1 core gene (*LOC_Os05g45740*) was annotated to starch and sucrose metabolism. These results further confirmed the high correlation between core genes and antioxidant properties as well as energy metabolism processes.

Discussion

High-antioxidant rice shows a wide range of biological functions in promoting human health⁹. Coloured rice generally exhibits higher antioxidant activity than nonpigmented rice due to higher contents of anthocyanins, phenolics, and flavonoids⁹. These three compounds play a crucial role in the synthesis of C=C aromatic groups (1740–1710 cm⁻¹) and C=C (stretching) alkanes (1640–1500 cm⁻¹), promoting the formation of organic compounds in coloured rice²⁴. Phenylalanine is a precursor of catechin and proanthocyanidin in the biosynthesis of phenylpropanoid compounds, and research has shown that coloured rice contains more phenylalanine¹⁸. In this study, the blue module was enriched in pathways related to the synthesis or degradation of phenylalanine, and WGCNA identified core genes related to phenylalanine in the blue module (Table 1; Fig. 3A). Phenolics have been reported to be the major hydrophilic antioxidants in rice and serve as an important indicator for measuring the antioxidant capacity of offspring samples in hybrid breeding. Rice with coloured husks generally contains higher levels of phenolic compounds and exhibits stronger antioxidant activity²⁵. In this study, the blue module was enriched in pathways related to the synthesis and metabolism of phenolic compounds (Table 1), indicating that purple rice grains are rich in phenolic compounds. Flavonoids are a group of polyphenolic compounds synthesized by the shikimic and malonic acid pathways that are present in higher amounts in coloured rice²⁶. As secondary metabolites, flavonoids exert a regulatory effect on the composition and function of endophytic fungal communities in black rice²⁷. Flavonoid biosynthesis is an important metabolic pathway during grain filling in rice²⁸. In this study, the blue module was enriched in pathways related to the synthesis and metabolism of flavonoids (Table 1), and 1 core gene related to flavonoids was also identified (Fig. 3A). Flavonoid biosynthesis was another metabolic pathway that was highly correlated with differentially expressed genes (Fig. 5A, B). Catalase (CAT) is a core antioxidant enzyme in most organisms and can catalyse the decomposition of hydrogen peroxide (H₂O₂), thereby controlling the abundance of this essential cell signalling molecule. CAT is the most abundant protein in plant peroxisomes and shows one of the highest catalytic rates known in biology²⁹. Our WGCNA identified core genes related to CAT in the blue module (Fig. 3A). In addition, studies have shown that black rice contains higher levels of amino acids and functions in energy metabolism, enriching its nutritional value over that of ordinary rice³⁰. In this study, 22 core genes related to energy metabolism were identified in the turquoise module (Fig. 3B), and the correlation analysis of the transcriptome and metabolome also indicated a strong correlation between energy metabolism and differentially expressed genes (Fig. 5A, B).

In this study, we used WGCNA to identify gene modules and core genes that were highly associated with antioxidant properties, revealing the molecular mechanisms underlying the antioxidant properties of rice grains after the filling stage. We focused on the turquoise and blue modules, which showed high correlations, although other gene modules that were not discussed in detail here may also contain pathways related to antioxidant properties and warrant further investigation of their biological significance.

Conclusion

In this study, we constructed a weighted gene coexpression network to perform an in-depth analysis of the two gene modules with the highest associations with the target trait. As a result, we identified representative core genes (*LOC_Os10g39140*, *LOC_Os10g38276*, and *LOC_Os05g45740*). We found that these core genes were involved in pathways related to antioxidant activity and energy metabolism, such as flavonoid biosynthesis and glycolysis/gluconeogenesis. These results provide clues about the molecular mechanism underlying the antioxidant properties of purple rice grains and offer theoretical support for the breeding of high-antioxidant functional rice.

Data availability

The data supporting the findings of this study are available from the corresponding authors upon request. The raw sequences data was uploaded to National Genomics Data Center (<https://ngdc.cnca.ac.cn/>). The accession numbers are uploaded and archived at NGDC with accession CRA010137.

Received: 12 July 2024; Accepted: 19 September 2024

Published online: 30 September 2024

References

- Xiong, Q. et al. Metabolites associated with the main nutrients in two varieties of purple rice processed to polished rice. *Metabolites* **13**, 7 (2023).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
- Tan, M. et al. Co-expression network analysis of the transcriptomes of rice roots exposed to various cadmium stresses reveals universal cadmium-responsive genes. *BMC Plant Biol.* **17**, 194 (2007).
- Zou, X. et al. Co-expression network analysis and hub gene selection for high-quality fiber in upland cotton (*Gossypium hirsutum*) using RNA sequencing analysis. *Genes* **10**, 119 (2019).
- Yang, Y., Sang, Z., Xu, C., Dai, W. & Zou, C. Identification of maize flowering gene co-expression modules by WGCNA. *Acta Agron. Sin.* **45**, 161–174 (2019).
- Hussain, I. et al. Alleviating effects of salicylic acid spray on stage-based growth and antioxidative defense system in two drought-stressed rice (*Oryza sativa* L.) cultivars. *Turk. J. Agric. For.* **47**, 79–99 (2023).
- Lu, H. et al. Superior antioxidant capacity and Auxin Production promote seedling formation of Rice seeds under submergence stress. *Agronomy* **13**, 171 (2023).
- Zhang, Y. et al. Foliar methyl jasmonate (MeJA) application increased 2-acetyl-1-Pyrroline (2-AP) content and modulated antioxidant attributes and yield formation in fragrant rice. *J. Plant. Physiol.* **1**, 153946 (2023).
- Fatchiyah, F. et al. Bioactive compound profile and their biological activities of endogenous black rice from Java and East Nusa Tenggara. *Cyta-J Food* **21**, 159–170 (2023).
- Li, P. et al. Genes and their molecular functions determining seed structure, components, and quality of rice. *Rice* **15**, 1–27 (2022).

11. Xiong, Q. et al. Biomarkers of the main nutritional components in purple rice during five successive grain filling stages. *Food Chem. X* **1**, 101528 (2024).
12. Xiong, Q. et al. Metabolomics and biochemical analyses revealed metabolites important for the antioxidant properties of purple glutinous rice. *Food Chem.* **389**, 133080 (2022).
13. Gunaratne, A. et al. Antioxidant activity and nutritional quality of traditional red-grained rice varieties containing proanthocyanidins. *Food Chem.* **138**, 1153–1161 (2013).
14. Finocchiaro, F. et al. Characterization of antioxidant compounds of red and white rice and changes in total antioxidant capacity during processing. *Mol. Nutr. Food Res.* **51**, 1006–1019 (2007).
15. Ghasemzadeh, A., Karbalaii, M., Jaafar, H. & Rahmat, A. Phytochemical constituents, antioxidant activity, and antiproliferative properties of black, red, and brown rice bran. *Chem. Cent. J.* **12**, 1–13 (2018).
16. Xiong, Q. et al. Metabolomics revealed metabolite biomarkers of antioxidant properties and flavonoid metabolite accumulation in purple rice after grain filling. *Food Chem. X* **18**, 100720 (2023).
17. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
18. Kim, D., Langmead, B. & Salzberg, S. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
19. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
20. Li, B. & Dewey, C. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
21. Wang, L., Feng, Z., Wang, X., Wang, X. W. & Zhang, X. DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2009).
22. Xie, C. et al. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322 (2011).
23. Livak, K. & Schmittgen, T. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
24. Moko, E. & Rahardiyana, D. Structure of stigmaterols in bran of red rice from Minahasa Regency, North Sulawesi, Indonesia. *Fuller. J. Chem.* **5**, 16–22 (2020).
25. Zhang, H., Shao, Y., Bao, J. & Beta, T. Phenolic compounds and antioxidant properties of breeding lines between the white and black rice. *Food Chem.* **172**, 630–639 (2015).
26. Pradipta, S., Siswoyo, T. & Ubaidillah, M. Nutraceuticals and bioactive properties of local Java pigmented rice. *Biodiversitas J. Biol. Divers.* **24**, 1 (2023).
27. Solanki, S. et al. Co-application of Silver nanoparticles and Symbiotic Fungus *Piriformospora indica* improves secondary Metabolite production in Black Rice. *J. Fungi* **9**, 260 (2023).
28. Wang, T. et al. Large-scale metabolome analysis reveals dynamic changes of metabolites during foxtail millet grain filling. *Food Res. Int.* **165**, 112516 (2023).
29. Baker, A. et al. A critical node in the regulation of cell fate. *Free Radic. Biol. Med.* **199**, 56–66 (2023).
30. Fatchiyah, F., Sari, D., Safitri, A. & Cairns, J. Phytochemical compound and nutritional value in black rice from Java Island, Indonesia. *Syst. Rev. Pharm.* **11**, 414–421 (2020).

Author contributions

Qiangqiang Xiong, Xiaobin Liu: Conceptualization, Junkai Zhu: methodology, Ying Zhu: software, Jinxia Cao, Rui Ding: formal analysis, Yanqiong Zhang: investigation, Huijuan Zhou: resources, Yangang Zhu, Yiming Ji: data curation. Changhui Sun: writing—original draft preparation, Jiao Zhang: writing—review and editing.

Funding

This research was funded by the Breeding of new conventional medium-grain *japonica* rice varieties for integrated rice cultivation and breeding (YZ2023051), Open Project of Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University (202303) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-73698-w>.

Correspondence and requests for materials should be addressed to Q.X. or X.L.

Reprints and permissions information is available at www.nature.com/reprints.

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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024