

STRUCTURAL GENES ENCODING F3'5'H IN 4D CHROMOSOME CONTROL THE BLUE GRAIN TRAIT IN WHEAT

Diddugodage Chamila Jeewani^{1,2*}, Wang Zhonghua¹, Jian Wenhui¹, Muththanthirige Don Lalith Chandana Nishantha¹, Nan Wenzhi¹ and Liu Tianxiang¹

¹State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi, China

²Grain Legumes and Oil Crops Research and Development Centre, Department of Agriculture, Angunakolapelessa, Sri Lanka

ABSTRACT

Blue grained wheat contained a higher quantity of natural anthocyanin compounds while normal commercial wheat does not have. Though the genes related to several colours of wheat have been identified, the major genes in blue wheat are still unrevealed. Hence, combining the SNP mapping, and transcriptome analysis, pivotal genes regulating blue grain trait was identified. SNP genotyping was carried out in an F2 blue and white wheat population. The blue trait was controlled by a gene/locus located between two SNP markers of IWB 18525 and IWB16381 on the 4D chromosome. Comparative transcriptome analysis revealed that 40 structural differentially expressed genes (DEGs) related to anthocyanin biosynthesis had significant expression differences between blue and non-blue samples. Among them, 12 DEGs expressed only in blue samples while 2 DEGs were specific to blue wheat. Only two F3'5'H genes located in 4D (Traes_4DL_27C195FDE, Traes_4DL_5A3D8F519) were consistent with the location results performed in SNP genotyping. Further, F3'5'H is considered as the main enzyme for Delphinidin compounds, cause for blue colouration. Hence, two genes encoding F3'5'H in the 4D chromosome preferentially account for the blue pigmentation in wheat.

Keywords: anthocyanin, blue wheat, F3'5'H, SNP genotyping, structural genes, transcriptome

INTRODUCTION

Common wheat cultivars usually produce either red or white grains, but some cultivars possess either purple or blue grains. Purple and blue grains are caused by different anthocyanins, which deposit in the pericarp (purple) or aleurone layer (blue) (Trojan *et al.*, 2014). Due to its high anthocyanin compounds, which contribute positively to human health, especially to prevent coronary heart diseases, certain cancers, oxidative stress, systemic inflammation (Kong 2003; Li and Beta 2011; Ross and Kasum 2002), and other age-related diseases, scientific community made great attention on blue or purple wheat recently. It has been shown that the purple colour was caused by the deposition of cyanidin 3-glucoside, followed by peonidin 3-glucoside in the pericarp of purple caryopses, and blue colour was caused by delphinidin 3-glucoside

in the aleurone layer of blue caryopses during caryopsis development (Abdelaal and Hucl 2003; Escribanobailon *et al.* 2004).

Anthocyanin biosynthesis initiates with the chalcone synthase (CHS) enzyme that exploits a p-coumaroyl-coa and three malonyl-coas to form chalcone, which is subsequently transformed to naringenin by chalcone isomerase (CHI). Then naringenin is converted to dihydroflavonol by flavanone-3-hydroxylase (F3H). dihydroflavonols (dihydrokaempferol, dihydromyricetin, and dihydroquercetin) are subjected to diverse modifications by the action of flavonoid 3-hydroxylases (F3H) and flavonoid 3',5'-hydroxylase (F3'5'H) which primarily cause colour differences among anthocyanin pigments (Ahn *et al.* 2015). Accordingly, the resulting dihydroflavonols act as immediate precursors for the synthesis of

*Corresponding author: jeewanididdugoda@yahoo.com

leucoanthocyanidins and flavonols by dihydroflavonol 4-reductase (DFR) and flavonol synthase (FLS), respectively. These colourless leucoanthocyanidins are transformed by anthocyanidin synthase (ANS) to coloured anthocyanidins. Subsequently, these anthocyanidins are converted by glycosyltransferases (GTS) to anthocyanin pigments (Holton and Cornish 1995).

Genes responsible for anthocyanin pigmentation in different colours of wheat have been studied and mapped into several chromosomes such as three genes (Rc-A1, Rc-B1, and Rc-D1) for red coleoptile, three genes for purple culm (Pc-A1, Pc-B1, Pc-D1), three genes for purple leaf sheath (Pls-A1, Pls-B1, Pls-D1), three homoeologous for purple leaf blades (Plb-A1, Plb-B1, Plb-D1), two genes for purple anther (Pan-A1 and Pan-D1) and two genes for purple pericarp, (Pp1 and Pp3). In addition to that several functional genes such as 6 CHS, 3CHI, 4F3H, 3DFR and 5ANS have been already identified in the wheat genome with red and purple wheat (Shoeva and Khlestkina 2015).

The blue wheat lines were usually developed from distant hybridization, in which alien chromosome or fragments were integrated into the wheat genome (Zeven 1991). Genes for the blue aleurone (*Ba*) colour have been identified in several wheat-related species such as *Thinopyrum ponticum* (*Ba1*) (Keppenne and Baenziger 1990), *Triticum monococcum* (*Ba2*) (Dubcovsky *et al.* 1996), *T. boeoticum* (*Ba2*) (Singh *et al.* 2007) and *Thinopyrum bucranium* (*bathb*) (Shen *et al.* 2013). Though the genes and expression related to the anthocyanin biosynthesis pathway in many plants have been identified in detail, the exact genes and expression of blue colour in wheat have not been detected and verified.

Red or purple wheat was controlled by MYB or bHLH transcription factors. It is still unknown if the blue seed colour is similarly controlled by transcription factors or differently. For the huge genome size and gene redundancy, map-based cloning and functional characterizing some genes related to the anthocyanin biosynthesis in wheat are complex

and find it difficult to understand the molecular mechanism of the blue grain phenotype.

Genetic maps are useful for discovering, dissecting, and manipulating the genes responsible for simple and complex traits in crop plants (Tanksley *et al.* 1992). Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the genome (Rafalski 2002). Accordingly, they are well-suited for genomics approaches such as genomic selection (Meuwissen *et al.* 2001) and association mapping (Myles *et al.* 2009). A newly discovered high-throughput SNP genotyping platform make a revolution in wheat genotyping by allowing large-scale utilization of genotypic data in a cost-effective manner (Akhunov *et al.* 2009). Transcriptome analysis based on high throughput RNA sequencing technologies (RNA-seq) is also a newly identified powerful approach to be used for characterizing the whole profile of gene expression in a given sample (Wang *et al.* 2009; Wolf 2013).

In this study, using a high throughput SNP mapping, the candidate regions and related molecular markers for wheat blue grain trait were identified. The expression profiles of anthocyanin biosynthesis-related genes were investigated using transcriptome analysis between blue and white wheat at two different physiological stages of colour development and the final aim was to find out candidate genes responsible for the blue grained wheat.

MATERIALS AND METHODS

Plant materials

An F₂ population, consisting of 101 individuals, developed from the blue wheat line 'Zhiluowumai' crossed with a white wheat line 'No.4045' was used as the SNP mapping population. For the transcriptome analysis, two blue grain wheat lines ('Zhiluowumai' and 'Xiaoyanwumai') and two white wheat lines ('A14' and 'Xiaoyan mutant') were selected as the plant materials. All the plants were grown in the experimental field of Northwest A & F University (Shaanxi, China) during the growing season of 2015. The altitude of the area was 525 m

and the climate was semi-humid prone to semi-arid with a mean temperature of 13°C and average annual rainfall of 600 mm.

Colour development of wheat grains was visually observed during seed development in both blue and white lines used for transcriptome analysis. According to the observations of the blue colour deposition, sampling dates were determined (Fig.1). Hence, 10 days post-anthesis (dpa) and 20 dpa grains were used to extract RNA samples for transcriptome RNA-seq and Quantitative Real-time PCR (qRT-PCR) analyses.

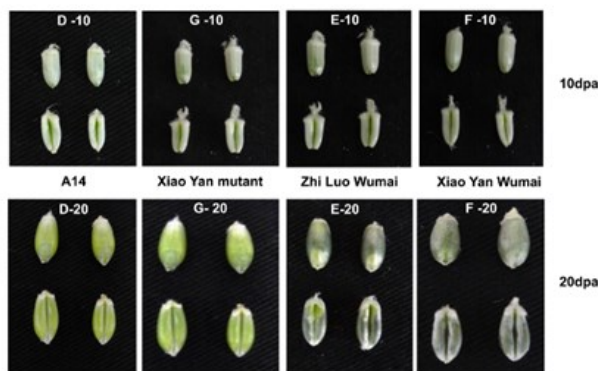


Figure 1. Wheat seed materials used for the transcriptome analysis.

Note: D-10: 'A14' at 10 dpa, D-20: 'A14' at 20 dpa, G-10: 'Xiaoyan mutant' at 10 dpa, G-20: 'Xiaoyan mutant' at 20 dpa, E-10: 'Zhiluwumai' at 10 dpa, E-20: 'Zhiluwumai' at 20 dpa, F-10: 'Xiaoyanwumai' at 10 dpa, F-20: 'Xiaoyanwumai' at 20 dpa

Genetic mapping of the blue grain gene

Genomic DNA of 'Zhiluwumai' and 'No.4045' parental lines and 101 individuals of the F₂ population derived from the above two parents were extracted from the young leaves using the CTAB method and were used for the genotyping. Genotyping was carried out using the recently developed wheat 90k iSelect SNP array comprised of nearly 90,000 gene-associated SNPs (Wang *et al.* 2014). This was performed at China golden marker company (cgmb.com.cn) adapting the manufacturer's recommendations as described in (Akhunov *et al.* 2009). The genotyping assays were done in Illumina

iScan reader using Genome Studio software version 2011.1 (Illumina). The linkage map was constructed using the MAP function in software QTL IciMapping, version 4.1 (Wang *et al.* 2012) and Kosambi mapping function. The ordering markers to the chromosomes referred to 90K SNP maps that were developed in bread wheat (Cavanagh *et al.* 2013; Wang *et al.* 2014). QTL mapping was performed with QTL IciMapping version 4.1 (Wang *et al.* 2012) using inclusive composite interval mapping of additive (ICIM-ADD) module. The map distances were calculated in Centimorgan (cM). Additive QTL was detected using a 1.0 cm step in scanning. The probability used in stepwise regression was 0.001.

RNA extraction and cDNA library preparation for transcriptome sequencing

Seeds of two blue grain wheat lines ('Zhiluwumai' and 'Xiaoyanwumai') and two white wheat lines ('A14' and 'Xiaoyan mutant') at 10 dpa and 20 dpa were collected and squeezed to remove the starch portion of the seed. They were immediately placed in RNA free Eppendorf tubes, which were kept in liquid nitrogen. All samples were collected in triplicate. Total RNA was extracted using the RNA Prep Pure Plant Kit (Tiangen, China) according to the manufacturer's protocol. RNA degradation and contamination were detected by electrophoresis in 1% agarose gels and RNA purity was checked by the Nano Photometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using the Qubit® RNA Assay Kit in Qubit® 2.0 Fluorimeter (Life Technologies, CA, USA). RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used to measure the integrity of RNA.

The mRNA libraries were prepared separately for each sample. Sequencing libraries were created using nebnex® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. The various steps including mRNA enrichment, fragmentation, second-strand cDNA synthesis, PCR amplification, size selection, clustering and sequencing, using Illumina hiseq™2500

platform were performed by the Novogene Bioinformatics Technology Co. Ltd, China. Finally, 150bp paired-end reads were generated.

Sequencing data processing and annotation

After completion of the sequencing, resulted image data were transformed to raw reads and those raw reads were stored in a FASTQ format. Then the data in FASTQ format were cleaned using in-house perl scripts. In this step, reads with adapter, poly-N and other low-quality reads were removed from the raw data. At the same time Phred quality score (Q), which is defined as a property that is logarithmically related to the base calling error probabilities (P)² (www.illumina.com) and GC content were calculated. Low-quality data of $Q < 20$ were removed and higher quality, clean reads from each wheat transcriptome library were used for the downstream analyses.

Reads mapping to the reference genome

Reference genome and gene model annotation files were downloaded from the wheat genome website directly (<https://www.wheatgenome.org>). The index of the reference genome was built using Bowtie v2.2.3 and paired-end clean reads were aligned to the reference genome using TopHat v2.0.12. We selected TopHat as the mapping tool for that TopHat can generate a database of splice junctions based on the gene model annotation file and thus a better mapping result than other non-splice mapping tools.

Quantification of gene expression level

The expression levels of the mapped genes were estimated from the wheat transcriptome sequencing data based on the number of raw reads. Version 0.6.1 of htseq was used to count the read numbers mapped to each gene (Anders *et al.* 2015). Then the reads for each gene were normalized by using fragments per kilobase of transcript per million mapped reads (FPKM). The FPKM of each gene was calculated based on the length of the gene and reads count mapped to this gene. This FPKM was considered the effect of sequencing depth and gene length for the read count at the same time and is currently the most commonly used

method for estimating gene expression levels (Trapnell *et al.* 2010).

Functional analysis of differentially expressed genes

Differential expression analysis of two conditions/groups (two biological replicates per condition) was performed using the Deseq R package (1.18.0). Deseq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P-values were adjusted using Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with the adjusted P-value < 0.05 found by Deseq were assigned as differentially expressed.

GO (Gene Ontology) enrichment analysis of differentially expressed genes was carried out using the Goseq R package. This package can verify the bias of gene length. GO terms with corrected P-value less than 0.05 were considered significantly enriched by differentially expressed genes.

Quantitative Real-time PCR analysis

To validate the accuracy and repeatability of the gene expression data obtained from RNA seq, qRT-PCR was carried out on a few randomly selected DEGs that were prepared from the total RNA. Total RNA from the seed coat of all 4 varieties at two different physiological stages were extracted as described above and cDNA was synthesized using the PrimeScript™ RT Reagent kit with gDNA Eraser (TaKaRa, China). The specific real-time PCR primers were designed using the Primer 5.0 program. qRT-PCRs were carried out on a Step One Plus™ real-Time PCR System (USA) in a 20 μ l reaction volume containing 1:10 diluted cDNA templates, 0.4 μ l of each primer, deionized water, Real star green power mixture and ROX reference dye (GenStar, China). The qrt-PCR conditions were as follows: denaturing at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, 60°C for 45 s and 72°C for 15 s. The relative gene expression level of each target gene was analyzed using the comparative CT method ($2^{-\Delta\Delta C_T}$ method)

(Schmittgen and Livak, 2008). The qrt-PCR reactions were normalized to the Ct values for *TaActin* in wheat. All the reactions were measured in three replicates.

RESULTS AND DISCUSSION

Wheat grain colouring is caused by anthocyanin residue in the pericarp (purple) or aleurone layer (blue) (Trojan *et al.* 2014). At present, major genes related to red and purple wheat have been mapped and identified into several chromosomes (Liu *et al.* 2016; Shoeva and Khlestkina 2015). However, the genes that control the blue grain trait of wheat have not been identified yet. The blue wheat lines were usually developed from distant hybridization, in which alien chromosome or fragments were integrated into the wheat genome (Zeven 1991).

Genetic mapping of the blue grain trait

The analysis with the 90k iSelect Infinium assay produced data for 81,587 SNP markers. Out of that, a total of 10,305 (12.6%), were polymorphic between the parental wheat lines of 'No.4045' and 'Zhiluowumai'. However, among them, 403 markers failed in No.4045 line and 1017 failed in 'Zhiluowumai'. About 2468 markers were heterozygous to No.4045 line, while 2248 markers were heterozygous to 'Zhiluowumai'. Further, 424 markers were deleted while making the linkage map. Therefore, the genetic map was developed based on 3,745 markers. Three wheat genomes of A, B and D were covered by 1487, 1876 and 382 SNP markers, respectively. Their average genetic distances were 1.7cM, 1.3cM and 1.9cM for A, B and D genomes respectively (Table 1). A gene/locus on the 4D chromosome was identified as controlling the blue colour of the wheat

grains and was flanked by two SNP markers *IWB18525* and *IWB16381*, within a 2.16 cM genetic distance. The gene locus was named *Ba* (Fig. 2). A high-density SNP map had not been used for mapping blue grain genes in wheat so far. Furthermore, the blue aleurone trait in hexaploid wheat has not been mapped previously using any other mapping methods and this research first time reports it.

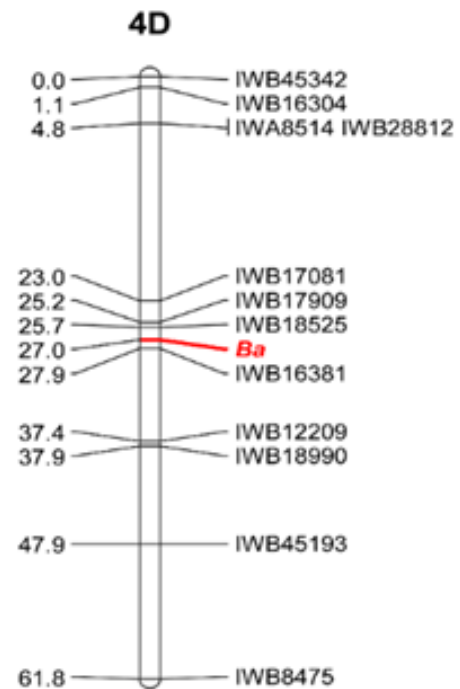


Figure 2. A diagrammatic representation of the 4D chromosome, showing the location of the locus/gene for blue grain colour in F₂ ('Zhiluowumai' [blue] x 'No.4045' [white]) mapping population of wheat.

Strategy for screening candidate genes involved in the blue grains

With the seed development, major colour changes were visually observed in both blue lines by changing the colour from green to blue at 20 dpa while both white lines remained green in colour. This inferred the first screening condition of candidate genes, in which the blue grain-related genes were differentially expressed before and after 20 dpa in the blue line. The second selection was to exclude the differentially expressed genes (DEGs) that commonly occurred in all 4 lines. Therefore, the samples collected at 10 dpa and 20 dpa were used for screening candidate

Table 1. Description of the identified SNP markers that covered the wheat genome

Genome	Genetic Distance-cM	Number of SNP markers	Average Genetic Distance-cM
A	2568	1487	1.72
B	2504.82	1876	1.33
D	753.17	382	1.97
Total	5825.99	3745	1.55

genes with transcriptome analysis. As a result, we had two blue grain samples and 6 non-blue samples as the background controls for transcriptome analysis. The third selection of candidates would focus on the DEGs, involved in the anthocyanin biosynthesis pathway based on the transcriptome analysis.

The blue aleurone lines used in the research were translocation lines, developed from common wheat crossed with *Agropyron elongatum*. Because of the crossed in alien chromosome segments, comparative transcriptome analysis of blue and white wheat could be effective to identify the candidate genes related to the blue grain trait in wheat. It has been reported that large numeric DEGs could be screened out based on transcriptome analysis. It is critical to screen down the unrelated DEGs. As we know that the blue grain colour was affected by genetic background and development stages, combing the factors two blue samples and six non-blue samples were used for the transcriptome analyses. It provides an effective decrease in the background DEG numbers.

Transcriptome sequencing and bioinformatics analysis of differentially expressed genes

The sum of 397,450,110 raw reads was generated by RNA sequencing, which consisted of the averages of 50,129,381 and 49,233,147 raw reads for white and blue groups, respectively. The number of reads from each pool ranged from 44 to 57 million (Table 2). After removing the lower quality reads such as Phred quality scores <20, N reads (The reads,

which could not determine the base information) and adapter sequences, 367 million cleaned reads with 55.14G of cleaned bases were selected to use for subsequent analysis. The percentage of total raw reads to total clean reads was 92.6% (185,671,940) and 92.4% (181,941,724) for white and blue groups, respectively. These data were used for further bioinformatics analysis to screen out differential expressed genes between the groups of blue and non-blue samples.

A total of 126,792 genes was found in blue and white wheat pools. Gene expression differences were assessed by pairwise comparison of two physiological stages of each variety. It indicated that among the total DEGs, 3855 were found to be up-regulated while 5585 were down-regulated compared to each stage in each variety (Fig. 3).

A total of 4946 DEGs was screened out as relatively high level expressed (FPKM value>1) genes in whole data pools, out of which 376 DEGs were commonly expressed in all four lines (Fig. 4a). At first, the differential expression was compared between 10 dpa and 20 dpa samples within each line. Then the DEGs were compared separately within white and blue line groups. As a result, 500 DEGs were commonly expressed in the blue wheat lines (Fig. 4b) and 1316 DEGs were common in white lines (Figure 4c). Combining the DEGs between blue and white groups 376 DEGs were common to all lines, while 124 DEGs were only in blue lines (Fig. 4d). Among the 124 commonly expressed DEGs in both blue lines, 75 DEGs were up-regulated

Table 2. Summary of RNA sequencing data

Sample ID	Raw Reads	Clean Reads	Clean Bases	Total Mapped	Total mapping Ratio (%)	Q20 (%)	Q30 (%)	GC Content (%)
D-10	57,679,132	53,689,772	8.05G	38,597,746	71.89%	96.69	92.25	54.5
D-20	45,581,164	42,123,340	6.32G	26,321,886	62.49%	95.77	90.89	54.28
G-10	51,867,850	47,741,050	7.16G	31,810,796	66.63%	95.89	91.12	53.38
G-20	45,389,376	42,117,778	6.32G	27,061,531	64.25%	96.39	91.79	56.75
E-10	49,319,576	45,562,436	6.83G	29,825,004	65.46%	96.18	91.53	54.59
E-20	53,213,544	49,049,412	7.36G	30,478,869	62.14%	95.96	91.25	54.79
F-10	44,873,124	41,554,242	6.23G	28,761,688	69.21%	96.33	91.68	54.41
F-20	49,526,344	45,775,634	6.87G	28,133,833	61.46%	96.11	91.42	56.77
Total	397,450,110	367,613,664	55.14G					

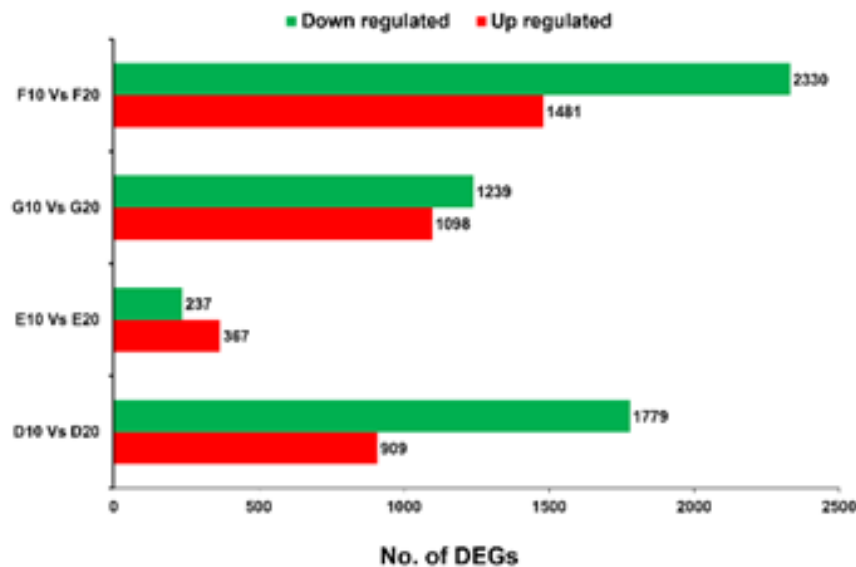


Figure 3. Comparative up regulating and downregulating DEGs in blue and white wheat lines.

Note: D-10: 'A14' at 10 dpa, D-20: 'A14' at 20 dpa, G-10: 'Xiaoyan mutant' at 10 dpa, G-20: 'Xiaoyan mutant' at 20 dpa, E-10: 'Zhiluowumai' at 10 dpa, E-20: 'Zhiluowumai' at 20 dpa, F-10: 'Xiaoyanwumai' at 10 dpa, F-20: 'Xiaoyanwumai' at 20 dpa

and 49 DEGs were down-regulated. The most available up-regulated DEGs in blue wheat pools included NAC transcription factor, chalcone synthase, hydroxylase and serpin domain-containing protein.

GO annotation was done for the blue specific 124 DEGs and they were further categorized into three main categories of biological process, cellular component and molecular function (Fig. 5). We obtained 37 clusters, which consist of 17, 9 and 11 clusters for biological process, molecular function and cellular component, respectively. The topmost score recording clusters in cellular component were 'cell' and 'cell part'. 'Metabolic process' and 'catalytic activity' expressed the highest number of GO terms of the biological process and molecular function respectively. Go graphs were created using WEGO software (Ye et al. 2006).

Screening candidate transcription factor genes

Blue grain trait visually started to occur on seed coat at 20 dpa, regulated by different transcription factors. The structural genes and

regulatory genes involved in anthocyanin biosynthesis in other plants have been reported.

Regulatory genes are capable of controlling the developmental or tissue-specific expressions exhibited by anthocyanin structural genes. A set of proteins, including MYB, bHLH and WD40 form a regulatory complex, which controls the transcription of anthocyanin structural genes (Ramsay and Glover 2005). Based on the transcriptome analysis 410 transcription factor genes were annotated as DEGs, in which 34 MYB, 12 bHLH and 7 WD40 DEGs were detected among the 8 samples. Only one MYB (Traes_1AL_01E24503F) and one bHLH (Traes_2AL_67BC3AFBE) were screened out as the candidate transcription factor genes which differentially expressed between blue samples and non-blue samples.

The MYB gene (Traes_1AL_01E24503F) showed 2.7 to 2.8 times higher expression in 20 dpa samples than in 10 dpa in blue wheat lines. However, the bHLH gene had an 8.6 to 23.8 times higher expression level in 20 dpa samples than in 10 dpa samples.

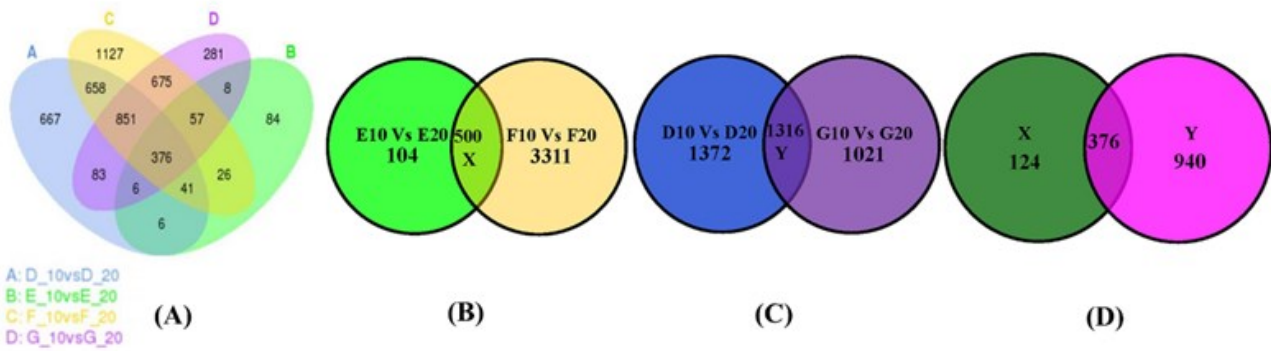


Figure 4. Venn diagrams of DEGs.

Note: (a) DEGs among the four wheat lines of blue and white which have the gene expression (FPKM) > 1. (b) DEGs between two blue wheat lines. (c) DEGs between two white wheat lines. (d) DEGs common for blue and common for white

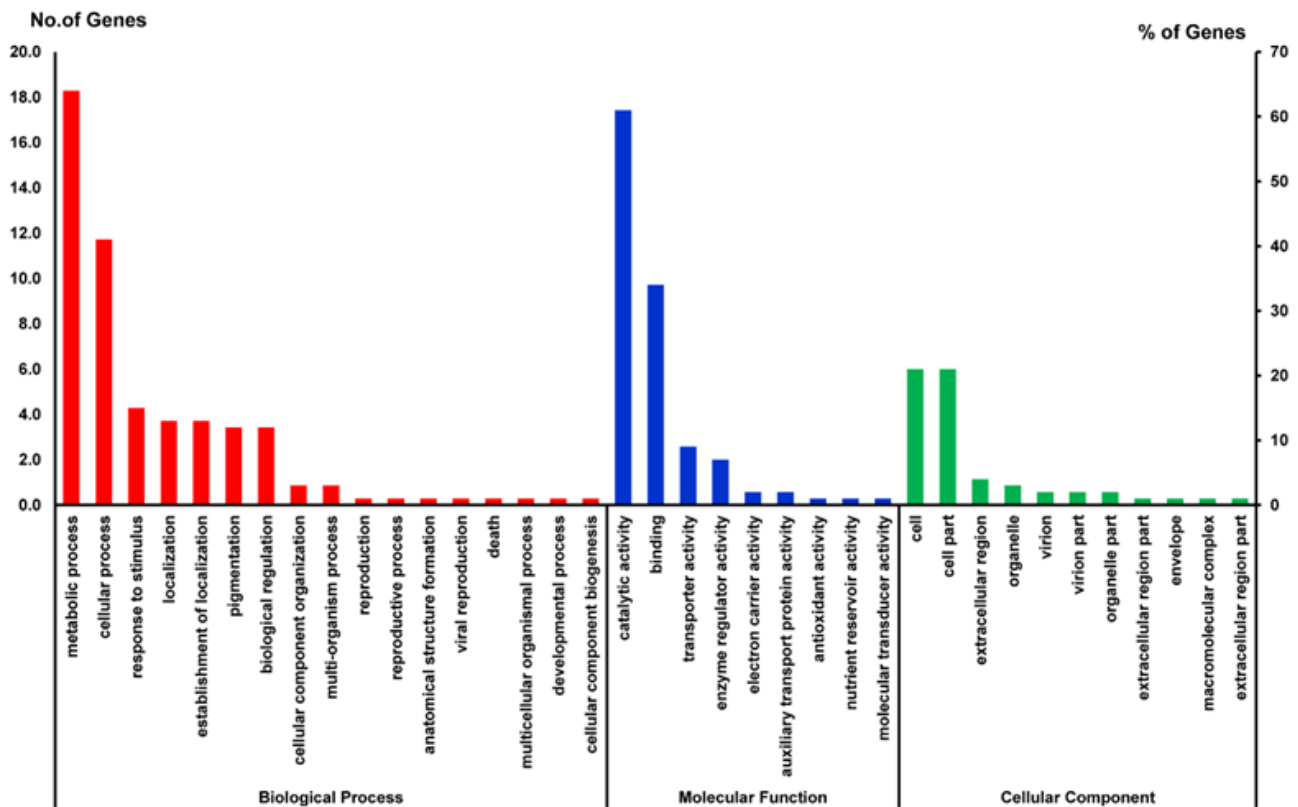


Figure 5. Gene Ontology classification of DEGs in blue wheat gene pool compared to white wheat.

Note: Red colour indicates a biological process, the blue colour indicates the molecular function and green colour indicates cellular components

Candidate structure genes involved in anthocyanin biosynthesis

Based on the bioinformatics analysis of the transcriptome, 40 structural DEGs involved in anthocyanin biosynthesis pathway were identified (Fig.6). Among the identified transcripts, 12 DEGs were recorded highly up-regulated at 20 dpa only in blue wheat. Those candidate genes included 5 CHS genes

(Traes_2AS_0EA2792B8, Traes_2BL_07EC87598, Traes_2BS_4AC3D17E8, Traes_2AL_ED4D3BEC1, Traes_2DS_8827E95F0), two F3'5'H genes (Traes_4DL_27C195FDE, Traes_4DL_5A3D8F519), two DFR genes (TRAES3BF026100070CFD_g, Traes_3AL_197871859), two ANS genes

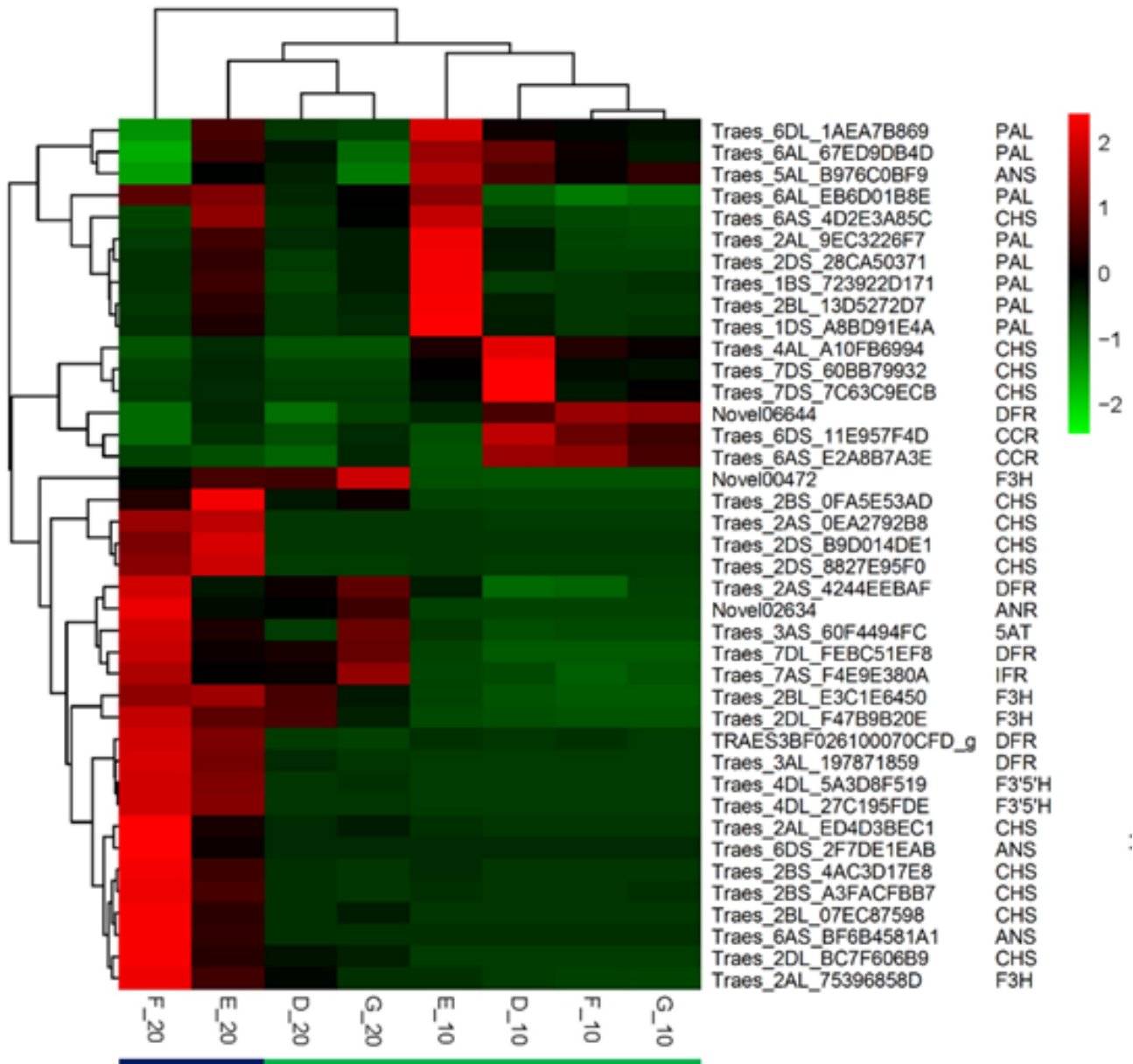


Figure 6. Heat map of the expression of structural genes related to flavonoid biosynthesis in blue and non-blue samples.

Note: Differentially Expressed Gene IDs are indicated on the right side of the heat map. Blue and green lines under the heat map indicated the colour of the seed at that moment

(Traes_6AS_BF6B4581A1, Traes_6DS_2F7DE1EAB) and one Anthocyanin 5-aromatic acyltransferase gene (Traes_3AS_60F4494FC).

Of the five CHS expressed only in blue wheat, 2 CHS (Traes_2AS_0EA2792B8 and Traes_2AL_ED4D3BEC1) had alternative splicing events at 5' first exon (TSS) and 3' last exon (TTS) while another two CHS, (Traes_2BL_07EC87598 and Traes_2BS_4AC3D17E8) had SNP with Cytosine into Adenine and Thymine, respectively.

Two genes annotated as Flavonoid 3', 5'-hydroxylase (Traes_4DL_5A3D8F519 and Traes_4DL_27C195FDE) were remarkably highly up-regulated during blue colouration in blue wheat varieties at 20 dpa. Those two DEGs had increased relative expression of 72–128 fold in blue samples. Traes_4DL_27C195FDE gene had two alternative splicing events at TSS and TTS.

Up-regulation of two DFR encoding genes (TRAES3BF026100070CFD_g and Traes_3AL_197871859) was evident in blue samples. TRAES3BF026100070CFD_g had shown alternative splicing at TSS and TTS, while Traes_3AL_197871859 exhibited a SNP. Two ANS of Traes_6AS_BF6B4581A1, Traes_6DS_2F7DE1EAB and one Anthocyanin 5-aromatic acyltransferase gene (Traes_3AS_60F4494FC) showed very high expression in both blue samples while not expressed in other samples. No alternative splicing and SNPs were detected in the three genes.

To confirm the expression levels of these genes involved in the anthocyanin biosynthesis pathway, quantitative Real-Time PCR analyses were performed in three biological replicates on six selected genes including CHS, F3H, F3'5'H, DFR and ANS. The results were consistent with the findings obtained by the RNA seq (Fig. 7).

The flavonoid biosynthetic pathway has been extensively studied at the genetic, molecular and biochemical levels in various plant spe-

cies, including maize, snapdragon, petunia (Holton and Cornish 1995), strawberry (Almeida et al. 2007), litchi (Lai et al. 2015) and Arabidopsis (Shirley et al. 1995; Bharti and Khurana 1997). In this research, most of the genes involved in anthocyanin biosynthesis of blue grain were identified. It displayed a profile of gene networks involved in the blue seed pigment biosynthesis.

Anthocyanins in various colours of bread wheat were identified and quantified (Abdel-Aal et al. 2006). Delphinidin 3-O-rutinoside, delphinidin 3-O-glucoside, and malvidin 3-O-glucoside are predominant in blue wheat, whereas cyanidin 3-O-glucoside, peonidin 3-O-galactoside, and malvidin 3-O-glucoside are common in purple wheat (Ficco et al. 2014). F3'5'H is the main enzyme directed to the production of delphinidin compounds, which give blue colouration (Carolyn and Mark 2012) (Fig. 8). We found two F3'5'H genes (Traes_4DL_5A3D8F519 and Traes_4DL_27C195FDE), located in 4D and significantly higher expressed at 20 dpa in blue wheat. They were not expressed at 10 dpa. qRT-PCR results also validated the expression of F3'5'H in blue wheat.

Genetic research showed that the blue grain was controlled by one or two genes. Transcription factors and structural genes screened out in this research are located on 8 different chromosomes (2A, 2B, 2D, 3A, 3B, 4D, 6A, 6D-TraesS3BF026100070CFDg, Traes3AL197871859, Traes2AS0EA2792B8, Traes2BL07EC87598, Traes2BS4AC3D17E8, Traes2ALE4D3BEC1, Traes2DS8827E95F0, Traes4DL27C195FDE, Traes4DL5A3D8F519, Traes_6AS_BF6B4581A1, Traes6DS2F7DE1EAB, Traes3AS60F4494FC). The SNP genotyping performed using wheat 90k SNP chips identified the genetic location of the blue trait in the 4D chromosome. It has been reported purple grain was caused by transcription factors (Jiang et al. 2018). Based on our mapping result, the two-transcription factor DEGs were excluded as they are located in 1A and 2A chromosomes, and two F3'5'H DEGs were suggested as the candidate control genes for the blue trait. It potentially displayed a novel control mechanism of the blue grain trait.

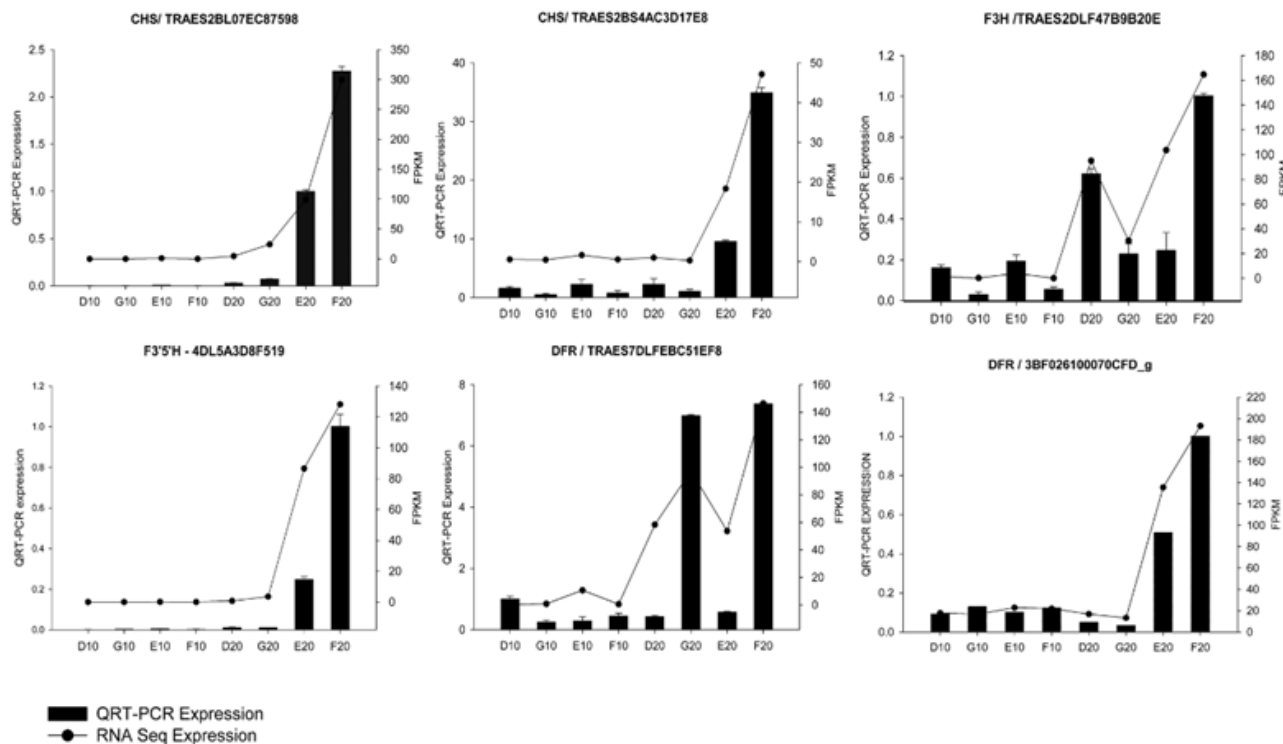


Figure 7. qRT-PCR validation of differential gene expression in anthocyanin biosynthesis.

Note: The left ‘Y’ axis indicates relative gene expression levels determined by qRT-PCR. The right ‘Y’ axis indicates gene expression levels calculated by the RNA sequencing. Bars represent the standard error

There are many useful genes such as several disease-resistant genes of *Sr31* and *Sr38* resistant to stem rust (Knott 1989), *Bdv2* gene resistance to barley yellow dwarf virus (Crasta et al. 2000), *Lr41* resistance to leaf rust (Sun et al. 2009), *Yr26* gene resistant to all important races of *Puccinia striiformis* f. sp. Tritici, (Zhang et al. 2013) is transferred from distant hybridization and many elite lines are translocation lines in wheat. It causes a very difficult problem that the inserted fragment usually unpaired in meiosis and as a result, no cross events will occur in the fragments. Due to this reason, the fine-mapping is impossible in this translocated area and cloning of the gene is not practicable. It raises a common problem of wheat that how to clone the particular translocated gene. This research provides a good example to solve this problem. By mapping, we can identify the location of the particular gene and by transcriptome analysis, we can find the genes related to that chromosome and trait. This could be applied

to similar cases.

The blue wheat lines we used for this research were translocated lines and it was based on the evidence of fluorescence in situ hybridization (Jia et al. 2001). By comparing our mapping results with the wheat consensus map, we can further confirm that they are truly translocated lines. The functional analysis of the F3’5’H genes is needed to be carried out. Transgenic expression assay of these genes should provide clear evidence for understanding the molecular mechanism of blue grain trait in wheat.

CONCLUSION

The SNP genotyping performed using 90k iSelect Infinium assay identified the location of the gene/locus, which control the wheat blue grain trait in the 4D chromosome flanked by two SNP markers IWB18525 and IWB16381, within a 2.16 cM genetic distance. In our transcriptome analysis revealed that 40 structural DEGs were related to anthocyanin

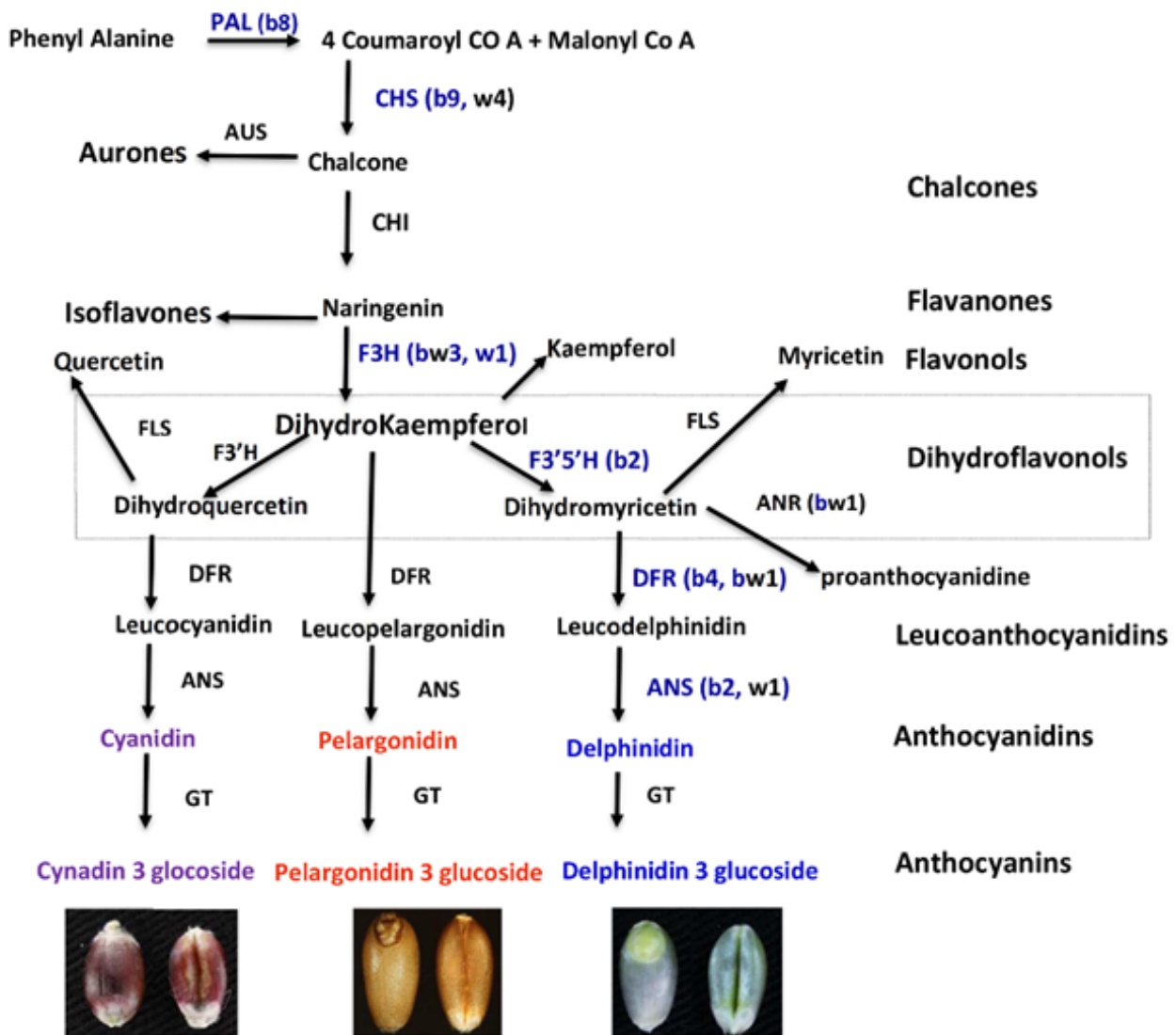


Figure 8. Simplified flavonoid biosynthesis pathway in blue and white wheat.

Note: Numbers with the “b” indicate the number of DEGs identified in blue wheat and “w” indicates the number of DEGs identified in non-blue wheat. Numbers with the “bw” indicate the number of DEGs in both blue and non-blue wheat lines. Pathway enzymes are indicated in abbreviated as follows: PAL: phenylalanine ammonia-lyase; CCR: Cinnamoyl CoA reductase; CHS: Chalcone Synthase; CHI: Chalcone Isomerase; F3H: Flavanone 3-hydroxylase; F3'5'H: Flavonoid 3',5'-hydroxylase; F3'H: Flavonoid 3'-hydroxylase; DFR: Dihydro Flavonol-4-Reductase; ANS: Anthocyanidin Synthase; GT: Glycosyltransferase; FS: Flavone Synthase; FLS: Flavonol Synthase; IFS: Iso Flavone Synthase; IFR: Iso Flavone Reductase; AUS: Golden Grass Synthase; ANR: Anthocyanidin reductase (Modified from Holton and Cornish 1995).

biosynthesis and had obvious expression differences between blue and white wheat. Among them, 12 DEGs expressed only in blue wheat at 20 dpa, highlight the importance of them to produce blue colouration. This research successfully found the molecular processes involved in the biosynthesis of antho-

cyanin in blue wheat. Delphinidin compounds are predominant in blue wheat and F3'5'H is the crucial enzyme that leads to the production of Delphinidin. Hence, two DEGs, which were located in 4D and annotated as F3'5'H are the most important and key structural gene for the blue colouration in wheat. This was

further confirmed by the results of qRT-PCR and mapping. It has been reported that purple grain was caused by transcription factors. Based on our mapping result, the two-transcription factor DEGs were omitted, and two F3'5'H DEGs were suggested as the candidate control genes for the blue trait. It potentially displayed a novel control mechanism of the blue grain trait. Network of blue grain identified. Our research provides a good example of how to find out the translocated genes in a particular trait. This study filled the several gaps that remained unrevealed on anthocyanin biosynthesis of blue wheat and will favour the development of new biotechnological tools for crop improvement.

ACKNOWLEDGMENTS

The authors are grateful to the China Government Scholarship Council.

Author Contribution

DCJ and WZ conceptualized and designed the study. DCJ and JW performed the experiments. DCJ, MDLCN, JW and LT analyzed and interpret the data. DCJ and WZ contributed in drafting the manuscript and WZ critically revised the manuscript.

REFERENCES

- Abdel-Aal el SM, Young JC and Rabalski I 2006 Anthocyanin composition in black, blue, pink, purple, and red cereal grains. *J Agric Food Chem*, 54:4696-4704.
- Abdelaal EM and Hucl P 2003 Composition and stability of anthocyanins in blue-grained wheat. *Journal of Agricultural and Food Chemistry*, 51:2174-2180.
- Ahn JH, Kim J-S, Kim S, Soh HY, Shin H, Jang H, Ryu JH, Kim A, Yun KY and Kim S 2015 De novo transcriptome analysis to identify anthocyanin biosynthesis genes responsible for tissue-specific pigmentation in zoysiagrass (*Zoysia japonica* Steud.). *PloS one*, 10:e0124497.
- Akhunov E, Nicolet C and Dvorak J 2009 Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet*, 119:507-517.
- Almeida JR, D'Amico E, Preuss A, Carbone F, de Vos CH, Deiml B, Mourgues F, Perrotta G, Fischer TC, Bovy AG, Martens S and Rosati C 2007 Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria xanassa*). *Archives of biochemistry and biophysics*, 465:61-71.
- Anders S, Pyl PT and Huber W 2015 HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31:166-169. doi:10.1093/bioinformatics/btu638
- Bharti AK and Khurana JP 1997 Mutants of *Arabidopsis* as tools to understand the regulation of phenylpropanoid pathway and UV-B protection mechanisms. *Photochem Photobiol*, 65:765-776.
- Carolyn AW and Mark DR 2012 Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany*, 63: 5741-5749.
- Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S et al 2013 Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America*, 110:8057-8062.
- Crasta OR, Francki MG, Bucholtz DB, Sharma HC, Zhang J, Wang RC et al 2000 Identification and characterization of wheat-wheatgrass translocation lines and localization of barley yellow dwarf virus resistance. *Genome*, 43: 698-706.
- Dubcovsky J, Luo M, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A and Dvorak J 1996 Genetic map of diploid wheat, *Triticum monococcum* L., and its comparison with maps of *Hordeum vulgare* L. *Genetics*, 143:983-999.
- Escribanobailon MT, Santosbuelga C and Rivasgonzalo JC 2004 Anthocyanins in cereals. *Journal of Chromatography A*,

- 1054:129-141.
- Ficco DBM, Mastrangelo AM, Trono D, Borrelli GM, De Vita P, Fares C, Beleggia R, Platani C and Papa R 2014 The colours of durum wheat: a review. *Crop and Pasture Science*, 65:1-15.
- Holton TA and Cornish EC 1995 Genetics and Biochemistry of Anthocyanin Biosynthesis. *The Plant Cell*, 7:1071-108.
- Jia Y, Bin L, Mei M, Ping Z, Zhong L and Sheng L 2001 Identification of Blue-grained Wheat Translocation Lines Using Fluorescence in Situ Hybridization. *Acta Botanica Sinica*, 43:164-168.
- Jiang W, Liu T, Nan W, Jeewani DC, Niu Y, Li C, Wang Y, Shi X, Wang C, Wang J, Li Y, Gao X and Wang Z 2018 Two transcription factors *TaPpm1* and *TaPpb1* co-regulate anthocyanin biosynthesis in purple pericarps of wheat. *Journal of Experimental Botany*, 69:2555-2567.
- Keppen VD and Baenziger PS 1990 Inheritance of the blue aleurone trait in diverse wheat crosses. *Genome*, 33:525-529.
- Knott DR, 1989 The effect of transfers of alien genes for leaf rust resistance on the agronomic and quality characteristics of wheat. *Euphytica*, 44: 65-72.
- Kong J, 2003 Analysis and biological activities of anthocyanins. *Phytochemistry* 64:923-933.
- Lai B, Hu B, Qin YH, Zhao JT, Wang HC and Hu GB 2015 Transcriptomic analysis of Litchi chinensis pericarp during maturation with a focus on chlorophyll degradation and flavonoid biosynthesis. *BMC genomics*, 16:225.
- Li W and Beta T 2011 Evaluation of antioxidant capacity and aroma quality of anthograin liqueur. *Food chemistry*, 127:968-975.
- Liu D, Li S, Chen W, Zhang B, Liu D, Liu B and Zhang H 2016 Transcriptome Analysis of Purple Pericarps in Common Wheat (*Triticum aestivum* L.). *PLoS One*, 11:e0155428.
- Meuwissen THE, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819-1829.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z and Costich DE et al 2009 Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*, 21:2194-2202.
- Rafalski A, 2002 Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5: 94-100.
- Ramsay N and Glover BJ 2005 MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, 10:63-70.
- Ross JA and Kasum CM 2002 Dietary flavonoids: bioavailability, metabolic effects, and safety. *Annual review of nutrition*, 22:19-34.
- Schmittgen TD and Livak KJ 2008 Analyzing real-time PCR data by the comparative C T method. *Nature Protocols*, 3:1101-1108.
- Shen Y, Shen J, Dawadondup, Zhuang L, Wang Y, Pu J, Feng Y, Chu C, Wang X and Qi Z 2013 Physical localization of a novel blue-grained gene derived from *Thinopyrum bessarabicum*. *Molecular Breeding*, 31:195-204.
- Shirley BM, Kubasek WL, Storz G, Bruggemann E and Koornneef M 1995 Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. *Plant J*, 8:659-671.
- Shoeva OY and Khlestkina EK 2015 The Specific Features of Anthocyanin Biosynthesis Regulation in Wheat. In: *Advances in wheat genetics: From genome to field*. Proceedings of The 12th international wheat genetics symposium, SpringerTokyo, pp 147-157.
- Singh K, Ghai M, Garg M, Chhuneja P, Kaur P, Schnurbusch T, Keller B and Dhaliwal HS 2007 An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* × *T. monococcum* RIL population. *Theoretical and Applied Genetics*, 115:301-312.
- Sun X, Bai G and Carver BF 2009 Molecular markers for wheat leaf rust resistance gene *Lr41*. *Molecular Breeding*, 23:311-321.

- Tanksley SD, Ganai MW, Prince JP, De Vicente MC, Bonierbale MW and Broun P et al, 1992
- High-density molecular linkage map of the tomato and potato genomes. *Genetics*, 132:1141-1160.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ and Pachter L 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28:511-515.
- Trojan V, Musilová M, Vyhnánek T, Klejdus B, Hanáček P and Havel L 2014 Chalcone synthase expression and pigments deposition in wheat with purple and blue coloured caryopsis. *Journal of Cereal Science*, 59:48-55.
- Wang JK, Li HH, Zhang LY and Meng L 2012 QTL IciMapping version 3.2. <http://www.isbreeding.net> (accessed on 2016/12/16).
- Wang S, Wong D, Forrest KL, Allen AM, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG and Cattivelli L 2014 Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant biotechnology journal*, 12:787-796.
- Wang Z, Gerstein M and Snyder M 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10:57-63.
- Wolf JB 2013 Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular ecology resources*, 13:559-572.
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L and Wang J 2006 WEGO: a web tool for plotting GO annotations. *Nucleic acids research* 34 (Web Server issue):W293-297. doi:10.1093/nar/gkl031
- Zeven AC 1991 Wheats with Purple and Blue Grains—a Review. *Euphytica*, 56:243–258.
- Zhang X, Han D, Zeng Q, Duan Y, Yuan F and Shi J et al. 2013. Fine Mapping of Wheat Stripe Rust Resistance Gene Yr26 Based on Collinearity of Wheat with *Brachypodium distachyon* and Rice. *Plos One*, 8(3): e57885.