

# Unique Transcriptome And Gene Expression Analysis of Flag Leaves of A Super-Hybrid Rice WFYT025

**Qin Cheng**

Ministry of Education, Jiangxi Agricultural University

**Shiying Huang**

Ministry of Education, Jiangxi Agricultural University

**Peng Wang**

Ministry of Education, Jiangxi Agricultural University

**Shilei Liu**

Ministry of Education, Jiangxi Agricultural University

**Guangliang Wu**

Ministry of Education, Jiangxi Agricultural University

**Yue Li**

Ministry of Education, Jiangxi Agricultural University

**Lan Lin**

Ministry of Education, Jiangxi Agricultural University

**Yanning Wang**

Ministry of Education, Jiangxi Agricultural University

**Qi Zhong**

Ministry of Education, Jiangxi Agricultural University

**Tao Huang**

Ministry of Education, Jiangxi Agricultural University

**Mengmeng Yang**

Ministry of Education, Jiangxi Agricultural University

**Xinyue He**

Ministry of Education, Jiangxi Agricultural University

**Haohua He**

Ministry of Education, Jiangxi Agricultural University

**Jianmin Bian** (✉ [jmbian81@126.com](mailto:jmbian81@126.com))

Ministry of Education, Jiangxi Agricultural University

**Keywords:** Rice, Super-hybrid Rice, Heterosis, Grain number, Flag leaf, RNA-seq

**Posted Date:** June 16th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-490906/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** The health and physiology of flag leaves are closely related to rice yield, and flag leaves play an important role in providing photosynthetic products during grain filling, many breeding studies have tried to improve the performance of flag leaves. However, there are few studies on the heterosis of rice flag leaves up to now.

**Results:** Thus, the present research is focused on the flag leaves heterosis of a widely used late-cropping *indica* super hybrid rice combination WFYT025 in China using a high-throughput next-generation RNA-seq strategy under different environment with two stages, trying to find some genes related to photosynthesis, transpiration, and development of seeds. Mid-parent heterosis (MPH) and higher parent heterosis (HPH) were estimated for the heterosis of flag leaf. Under the environment of middle rice, the number of genes up-regulated in CHT025, WFB and WFYT025 were 892,1,273 and 819, down-regulated in CHT025, WFB and WFYT025 were 616,1934 and 2196, respectively. Among the SDGhps on the first day after flowering, 10.9% had a dominant effect, 41.81% had a partial dominant effect, 22.07% had an additive effect and the remaining 25.22% had an over-dominant effect. Meanwhile, on the tenth day after flowering, there were 491 genes, accounting for 27.16%, showed over-dominance; 222 genes, accounting for 12.28%, showed dominance; 760 genes, accounting for 42.04%, showed partial dominance; and 335 genes, accounting for 18.52%, showed additive effect.

**Conclusion:** The co-expressed gene sets via weighted gene co-expression network analysis (WGCNA) were identified, and total of 5,000 highly expressed genes were divided into 24 co-expression groups. In the two stages, we found 9 identical transcription factors. Except for 5 reported TFs, the other 4 TFs may play an important role in grain number and photosynthesis heterosis.

## Background

As a principal organ in rice, leaves are involved in many fundamental physiological functions such as photosynthesis and transpiration [1]. Among the rice and other major cereals, the uppermost three leaves, especially the flag leaf, are the main source of the carbohydrates that eventually accumulate in the grains [2, 3, 4, 5]. The top three leaves on the stem are the primary source of carbohydrate production. Particularly, the flag leaves produce over 50% of the carbohydrates that are accumulated into grains [6], which finally influences grain yield. The phenotype of the flag leaves, in particular its size and angle, is regarded as the major determinant of plant architecture and strongly affects high-yield performance, such as 1,000-grain weight [7, 8].

Previous studies on rice leaf phenotype mainly focused on the location of quantitative trait loci (QTL). Few genes related to leaf phenotype were identified or cloned. Longitudinal cell division, cell elongation and cell arrangement determine rice leaf elongation. For example, the dwarf mutant dwarf and *gladius leaf 1 (dgl1)* have shorter leaves and rounded tips compared to wild type (WT). The *dgl1* phenotype was abnormal due to poor elongation of longitudinal epidermal cells and coarse cell deformation [9]. The

variation of leaf width in rice was mainly caused by changes in vascular bundles and cell division. The *NARROW LEAF 1* (*NAL1*) mutant has narrower leaves than the WT, and its abnormal phenotype is caused by the decrease in the number of longitudinal veins. *NAL1* affects vascular morphology and polar auxin transport in rice, and plays an important role in controlling lateral leaf growth [10]. In *NAL2/3* double mutant, narrowed curly leaves are caused by less longitudinal veins and reduced lateral axial branch. *NAL2* and *NAL3* are paralogs, encoding the same *WUSCHEL-RELATED HOMEODOMAIN 3A* (*OsWOX3A*; *OsNS*) transcriptional activator, which is related to organ development, tiller development, and so on [11]. The leaf width phenotype of the *NAL1* mutant is reduced, and microscopic analysis shows that compared with WT, the leaves of the *NAL1* mutant have fewer longitudinal veins [13]. The *ABNORMAL VASCULAR BUNDLES* (*AVB*) mutant has narrower leaves than the wild type. The phenotypic abnormality of *AVB* leaves is caused by a decrease in the vascular bundles of gas producing organs [14].

High-throughput RNA sequencing has been used to search for heterosis in rice to avoid defects of methods with low throughput, high cost, low sensitivity, clonal preference, and high background noise. Several studies have investigated gene transcript abundance during leaf senescence in crops such as cotton and maize [15, 16] and in a range tissues during grain filling in rice [17], including flag leaf tissue, and transcript data are publicly available. However, there are few studies on the heterosis of rice flag leaves and the mechanism of the flag leaves to improve yields is not well understood. It is very necessary to study the heterosis of rice flag leaves. In this study, we focused on a super rice variety, WFYT025, which is widely used in China. Therefore, we took the flag leaves of WFYT025 and its two parents for high-throughput transcriptome sequencing, trying to find some genes related to photosynthesis or transpiration and development of seeds. The aim of our study is (1) we investigated the length, width, and leaf area of flag leaves for WFYT025 and its parents. MPH and HPH were estimated for the heterosis of flag leaf. (2) We selected the flag leaves of hybrid rice WFYT025 and its parents for transcriptome sequencing, the first day and the tenth days after flowering in the environment of early rice and middle rice, respectively. (3) we referred to further investigate the gene regulatory network. Several major sub-networks were found to represent interactions among genes with similar expression profiles via gene regulatory network analysis.

## Methods

### Plant materials and growth conditions

The hybrid WFYT025 along with its parental lines Changhui T025 (CHT025) and Wufeng B (WFB) were planted in the experimental field of Jiangxi Agricultural University. WFYT025 is a super-hybrid rice combination derived from the cross between female parent WFB and male parent CHT025. WFYT025 and the two parents were sown at the experimental plot in Jiangxi Agricultural University in a completely randomized block design with three replications in summer 2019. Each plot consisted of 40 rows, with each row consisting of 10 plants, each separated from its neighbour by 20 cm. Crop management followed normal procedures for rice. These three lines were selected in this study to measure phenotypic traits and conduct transcriptome analyses [18]. We took the grains and sword leaves from the first day,

the fifth day, the tenth day, the fifteenth day and the twenty-first day after flowering of rice. By measuring the 1000 grain weight of the grains, we finally selected the sword leaves from the first day and the tenth day for sequencing. The flag leaves were collected and stored at  $-80^{\circ}\text{C}$  for RNA-Seq analysis, and each sample had at least three biological replications to minimize systematic errors.

### **Pathway enrichment analysis**

KEGG (Analysis of Kyoto Encyclopedia of Genes and Genomes) is the major public pathway-related database. Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in DEGs (different expression genes) comparing with the whole genome background [19]. Pathway enrichment analysis was performed using the OmicShare tools, a free online platform for data analysis ([www.omicshare.com/tools](http://www.omicshare.com/tools)). Significantly enriched pathways in DEGs comparing to the genome background were defined by hypergeometric test. The calculated p-value was gone through FDR Correction, taking  $\text{FDR} \leq 0.05$  as a threshold. Pathways meeting this condition were defined as significantly enriched pathways in DEGs.

### **GO enrichment analysis**

Gene Ontology (GO) is an international standardized gene functional classification system which offers a dynamic-updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component, and biological process. The basic unit of GO is GO-term. Each GO-term belongs to a type of ontology.

GO enrichment analysis provides all GO terms that significantly enriched in DEGs comparing to the genome background, filter the DEGs that correspond to biological functions. GO enrichment analysis was performed using the OmicShare tools, a free online platform for data analysis ([www.omicshare.com/tools](http://www.omicshare.com/tools)). Firstly all DEGs were mapped to GO terms in the Gene Ontology database (<http://www.geneontology.org/>), gene numbers were calculated for every term, significantly enriched GO terms in DEGs comparing to the genome background were defined by hypergeometric test [20]. The calculated p-value was gone through FDR Correction, taking  $\text{FDR} \leq 0.05$  as a threshold. GO terms meeting this condition were defined as significantly enriched GO terms in DEGs. This analysis was able to recognize the main biological functions that DEGs exercise.

### **Quantitative real-time PCR (qRT-PCR) validation**

To validate the RNA-seq results, different expression patterns of several genes were confirmed by quantitative real-time RT-PCR (qRT-PCR) [21]. For qRT-PCR, 1  $\mu\text{g}$  of total RNA was used to synthesized cDNA using PrimeScript™ RT reagent Kit (Perfect Real Time) (TaKaRa). The qRT-PCR was carried out using SYBR® Premix Ex Taq II (Tli RNaseH Plus; TAKARA BIO Inc., Shiga, Japan) and determined in LightCycler 480 (Roche, Basel, Switzerland) according to the manufacturer's instructions. The qRT-PCR reactions were amplified for  $95^{\circ}\text{C}$  for 30 s, followed by 40 cycles of  $95^{\circ}\text{C}$  for 5 s,  $55^{\circ}\text{C}$  for 30 s and  $72^{\circ}\text{C}$  for 30 s. All reactions were performed with three independent biological replicates for each sample and

three technical replicates for each biological replicate were analyzed. The relative gene expression was calculated by the software of ABI7500 Real-Time PCR System using the  $2^{-\Delta\Delta Ct}$  method.

### **The mode of inheritance analysis**

For statistical analysis, the analysis of variance (ANOVA) was usually by the model:  $y = u + (GA) + (GD) + (SR) + e$ , where  $y$  is the acquired gene expression,  $u$  is the overall mean,  $GA$  is the additive effect,  $GD$  is the dominant effect,  $SR$  is the replication effect, and  $e$  is the residual error [18].  $H_p = [d]/[a]$ , referred to as the dominance ratio or potency (where  $[a]$  and  $[d]$  represent  $GA$  and  $GD$ , respectively), was also calculated to measure the non-additivity of the F1 hybrid relative to its parents [22]. Considering gene expression levels as quantitative traits, we adopted traditional quantitative genetic parameters, such as composite additive effect  $[a]$  and composite dominance effect  $[d]$ , to estimate our expression profile. DGHP were classified according to the dominance ratio  $H_p (= [d]/[a])$ , based on 99.8% confidence intervals constructed for  $[d] - [a]$  ( $[d] > 0$ ) and  $[d] + [a]$  ( $[d] < 0$ ). According to the value of  $H_p (= [d]/[a])$ , we considered that these genes belonged to partial dominance ( $-0.8 < H_p \leq -0.2$  or  $0.2 < H_p \leq 0.8$ ), over-dominance ( $H_p \leq -1.2$  or  $H_p > 1.2$ ), dominance ( $-1.2 < H_p \leq -0.8$  or  $0.8 < H_p \leq 1.2$ ) and additive effect ( $-0.2 < H_p \leq 0.2$ ) [23, 24].

### **Normalization of gene expression levels and identification of differentially expressed genes**

Sequencing reads were mapped to the reference sequences. The expression level of each gene was measured by fragments per kilobase of exon model per Million mapped reads (FPKM). To determine the time-dependent transcriptional differences between early rice and middle rice, the differential expression genes (DEGs) at 1 and 10 days after anthesis were determined by comparing the expression levels. To correct for multiple testing, the false discovery rate (FDR) was calculated to adjust the threshold of  $p$ -value [25]. The standard of different expression between compare is a minimal 2-fold difference in expression ( $|\log_2 \text{Ratio}| \geq 1$ ) and a  $FDR \leq 0.005$  [23].

### **Weighted gene co-expression network analysis (WGCNA)**

The gene co-expression network analysis used the raw expression counts (RAW counts) of all Rice reference genes in the RNA-seq data of the 36 leaf groups obtained from the previous analysis. Firstly, the genes with low expression were eliminated. The elimination standard was that if the genes were not expressed in more than 80% of the samples, the genes with low expression were considered. Then select median absolute deviation from the genes with high expression (mad) 5000 genes of maximum value for total express network analysis using WGCNA (<https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/>). In the process of WGCNA analysis, 20 soft power was selected. GO enrichment analysis through the agriGOv2 (<http://systemsbiology.cau.edu.cn/agriGOv2/>), its parameters as follows: Statistical Test Method as HyperGeometry, Multi\_Test Adjustment Method as Hochberg FDR, Significance level as "0.05" and Minimum number of mapping entries as "5".

## **Results**

## Phenotype analysis for WFYT025 and its parents

In this study, we investigated the length, width, and leaf area of flag leaves for WFYT025 and its parents (Table 1). In addition, the 1,000-grain weight of WFYT025 and its two parents on the first day and the tenth days after flowering were measured (Table 1). In the environment of early rice, the 1,000-grain weight of WFYT025 was 7.30g, WFB was 6.70g, and CHT025 was only 5.68g on the first day. On the tenth day after flowering, of WFYT025 was 20.42g, WFB was 24.86g, and CHT025 was only 14.77g. As the same time, the fresh weight of WFB and its progeny WFYT025 was significantly higher than that of CHT025 both on the first and tenth day after flowering.

Table 1  
Phenotypic Analysis of WFYT025 and its parents

| Environment | Traits                          | WFB   | CHT025 | WFYT025 | MPH (%) | HPH (%) |
|-------------|---------------------------------|-------|--------|---------|---------|---------|
| early rice  | 1D leaf length(cm)              | 28.17 | 25.20  | 30.40   | 13.9    | 7.9     |
|             | 1D leaf width(cm)               | 2.02  | 2.18   | 2.38    | 13.3    | 9.1     |
|             | 1D leaf area(cm <sup>2</sup> )  | 42.68 | 41.20  | 54.26   | 29.3    | 27.1    |
|             | 1D TKW(g)                       | 6.70  | 5.68   | 7.30    | 17.9    | 8.9     |
|             | 10D leaf length(cm)             | 31.80 | 30.15  | 30.18   | -2.6    | -5      |
|             | 10D leaf width(cm)              | 2.07  | 2.33   | 2.35    | 6.8     | 0.8     |
|             | 10D leaf area(cm <sup>2</sup> ) | 49.37 | 52.69  | 53.19   | 4.2     | 0.1     |
|             | 10D TKW(g)                      | 24.86 | 14.77  | 20.42   | 2.81    | -17.8   |
| middle rice | 1D leaf length(cm)              | 27.00 | 38.50  | 44.10   | 34.7    | 14.5    |
|             | 1D leaf width(cm)               | 1.93  | 2.00   | 2.10    | 6.9     | 5       |
|             | 1D leaf area(cm <sup>2</sup> )  | 39.08 | 57.75  | 69.46   | 43.4    | 20.2    |
|             | 1D TKW(g)                       | 5.03  | 4.25   | 4.84    | 4.3     | -3.7    |
|             | 10D leaf length(cm)             | 31.37 | 40.78  | 43.68   | 21.1    | 7.1     |
|             | 10D leaf width(cm)              | 2.11  | 2.22   | 2.28    | 5.3     | 2.7     |
|             | 10D leaf area(cm <sup>2</sup> ) | 49.64 | 67.90  | 74.69   | 27.1    | 10      |
|             | 10D TKW(g)                      | 29.06 | 14.64  | 21.10   | -3.4    | -27.3   |

MPH and HPH were estimated for the heterosis of flag leaf (Table 1). It indicated that MPH showed negative effect on the tenth days leaf length in the environment of early rice, and the MPH values for all the traits varied from 5.3 to 43.4%. At the same time, HPH on the first day leaf area remained highly significant both in the environment of middle rice and early rice.

## **RNA sequencing of WFYT025 and its parents**

To further explore the possible genes and mechanisms involved in the heterosis of WFYT025. We selected the flag leaves of hybrid rice WFYT025 and its parents for transcriptome sequencing, the first day and the tenth days after flowering in the environment of early rice and middle rice, respectively.

As a result, a total of 1,725.9 million reads were generated, and an average of 47.94 million reads were generated for each sample. The total number of valid reads is 155,921 million, and the average valid read for each sample is 43.31 million (Supplementary Table S1). We randomly selected 12 genes using the specific primers (Supplementary Table S2), and validated their accuracy and reproducibility using qRT-PCR (Supplementary Fig. S1).

## **Analysis of differentially expressed genes**

To analyze genes expression in different stages under different environment, genes expression level was mainly measured by FPKM (Fragments Per Kilobase of exon model Per Million mapped reads) or RPKM (reads Per Kilobase of exon model Per Million mapped reads) (Fig. 1A). In the leaves of CHT025 under the early rice environment, the number of DEGs ( $p < 0.05$ ) on the first day and the tenth days were 3,072 in total, among which 1,937 were upregulated and 1,135 were downregulated. At the same time, the number of DEGs in WFB leaves in the first day and the tenth days were 5,039 in total, among which 3,244 were upregulated and 1,795 were downregulated. In the leaves of WFYT025, the number of DEGs in the first day and the tenth days were 3,757, among which 1,937 were upregulated and 1,135 were downregulated (Fig. 1B). Under the environment of middle rice, the number of genes up-regulated in CHT025, WFB and WFYT025 were 892, 1,273 and 819, respectively, down-regulated in CHT025, WFB and WFYT025 were 616, 1,934 and 2,196, respectively (Fig. 1C).

## **O enrichment analysis of hybrid rice and its parents**

GO enrichment analysis of these DEGs has identified several significant biological processes at different stages under different environment. In the environment of early rice, “chloroplast stroma” and “oxidation-reduction” process were significant enriched in CHT025 (Fig. 2A). Meanwhile, “closed chloroplast”, “ATP binding”, “cytosol” and “mitochondrion” were significant enriched in WFB (Fig. 2B). At the same time, “cytosol”, “chloroplast” and “oxidation-reduction” process were significant enriched in WFYT025 (Supplementary Fig. S2A). In the environment of middle rice, “protein serine/threonine kinase activity”, “protein phosphorylation”, “extracellular region” and “plasma membrane” were significant enriched in CHT025 (Supplementary Fig. S2B). “Closed protein serine/threonine kinase activity”, “ATP binding”, “plasma membrane” and “defense response” were significant enriched in WFB (Supplementary Fig. S2C). Meanwhile, “ATP binding”, “integral component of membrane”, “plasma membrane” and “protein serine/threonine kinase activity” were significant enriched in WFYT025 (Supplementary Fig. S2D). This suggests that the considerable differences in flag leaves between the first day and the tenth day, may be related to photosynthetic efficiency.

## **Analysis of KEGG enrichment of DEGs**

Functional enrichment analysis was performed for all these DEGs during rice seed development. In total, 131 pathways were identified in 1,502 DEGs (includes the first day and the tenth day of CHT025 in the environment of early rice. KEGG enrichment analysis showed that the top 20 most enrichment pathways included “ubiquinone”, “other terpenoid-quinone biosynthesis”, “glycerolipid metabolism”, “taurine”, “hypotaurine metabolism” and “tryptophan metabolism” and so on (Fig. 3A). Meanwhile, 2,663 DEGs in WFB were classified into 137 pathways, and the top 20 most enrichment pathways were mainly concentrated “starch”, “sucrose metabolism”, “fatty acid metabolism”, “nicotinate”, “nicotinamide metabolism” and “fatty acid elongation” and so on (Fig. 3B). 1,911 DEGs in WFYT025 were classified into 133 pathways, and the top 20 most enriched pathways were mainly concentrated “ribosome”, “purine metabolism”, “nitrogen metabolism” and “ribosome biogenesis in eukaryotes” and so on (Supplementary Fig. S3). While in the environment of middle Rice, “plant-pathogen interaction”, “cutin”, “suberine”, “wax biosynthesis”, “brassinosteroid biosynthesis” and “flavonoid biosynthesis” pathways were enriched between the first day and the tenth day in CHT025. However, “plant-pathogen interaction”, “diterpenoid biosynthesis”, “phenylalanine metabolism” and “flavonoid biosynthesis” were enriched in WFB. While “plant-pathogen interaction”, “amino sugar”, “nucleotide sugar metabolism”, “diterpenoid biosynthesis” and “AGE-RAGE signaling” pathway in diabetic complications were enriched in WFYT025.

### **Identification of gene co-expression modules of flag leaf of WFYT025**

After obtaining the gene expression profile of flag leaves, we referred to further investigate the gene regulatory network. Several major sub-networks were found to represents interactions among genes with similar expression profiles via gene regulatory network analysis, which were designated as co-expression modules [27]. The co-expressed gene sets via WGCNA were identified, and total of 5,000 highly expressed genes were divided into 24 co-expression groups (Fig. 4). After that, we further analyzed the correlation between the gene expression profile of the co-expression module and the three phenotypic data (TKW; length; width) (Supplementary Fig. S4). Since the aim is to find genes related to grain development, the co-expression module with high correlation with TKW (the color of the module is Midnightblue, and the gene expression profile in this module is moderately negatively correlated with the thousand-grain weight phenotype) was focused on. We found the overall gene expression profile of the co-expression module MidnightBlue, which had the highest correlation with TKW, and the co-expression module contained a total of 106 genes (Supplementary Fig. S5). The GO enrichment analysis of these genes revealed several significant biological processes. Modules associated with TKW showed biological processes that are associated to “protein amino acid phosphorylation” and “protein modification process” and so on (Fig. 5). Overall, these modules represent the specific gene regulatory processes at each stage and are the indicators of operated programs of flag leaves development.

## **Discussion**

In the cycle of plant life, leaves play a significant role in grain yield, flag leaves length has been recognized as an important factor that determines plant type for high-yield potential in rice [28]. However, most of the previous studies on rice yield focused on rice grains, and few focused on the effect of rice

flag leaves on rice yield. In our study, we investigated the relationship between transcriptional profiles and heterosis in super hybrid rice WFYT025 by RNA-Seq.

### **The genetic basis of heterosis**

SDG (significant difference gene) was defined as the significant difference gene between first days and tenth days after anthesis. SDG<sub>S</sub> (significant difference genes) were defined as the significant difference genes between first days and tenth days after anthesis. Meanwhile, SDG that existed in hybrid but not in parents were defined as SDG<sub>hp</sub>.

We have been able to identify a number of SDG<sub>hp</sub>s underlying flag leaves between hybrid WFYT025 and paternal line CHT025, confirming the suggestion that heterosis is a polygenic phenomenon [29]. Among the SDG<sub>hp</sub>s on the first day after flowering, 10.9% had a dominant effect, 41.81% had a partial dominant effect, 22.07% had an additive effect and the remaining 25.22% had an over-dominant effect. Meanwhile, on the tenth day after flowering, there were 491 genes, accounting for 27.16%, showed over-dominance; 222 genes, accounting for 12.28%, showed dominance; 760 genes, accounting for 42.04%, showed partial dominance; and 335 genes, accounting for 18.52%, showed additive effect. Thus, partially dominance was the major contributor to the flag leaf heterosis of WFYT025 (Fig. 6).

However, on tenth day after flowering, 43.19% of SDG<sub>hp</sub>s (between hybrid WFYT025 and paternal line CHT025) was partially dominance, 16.65% was dominance, 32.17% was over-dominance, and the remaining 7.99% was additive effect on the first day after anthesis. In contrast, on the 10th day after anthesis, 52.83% of the genes were super dominant, 26.54% partial dominant, 9.56% dominant and 11.07% additive. It may be caused by too high temperature, which is inconsistent with the results of other environments (Fig. 7).

### **Transcription factors probably underlying heterosis**

It is well known that the analysis of expression patterns of important transcription factors is helpful to the study of heterosis in rice. In CHT025, we found 13 transcription factors of WRKY family in SDG<sub>hp</sub>s under the environment of early rice and 16 genes of WRKY family in SDG<sub>hp</sub>s of under the environment of middle rice. Transcription factors including *WRKY3*, *WRKY68* and *WRKY77* were found in both environments. Recently, a research reported that WRKY family genes are involved in the process of seed growth and maturing in Arabidopsis and rice [30]. However, the expression level of *WRKY68* in WFYT025 and its parents in the tenth days after flowering was significantly higher than that in the first day after flowering. The expression level of *WRKY68* in WFYT025 was significantly higher than that of its parents. In the two stages, we found 9 identical transcription factors. Except for 5 reported TFs, the other 4 TFs may play an important role in grain number and photosynthesis heterosis.

### **The key metabolic pathways of heterosis**

In order to eliminate the impact of environment, we studied the metabolic pathway with same  $SDG_{hp}$  ( $SSDG_{hp}$ ) in two environments. There were 312  $SSDG_{hp}$ s in total. These  $SSDG_{hp}$ s were mainly concentrated in the phosphorus metabolic process, phosphorylation, and plasma membrane and so on. During senescence in annual crop plants, P (phosphorus) is mobilized from leaves and other vegetative tissues and translocated to developing seeds, which are a strong P sink during the reproductive growth phase [31, 32, 33]. Previous studies have shown that developing seeds are strong sinks for P, and the remobilization of P from vegetative tissues results in insufficient P for other processes necessary for continued growth such as photosynthesis. In rice (*Oryza sativa* L.), P is predominantly mobilised from leaves during the later stages of grain filling [34, 35]. This might be the reason why the filling speed of rice is faster from the first day to the tenth day after flowering than from the eleventh day to the twenty-first day.

## Abbreviations

Mid-parent heterosis (MPH), higher parent heterosis (HPH), weighted gene co-expression network analysis (WGCNA), quantitative trait loci (QTL), *dwarf and gladius leaf 1* (*dgl1*), wild type (WT), *NARROW LEAF 1* (*NAL1*), *NARROW AND ROLLED LEAF 1* (*NRL1*), *ABNORMAL VASCULAR BUNDLES* (*AVB*), *WUSCHEL-RELATED HOMEODOMAIN 3A* (*OsWOX3A*; *OsNS*), Changhui T025 (CHT025), Wufeng B (WFB), different expression genes (DEGs), Gene Ontology (GO), Fragments Per Kilobase of exon model Per Million mapped reads (FPKM), Per Kilobase of exon model Per Million mapped reads (RPKM), significant difference genes ( $SDG_S$ ), significant difference gene (SDG), phosphorus (P).

## Declarations

### Ethics approval and consent to participate

WFYT025 along with its parental lines Changhui T025 (CHT025) and Wufeng B (WFB) obtained from the native cultivar line and all plants were grown in the test fields of Jiangxi Agriculture University. Sampling of plant materials were performed in compliance with institutional, national, and international guidelines.

### Consent for publication

Written informed consent for publication was obtained from all participants.

### Availability of data and materials

GSE175967, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE175967> provides access to all of data

### Competing interest

The authors declare that they have no conflict of interest.

## Funds

This research was supported by grant (20192BCB23010) from Project of Science and Technology Department of Jiangxi Provinc, grant (GJJ170241) from Project of Science and Technology Department of Department of Education of Jiangxi Province and grant (201910410003) from National Undergraduate Training Program for Innovation and Entrepreneurship.

## Author contributions

Q.C designed and performed experiments, analyzed data and wrote the manuscript. S.H participated in designing and performing experiment. P.W and S.L completed the manuscript with inputs in technical support, critical writing and suggestions regarding the manuscript. G.W, Y.L, L.L, Y.W, Q.Z, T.H, M.Y, X.H participated in performing experiments. All authors reviewed the manuscript.

## Acknowledgements

We thank the anonymous referees for their critical comments on this manuscript.

## References

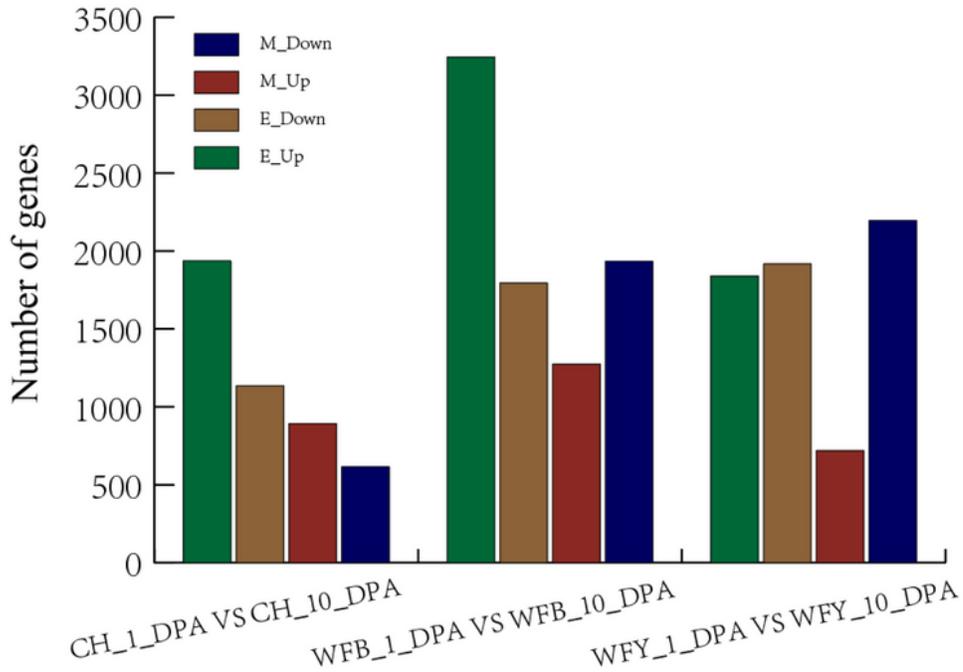
1. Wang P, Zhou GL, Cui KH, et al. Clustered QTL for source leaf size and yield traits in rice (*Oryza sativa* L.). *Molecular Breeding*. 2012; 29(1):99-113.20.
2. Ghosh S, Sahai VN, Saran S. Role of flag leaf on grain yield and spikelet sterility in rice cultivar. *Oryza*. 1990; 27:87–9.
3. Jebbouj R, Yousfi BE. Barley yield losses due to defoliation of upper three leaves either healthy or infected at boot stage by *Pyrenophora teres* f. *teres*. *Eur J Plant Pathol*. 2009;125(2):303–15.
4. Li ZK, Pinson SRM, Stansel JW, Paterson AH. Genetic dissection of the source-sink relationship affecting fecundity and yield in rice (*Oryza sativa* L.). *Mol Breed*. 1998;4(5):419–26.
5. Monyo JH, Whittington WJ. Genotypic differences in flag leaf area and their contribution to grain yield in wheat. *Euphytica*. 1973;22(3):600–6.
6. Li Z, Pinson S, Stansel JW, et al. Genetic dissection of the source-sink relationship affecting fecundity and yield in rice (shape *Oryza sativa* L.).*Molecular Breeding*.1998;4(5):419–426.
7. Cui KH, Peng SB, Xing YZ, et al. Molecular dissection of the genetic relationships of source, sink and transport tissue with yield traits in rice. *Tag. theoretical & Applied Genetics. theoretische Und Angewandte Genetik*.2003;106(4): 649.
8. Wang P, Zhou GL, Yu HH, et al. Fine mapping a major QTL for flag leaf size and yield-related traits in rice. *Theoretical & Applied Genetics*.2011;123(8):1319-1330.23.
9. Duan MJ, Sun SS. Profiling the expression of genes controlling rice grain quality. *Plant Molecular Biology*.2005; 59(1):165.

10. Qi J, Qian Q, Bu Q, et al. Mutation of the rice Narrow leaf1 gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiology*.2008;(4):1947–59.
11. Cho SH, Yoo SC, Zhang HT, et al. The rice narrow leaf2 and narrow leaf3 loci encode WUSCHEL-related homeobox 3A (OsWOX3A) and function in leaf, spikelet, tiller, and lateral root development. *New Phytologist*. 2013;198(4): 1071–1084.
12. Fujino K, Matsuda Y, Ozawa K, et al. NARROW LEAF 7 controls leaf shape mediated by auxin in rice. *Molecular Genetics & Genomics*. 2008; 279(5):499–507.
13. Hu J, Zhu L, Zeng DL, et al. Identification and characterization of NARROW AND ROLLED LEAF 1, a novel gene regulating leaf morphology and plant architecture in rice. *Plant Molecular Biology*. 2010;73(3):283–292.
14. Ma L, Sang, XC, Zhang T, et al. ABNORMAL VASCULAR BUNDLES regulates cell proliferation and procambium cell establishment during aerial organ development in rice. *New Phytologist*. 2017;213(1): 275–286.
15. Lin M, Pang CY, Fan SL, et al. Global analysis of the *Gossypium hirsutum* L. Transcriptome during leaf senescence by RNA-Seq. *BMC Plant Biology*.2015;15(1):43.
16. Zhang WY, Xu YC, Li WL, et al. Transcriptional Analyses of Natural Leaf Senescence in Maize. *Plos One*. 2014; 9(12):e115617.
17. Zhu T, Budworth P, Chen W et al. Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnol Journal*. 2003; (1):59–70.
18. Chen L, Bian JM, Shi SL, et al. Genetic analysis for the grain number heterosis of a super-hybrid rice WFYT025 combination using RNA-Seq. *Rice*.2018; 11(1):37.
19. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*.2008;36(1):D480-D484.
20. MD Young, Wakefield M J, Smyth G K, et al. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*.2010;11(2): R14-R14.
21. Hao YH, Wang HJ, Qiao SL, et al. Histone deacetylase HDA6 enhances brassinosteroid signaling by inhibiting the BIN2 kinase *Proc Natl Acad Sci U S A*. 2016;113(37):10418–23.
22. Griffing B. Use of a Controlled-Nutrient Experiment to Test Heterosis Hypotheses. *Genetics*.1990;126(3): 753–767.
23. Stuber CW, Edwards MD, Wendel JF. Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits. *Crop Science*.1987;27.
24. Bian JM, Jiang L, Liu LL, et al. Identification of Chromosome Segments Associated with Heterosis for Yield in × Rice Hybrids. *Journal of Rheumatology*.2010; 37(8):1673–9.
25. Rajkumar AP, Qvist P, Lazarus R, et al. Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC Genomics*. 2015; 25;16(1):548.
26. Audic S, Claverie J M. The significance of digital gene expression profiles. *Genome research*. 1997;7(10):986.

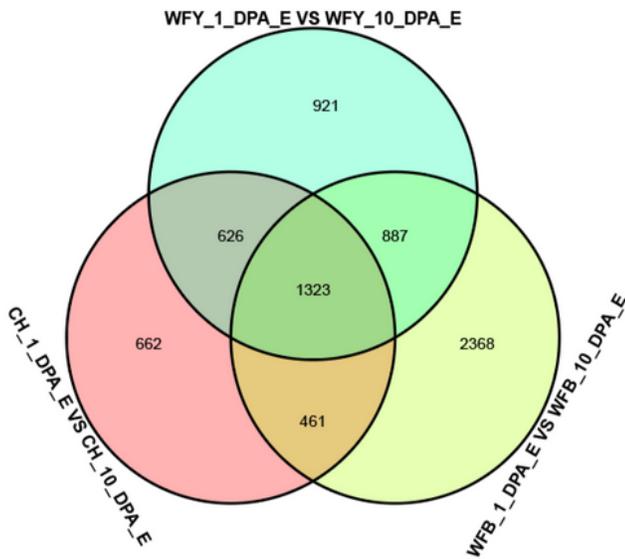
27. Garg R, Singh V K, Rajkumar M S, et al. Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *Plant Journal*. 2017; 91(6):1088–1107.
28. Yuan LP. Super-high yield hybrid rice breeding. *Hybrid Rice*. 1997;12: 1–6.
29. Kusterer B, Muminovic J, Utz HF, et al. Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. *Genetics*. 2007;175(4): 2009–2017.
30. Chen F, Hu Y, Alessandro V, et al. The WRKY transcription factor family in model plants and crops. *Critical reviews in plant sciences*. 2018; 36:311–335.
31. Gregersen PL, Holm PB, Krupinska K. Leaf senescence and nutrient remobilisation in barley and wheat. *Plant Biology*. 2010;10(s1): 37–49.
32. Kong LG, Guo HH, Sun MZ. Signal transduction during wheat grain development. *Planta*. 2015;241(4):789–801.
33. Kyla S, William P. Molecular Mechanisms of Phosphorus Metabolism and Transport during Leaf Senescence. *Plants*, 2015;4(4):773–798.
34. Wang FM, Terry R, Kwanho J, et al. The knowns and unknowns of phosphorus loading into grains, and implications for phosphorus efficiency in cropping systems. *Journal of Experimental Botany*. 2016; (5):122.
35. Julia C, Wissuwa M, Kretzschmar T, et al. Phosphorus uptake, partitioning and redistribution during grain filling in rice. *Annals of Botany*. 2016;118(6):1151–1162.

## Figures

A



B



C

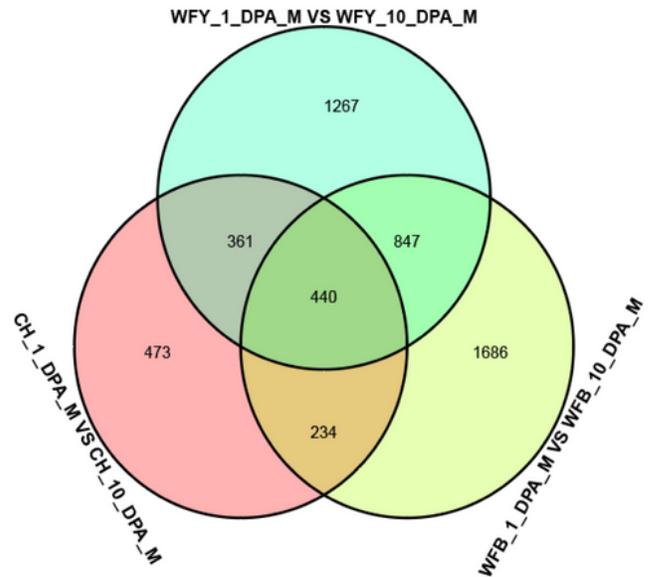
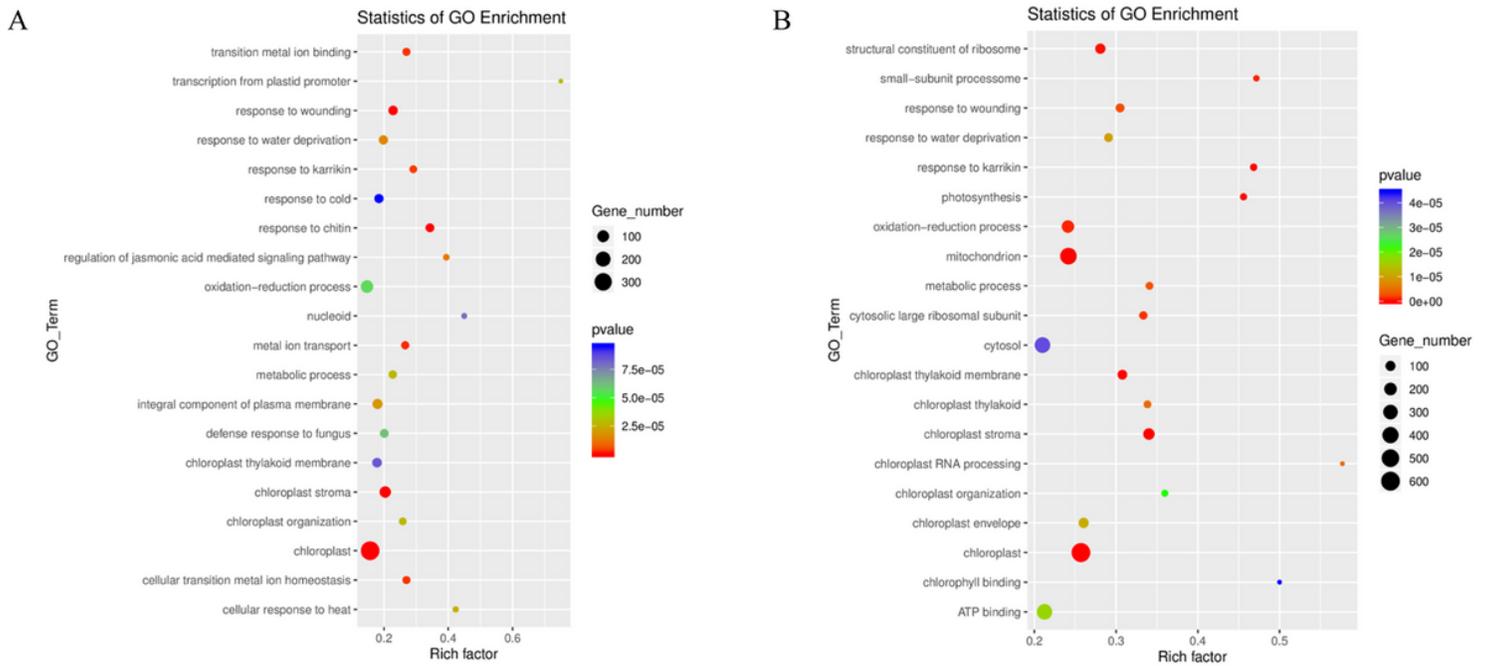


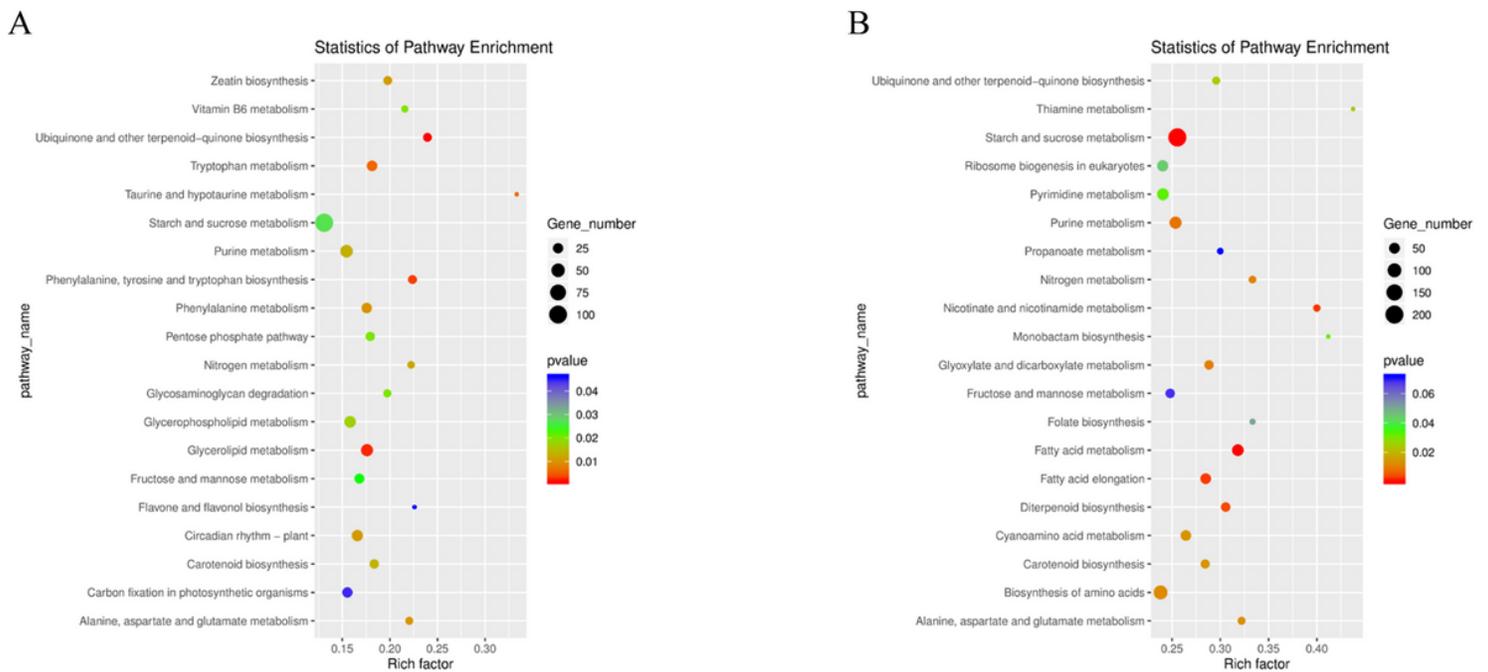
Figure 1

DEGs in super hybrid WFYT025 combination. (A) Number of DEGs between the hybrid and its parents in different stages under different environment. (B) Venn diagram of DEGs between the hybrid and its parents under the early rice environment (C) Venn diagram of DEGs between the hybrid and its parents under the middle rice environment. CH, WFY and WFB represent CHT025, WFYT025 and WFB, respectively.



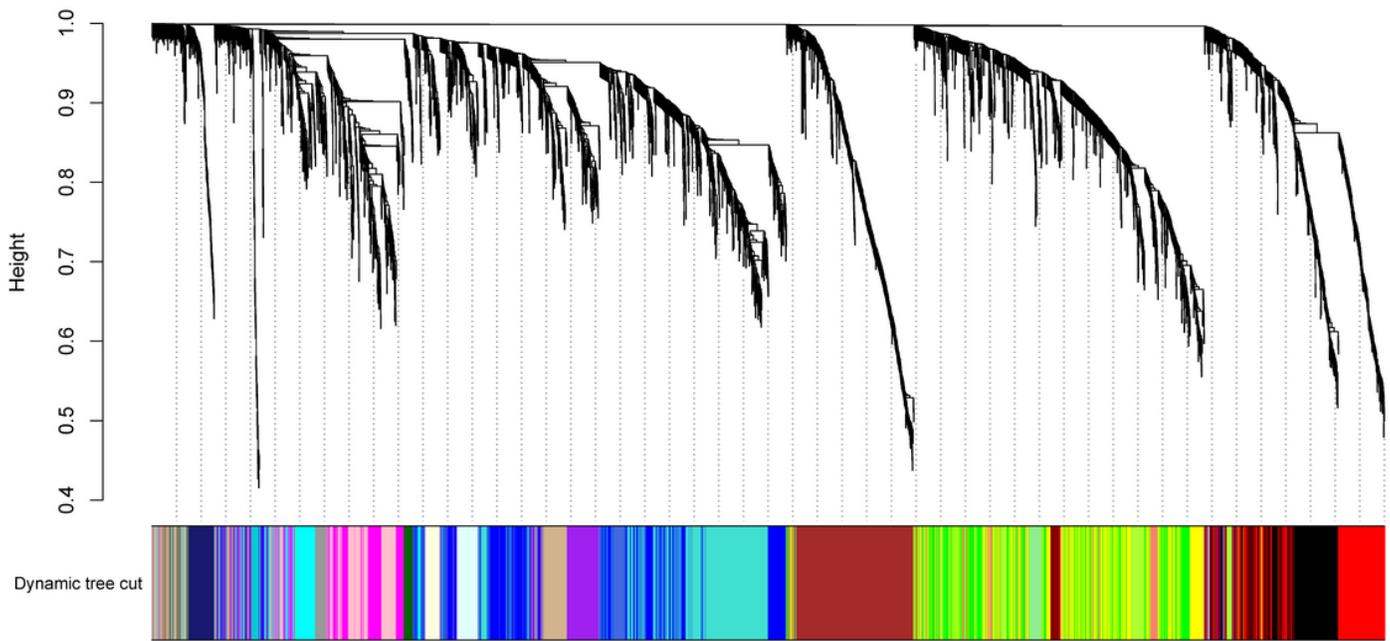
**Figure 2**

Gene ontology (GO) enrichment analysis of DEGs(A) GO enrichment analysis of DEGs on the first day and the tenth days under the environment of early rice in CHT025. (B) GO enrichment analysis of DEGs on the first day and the tenth days under the environment of early rice in WFB.



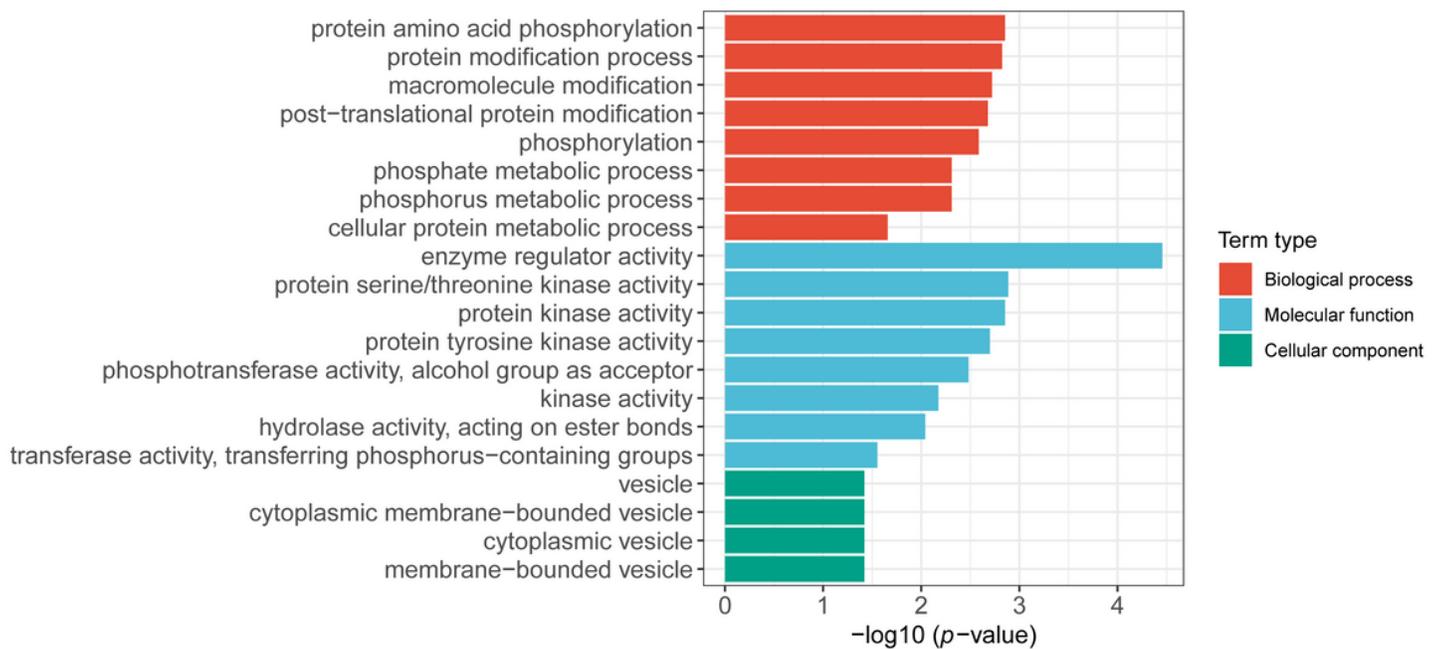
**Figure 3**

KEGG pathway assignments of DEGs. (A) KEGG analysis of DEGs between CHT025 in the environment of early rice. (B) KEGG analysis of DEGs between WFB in the environment of early rice. (A) (B) showed the top 20 most represented categories and the number of transcripts predicted to belong to each category.



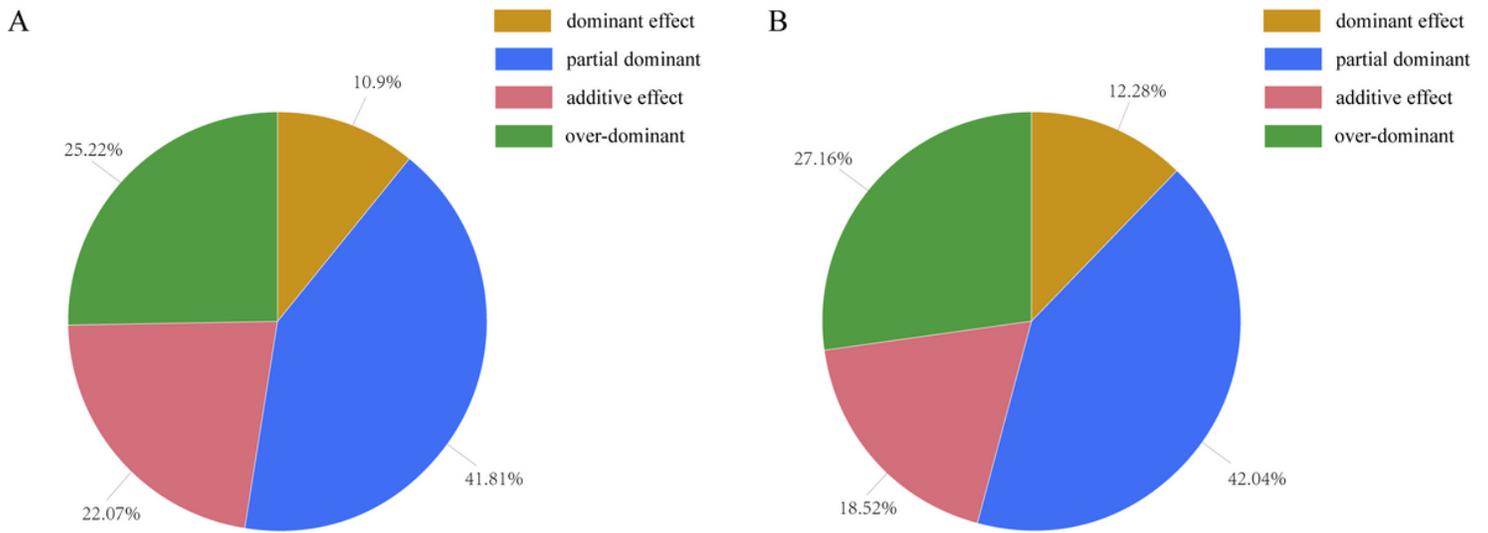
**Figure 4**

Analysis results of 5000 genes in co expression module.



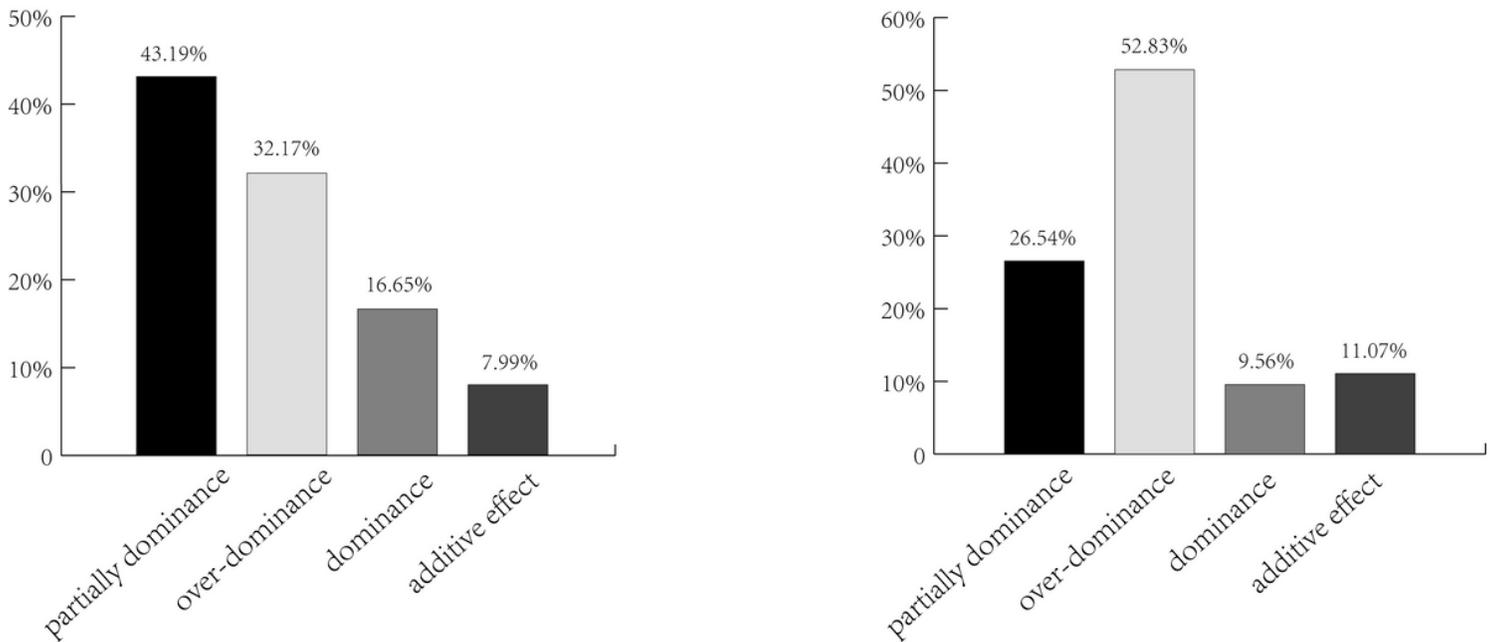
**Figure 5**

Go enrichment analysis of midnightblue module about 106 genes in the co expression module, showed statistically significant GO term (FDR < 0.05). The x-axis indicates the level of statistical significance, and the y-axis indicates the significantly enriched GO term.



**Figure 6**

Breakdown of the SDGhps according to the dominance ratio  $H_p$  under the environment of early rice. Depending on the principal of  $H_p = [d] / [a]$ ,  $H_p$  was classified as either positive or negative.



**Figure 7**

Breakdown of the SDGhps according to the dominance ratio  $H_p$  under the environment of middle rice. Depending on the principal of  $H_p = [d] / [a]$ ,  $H_p$  was classified as either positive or negative.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterials1.docx](#)