DOI: https://doi.org/10.4038/ta.v167i4.101 *Tropical Agriculturist, Vol.167 (4), 2019*

High throughput mapping and transcriptome analyses revealed the controlling gene of blue wheat grains

D.G.C.Jeewani[1,2], M. D. L. C. Nishantha[1], J. Wenhui[1], L. Tianxiang[1], N. Wenzhi[1] and W. Zhonghua[1]

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

[2]*Grain Legumes and Oil Crops Research and Development Center, Department of Agriculture, Angunakolapelessa, Sri Lanka*

Abstract

Blue grained wheat possesses a high amount of natural anthocyanin compounds with high antioxidant activity while the favorable characteristics which are not present in normal commercial wheat varieties. At present, major genes related to several colors of wheat have been identified. However, the major genes or control mechanism of anthocyanin biosynthesis in blue wheat is still unrevealed. Hence, combining the information of the chromosomal location of blue grain trait, and genes screened out from transcriptome analysis, we attempt to find the pivotal genes regulating blue grain trait in wheat. SNP genotyping was carried out in an F_2 blue and white wheat population. The results showed that blue grain was controlled by a gene/locus located between two SNP markers *IWB18525* and *IWB16381* on 4D chromosome. Comparative transcriptome analysis of two blue and eight non-blue wheat grain samples was carried out to identify the candidate genes controlling the blue grain trait. Forty 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 grain samples while two transcription factor DEGs were specific to blue wheat. These candidate genes were distributed in different chromosomes. Among them, 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. Meanwhile, F3'5'H is the main enzyme required for the production of Delphinidin compounds, which give the blue coloration. This study reports two candidate genes encoding F3'5'H in 4D chromosome responsible for the blue pigmentation in wheat.

Key words: Anthocyanin, Blue wheat, F3'5'H, SNP genotyping, Structural genes, Transcriptome

Introduction

Common wheat cultivars usually produce either red or white colored grains, but some cultivars possess either purple or blue grains. Purple and blue grains are caused by different anthocyanins, which are deposited in pericarp (purple) or aleurone layer (blue) (Trojan *et al*., 2014). Due to its high anthocyanin compounds, which contribute positively to the human health, specially 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 has paid a great attention on blue or purple wheat recently. It has been shown that the purple color was caused by the deposition of cyanidin 3-glucoside, followed by peonidin 3-glucoside in pericarp of purple caryopses, and blue color was caused by delphinidin 3-glucoside in aleurone layer of blue caryopses during caryopsis development (Abdelaal and Hucl, 2003; Escribanobailon *et al*., 2004).

Anthocyanin biosynthesis is initiated with the chalconesynthase (CHS) enzyme that exploits a p-coumaroyl-coA and three malonyl-coAsto form chalcone, which is subsequently transformed to naringenin by chalconeisomerase (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 color differences among anthocyanin pigments (Ahn *et al*., 2015). Accordingly, the resulting dihydro flavonols act as immediate precursors for the synthesis of leuco-anthocyanidins and flavonols by dihydro flavonol 4-reductase (DFR) and flavonol synthase (FLS), respectively. These colorless lecuo-anthocyanidins are transformed by antho-cyanidinsynthase (ANS) to colored antho-cyanidins. Subsequently, these antho-cyanidins are converted by glycosyl transferases (GTS) to anthocyanin pigments (Holton and Cornish, 1995).

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

Red or purple color of wheat is controlled by MYB or bHLH transcription factors. It is still unknown if the blue seed color is similarly controlled by transcription factors or differently. For the huge genome size and gene redundancy, map-based cloning and functional characterization of 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 wheat 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). Newly discovered high-throughput SNP genotyping platform has made 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).

The objectives of this study were to identify the candidate regions and related molecular markers for wheat blue grain trait using a high throughput SNP mapping, to identify the expression profiles of anthocyanin biosynthesis related genes by transcriptome analysis between blue and white wheat at two different physiological stages of color development and to find out candidate genes responsible for the blue grained wheat.

Materials and methods

Plant materials

An F_2 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 2015. The altitude of the area was 525 m and the climate was semi-humid prone to semi-arid with the mean temperature of 13^0C and average annual rainfall of 600 mm.

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

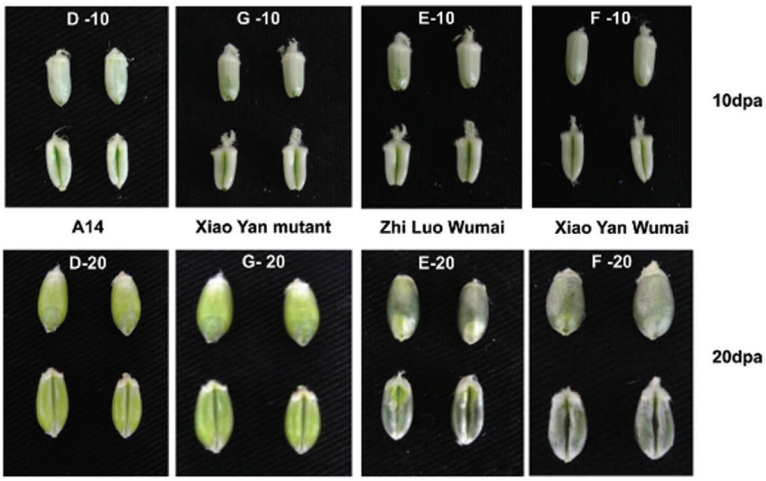

Figure 1. Wheat seeds used for the transcriptome analysis.

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: 'Zhiluowumai' at 10 dpa, E-20: 'Zhiluowumai' 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 'Zhiluowumai' and 'No.4045' parental lines and 101 individuals of the F_2 population derived from 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 iScanreader 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 were referred to a 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 ('Zhiluowumai' 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 RNase 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 manufacture's protocol. RNA degradation and contamination was tested 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 Bio analyzer 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 nebnext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. The various steps including mRNA enrichment, fragmentation, second strand cDNA synthesis, PCR amplification, size selection, clustering and sequencing, using Illumina hiseq^{tm} 2500 platform were performed by the Novogene Bioinformatics Technology Co. Ltd, China. Finally, 150 bp paired-end reads were generated.

Sequencing data processing and annotation

After completion of the sequencing, resulted image data were transformed to raw reads which were stored with an 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. 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. TopHat was used as the mapping tool as it was able to generate a database of splice junctions based on the gene model annotation file and thus a better mapping result could be obtained 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 kilo base 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 the 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 has the ability to verify the bias of gene length. GO terms with corrected p value less than 0.05 was 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 for a few randomly selected DEGs that were prepared from the total RNA. T NA from seed coat of all 4 varieties at two different physiological stages were extracted as described above and cDNA was synthesized using the Prime-Script™ RT Reagent kit with gDNA Eraser (TaKaRa, China). The gene specific real time PCR primers were designed using the Primer 5.0 program. A PCRs were carried out on -qRT Step One Plustm real-Time PCR System (USA) in a 20µl reaction vloume containing 1:10 diluted cDNA templates, 0.4µm 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 15s, 60°C for 45 s and 72°C for 15s. The relative gene expression level of each target gene was analyzed using the comparative CT method ($2^{-\triangle\triangle C}$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 color is produced by anthocyanin residing 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 into several chromosomes (Shoeva and Khlestkina, 2015; Liu *et al*., 2016). However, the genes those 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. The wheat genomes 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 4D chromosome was identified as controlling the blue color 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 as *Ba* (Figure 2). A high-density SNP map had not been used for mapping blue grain genes in wheat so far. Furthermore, blue aleurone trait in hexaploid wheat has not been mapped previously using any other mapping methods and this research reports it for the first time.

Table 1. Description of the identified SNP markers which 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

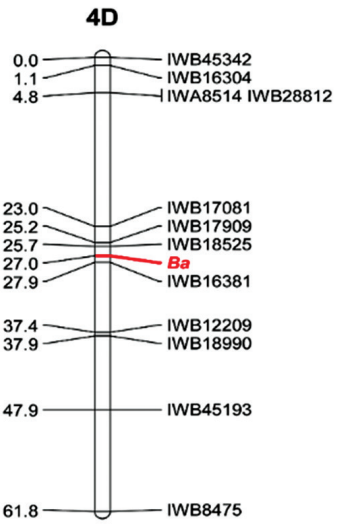

Figure 2. A diagrammatic representation of the 4D chromosome, showing location of the locus/gene for blue grain color in F_2 ('Zhiluowumai' [blue] x 'No.4045' [white]) mapping population of wheat

screening candidate genes involved in the blue grains

With the seed development, major color changes were visually observed in both blue lines by changing the color from green to blue at 20 dpa while both white lines remained green in color. 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 blue line. The second selection was to exclude the differentially expressed genes (DEGs) commonly occurred in all 4 lines. Therefore, the samples collected at 10 dpa and 20 dpa were used for screening candidate genes with transcriptome analysis. As a result, two blue grain samples and 6 non-blue samples were obtained 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 a transcriptome analysis. It is critical to screen the unrelated DEGs. As we know that the blue grain color 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 92.5% (272,788,618) and 92.2% (94,825,046) for non-blue and blue groups, respectively. These data were used for further bio-informatic analysis to screen out differential expressed genes between the groups of blue and non-blue samples.

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					

A total of 126,792 genes were found in blue and white wheat pools. Gene expression differences were assessed by pair wise 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 (Figure 3).

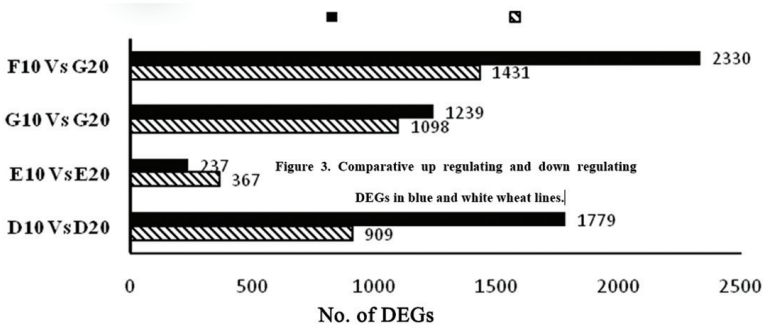

Figure 3. Comparative up regulating and down regulating DEGs in blue and white wheat lines.

Figure 3. Comparative up regulating and down regulating 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

A total of 4946 DEGs was screened out as relatively high level expressed (FPKM value>1) genes in the whole data pools, out of which 376 DEGs were commonly expressed in the all four lines (Figure 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 (Figure 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 (Figure 4D). Among the 124 commonly expressed DEGs in both blue lines, 75 DEGs were up-regulated and 49 DEGs were down-regulated. The most available up-regulated DEGs in blue wheat pools included NAC transcription factor, chalconesynthase, and hydroxylase and serpin domain containing protein.

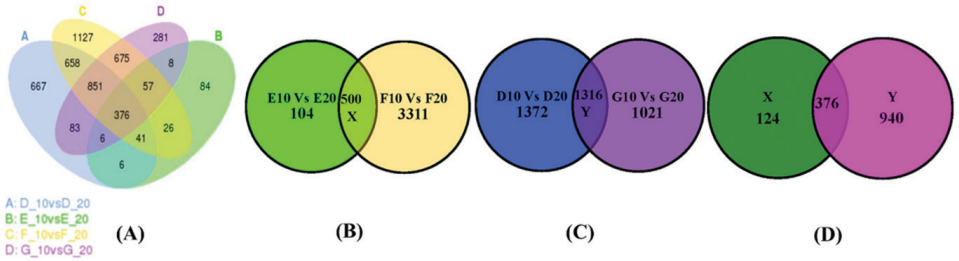

Figure 4. Venn diagrams of DEGs

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

GO annotation was done for the blue specific 124 DEGs and they were further categorized into three main categories namely biological process, cellular component and molecular function (Figure 5). Present study obtained 37 clusters, which consist of 17, 9 and 11clusters for biological process, molecular function and cellular component, respectively. The top most 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).

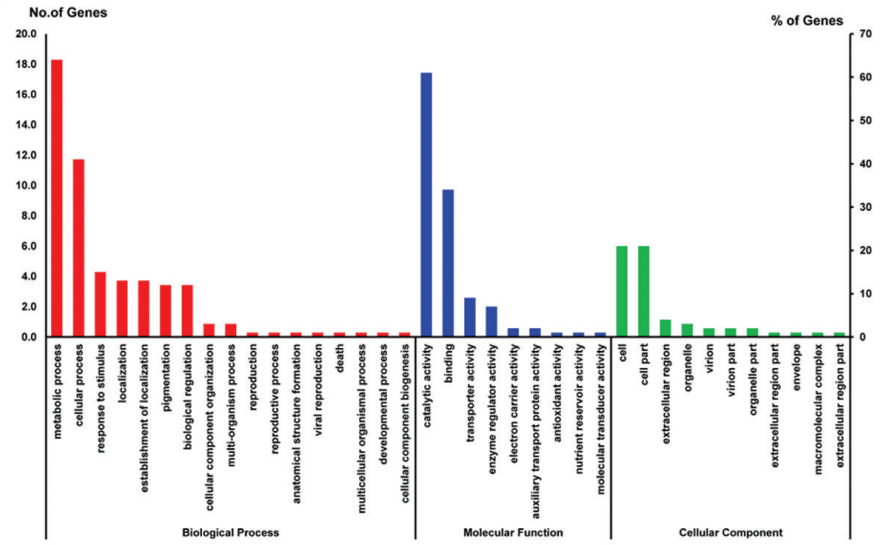

Figure 5. Gene Ontology classification of DEGs in blue wheat gene pool compared to white wheat

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

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 (Koes *et al*., 2005).

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 8.6 to 23.8 times higher expression level in 20 dpa samples than in 10 dpa samples.

Candidate structural genes involved in anthocyanin biosynthesis

Based on the transcriptome analysis, 40 structural DEGs involved in anthocyanin biosynthesis pathway were identified (Figure 6)

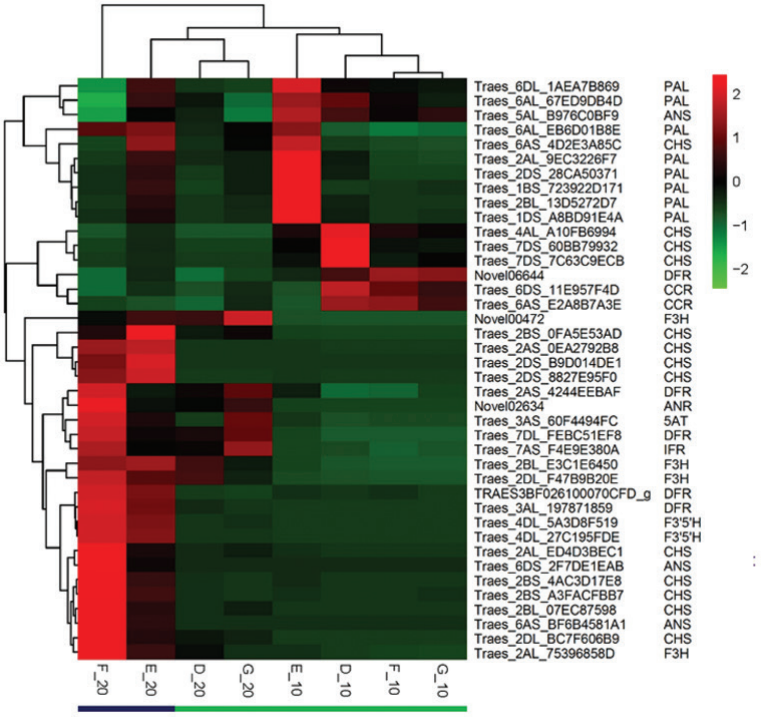

Figure 6. Heat map of the expression of structural genes related to flavonoid biosynthesis in blue and non-blue samples

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 color of the seed at that moment

Among the identified transcripts, 12 DEGs were recorded as highly up-regulated at 20 dpa only in blue wheat (Table 3). Those candidate genes included 5 CHS, two F3'5'H genes, two DFR genes, two ANS genes and one Anthocyanin 5-aromatic acyl transferase gene.

Table 3. Highly up-regulated DEGs identified only in blue wheat at 20 dpa

Name of the gene	DEG
CHS	Traes_2AS_0EA2792B8
	Traes_2BL_07EC87598
	Traes_2BS_4AC3D17E8
	Traes_2AL_ED4D3BEC1
	Traes_2DS_8827E95F0
F3'5'H	Traes_4DL_27C195FDE
	Traes_4DL_5A3D8F519
DFR	Traes3BF026100070CFD_g
	Traes_3AL_197871859
ANS	Traes_6AS_BF6B4581A1
	Traes_6DS_2F7DE1EAB
Anthocyanin 5-Aromatic Acyl transferase	Traes_3AS_60F4494FC

Of the five CHS genes 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 coloration 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 genes of Traes_6AS_BF6B4581A1, Traes_6DS_2F7DE1EAB and one Anthocyanin 5-aromatic acyl transferase gene (Traes_3AS_60F4494FC) showed very high expression in both blue samples while not expressed in other samples. No alternative splicing and SNPs was detected in the three genes.

To confirm the expression levels of these genes involved in 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 (Figure 7).

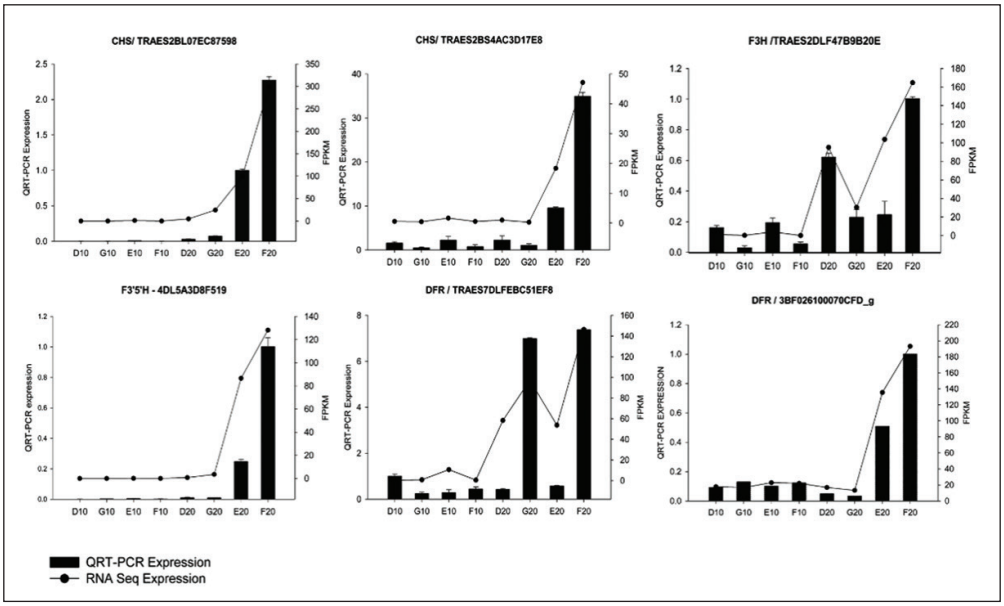

Figure 7. qRT-PCR validation of differential gene expression in anthocyanin biosynthesis

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.

The flavonoid biosynthetic pathway has been extensively studied at the genetic, molecular and biochemical levels in various plant species, including maize, snapdragon, Arabidopsis (Shirley *et al*., 1995; Bharti and Khurana, 1997), petunia (Holton and Cornish, 1995), strawberry (Almeida *et al*., 2007) and litchi (Lai *et al*., 2015). 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 colors of bread wheat were identified and quantified (Abdel-Aalel *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 the blue coloration (Figure 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.

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. The SNP genotyping performed using wheat 90k SNP chips, identified the genetic location of the blue trait in 4D chromosome. It has been reported that purple grain was regulated by transcription factors. 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.

There are many useful genes such as several disease resistant genes *Sr31* and *Sr38* resistant to stem rust (Knott, 1989), Bdv2 gene resistance to barley yellow dwarf virus (Crasta *et al*., 2000), *Lr41* resistance toleaf rust (Sun *et al*., 2009), Yr26 gene resistant to all important races of *Pucciniastriiformis* f. sp. *Tritici* (Zhang *et al*., 2013) are transferred from distant hybridization and many elite lines are translocation lines in wheat.

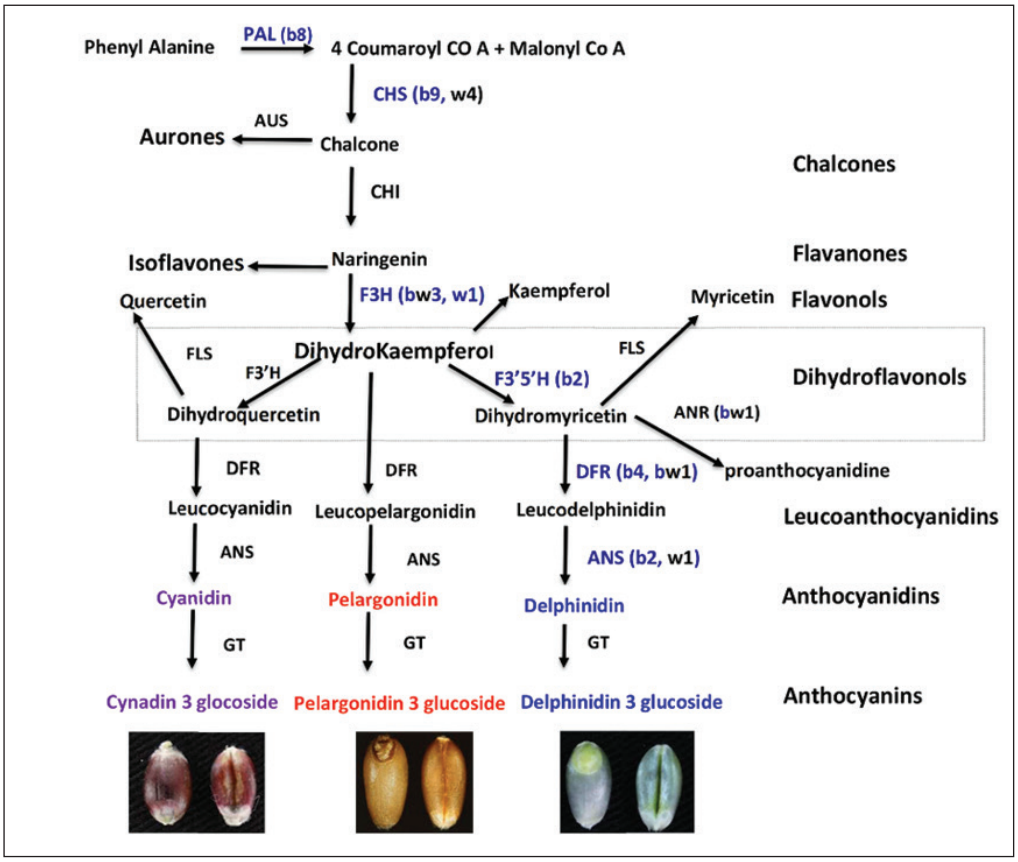

Figure 8. Simplified flavonoid biosynthesis pathway in blue and white wheats

Note: Numbers with "b" indicate number of DEGs identified in blue wheat and with "w" indicate number of DEGs identified in non-blue wheat. Numbers with "bw" indicate 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.

It causes a very difficult problem that the inserted fragment usually unpaired in meiosis and as a result, no cross over events will occur in the fragments. Due to this reason the fine mapping is impossible in this area and map based cloning of the gene is not practicable. It raises a common problem in wheat in cloning the particular translocated

gene. The present study provides a good solution to the problem. By mapping, location of a particular gene can be identified and by transcriptome analysis the genes related to that chromosome and trait can be confirmed. This could be applied for similar cases.

The blue wheat lines 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 mapping results with wheat consensus map, it can be confirmed that they are truly translocated lines. Further studies as the functional analysis of the F3'5'H genes have to be undertaken. Transgenic expression assay of these genes should provide a clear evidence for understanding molecular mechanism of blue grain trait in wheat.

Conclusions

The SNP genotyping performed using 90k iSelect Infinium assay, identified the location of the gene/locus, which controls the wheat blue grain trait in 4D chromosome. The transcriptome analysis revealed that 40 structural DEGs were related to anthocyanin biosynthesis and had obvious expression differences between blue and white wheat. Among them, 12 DEGs expressed only in blue wheat at 20 dpa, highlighted the importance of them to produce blue coloration. 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 coloration 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 the 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. Present study identified the blue grain network. This research provides a good example how to find out the translocated genes in a particular trait. Further, it filled the several gaps remained unrevealed on anthocyanin biosynthesis of blue wheat and will favor the development of new biotechnological tools for the crop improvement.

Acknowledgments

This study was financially supported by the Science and Technology Innovation Team Project of Shaanxi Province, China (grant number: 2014KCT-25), and the Cyrus Tang breeding fund. Authors express gratitude to both projects and China government Scholarship Council.

References

Abdel-Aalel, S.M., J.C.Young and I. Rabalski. 2006. Anthocyanin composition in black, blue, pink, purple, and red cereal grains. Journal of Agricultural and Food Chemistry.54:4696-4704.

Abdelaal, E.M. and P. Hucl. 2003. Composition and stability of anthocyanins in blue-grained wheat. Journal of Agricultural and Food Chemistry. 51:2174-2180.

Ahn, J.H., J.S. Kim, S. Kim, H.Y. Soh, H. Shin, H. Jang, J.H. Ryu, A. Kim, K.Y. Yun, S. Kim, K.S. Kim, D. Choi and J.H. Huh. 2015. De novotranscriptome analysis to identify anthocyanin biosynthesis genes responsible for tissue-specific pigmentation in zoysiagrass (*Zoysia japonica* Steud.). Plos one.10:e0124497.

Akhunov, E., C.Nicolet and J. Dvorak. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina goldengate assay. Theoritical and Applied Genetics. 119:507-517.

Almeida, J.R., E.D. Amico, A. Preuss, F. Carbone, C.H. De Vos, B. Deiml, F. Mourgues , G. Perrotta,T.C. Fischer, A.G. Bovy, S. Martens and C. Rosati. 2007. Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria xananassa*). Archives of biochemistry and biophysics. 465:61-71.

Anders, S., P.T. Pyl and W. Huber. 2015. Htseq-a Python framework to work with high-throughput sequencing data. Bioinformatics. 31:166-169.

Bharti, A.K and J.P. Khurana. 1997. Mutants of Arabidopsis as tools to understand the regulation of phenylpropanoid pathway and UV-B protection mechanisms. Photochemistry and Photobiology 65:765–776.

Crasta, O. R., M.G. Francki, D.B. Bucholtz, H.C. Sharma, J. Zhang, R.C. Wang, H.W. Ohm and J.M. Anderson. 2000. Identification and characterization of wheat-wheatgrass translocation lines and localization of barley yellow dwarf virus resistance. Genome. 43: 698-706.

Dubcovsky, J., M. Luo, G.Y. Zhong,R. Bransteitter, A. Desai, A. Kilian, A. Kleinhofs and J. Dvorák. 1996. Genetic Map of Diploid Wheat, *Triticum monococcum* L., and Its Comparison With Maps of Hordeum vulgare L. Genetics. 143:983-999.

Escribanobailon, M.T., C. Santosbuelga and J.C. Rivasgonzalo. 2004. Anthocyanins in cereals. Journal of Chromatography A. 1054:129-141.

Ficco, D.B.M., A.M.Mastrangelo, D.Trono, G.M. Borrelli, P. De Vita, C. Fares, R. Beleggia, C. Platani and R.Papa. 2014. The colours of durum wheat: a review. Crop and Pasture Science. 65:1-15.

Holton, T.A. and E.C. Cornish. 1995. Genetics and Biochemistry of Anthocyanin Biosynthesis. The Plant Cell. 7:1071-1083.

Jia, Y., L. Bin, M. Mei, Z. Ping, L. Zhong and L. Sheng. 2001. Identification of Blue-grained Wheat Translocation Lines Using Fluorescence in Situ Hybridization. Acta Botanica Sinica, 43:164-168.

Keppenne, V.D. and P.S. Baenziger. 1990. Inheritance of the blue aleurone trait in diverse wheat crosses. Genome. 33:525-529.

Knott, D. R. 1989. The effect of transfers of alien genes for leaf rest resistance on the agronomic and quality characteristics of wheat. Euphytica.44: 65-72.

Koes, R., W. Verweij and F. Quattrocchio. 2005. Flavonoids; a colorful model for the regulation and evolution of biochemical pathways. Trends in Plant Science. 10:236-242.

Kong, J. 2003. Analysis and biological activities of anthocyanins. Phytochemistry. 64:923-933.

Lai, B., B. Hu, Y.H.Qin, J.T.Zhao, H.C. Wang and G.B. Hu. 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 T. Beta. 2011. "Flour and bread from black, purple, and blue colored wheats," in Flour and Breads and their Fortification in Health and Disease Prevention, eds. V.R Preedy, R.R.Watson and V.B. Patel. Academic Press, London, UK, 59–67.

Liu, D., S. Li, W. Chen, B. Zhang, D. Liu, B. Liu and H. Zhang. 2016. Transcriptome Analysis of Purple Pericarps in Common Wheat (*Triticum aestivum* L.). Plos One. 11:e0155428.

Meuwissen, T.H.E., B.J.Hayes and M.E.Goddard. 2001. Prediction of Total Genetic Value usingGenome-Wide Dense Marker Maps. Genetics. 157:1819-1829.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich and E.S.Bukler. 2009. AssociationMapping: Critical Considerations Shift from Genotyping to Experimental Design. The PlantCell. 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 B.J. Glover. 2005. MYB-bhlh-WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science. 10:63-70.

Ross, J.A. and C.M. Kasum. 2002. Dietary flavonoids: bioavailability, metabolic effects, and safety. Annual review of nutrition. 22:19-34.

Shen, Y., J.Shen, L. Dawadondup, Y. Zhuang, J. Wang, J. Y. Pu, C. Peng, X, Chu, Wang and Z.Qi. 2013. Physical localization of a novel blue-grained gene derived from *Thinopyrum bessarabicum*. Molecular Breeding. 31:195-204.

Shirley, B.M., W.L. Kubasek, G. Storz, E. Bruggemann, and M. Koornneef. 1995. Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. The Plant Journal. 8:659–671.

Shoeva, O.Y. and E.K. Khlestkina. 2015. The Specific Features of Anthocyanin Biosynthesis Regulation in Wheat, in Advances in wheatgenetics: From genome to field, Proceedings ofThe 12th internationalwheat genetics symposium, eds. Y.Ogihara, S.Takumi & H.Handa,Springer,Tokyo. 147-157.

Singh, K., M. Ghai, M. Garg, P. Chhuneja, P. Kaur, T. Schnurbusch, B. Keller and H.S. Dhaliwal. 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., G. Bai and B.F. Carver. 2009. Molecular markers for wheat leaf rust resistance gene Lr41. Molecular Breeding. 23:311–321.

Tanksley, S.D., M.W.Ganal, J.P.Prince, M.C. De Vicente, M.W.Bonierbale,P. Broun.T.M. Fulton, J. J. Giovannoni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Paterson, O. Pineda, M. S. Roder, R. A. Wing, W. Wu and N. D. Young .1992. High density molecular linkage map of the tomato and potato genomes. Genetics. 132: 1141-1160.

Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, S.L. Salzberg, B.J. Wold and L. Pachter. 2010. Transcript assembly and quantification by RNA- Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 28:511-515.

Trojan, V., M.Musilová, T.Vyhnánek, B.Klejdus, P. Hanáček, and L. Havel. 2014. Chalcone synthase expression and pigments deposition in wheat with purple and blue colored caryopsis. Journal of Cereal Science. 59:48-55.

Wang, J.K., H.H. Li, L.Y. Zhang, and L. Meng. 2012. QTL icimapping version 3.2.http://www.isbreeding.net .Accessed on 16.05.2016.

Wang, S., D. Wong, K.L. Forrest, A.M. Allen, S. Chao, and B.E. Huang. 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., M. Gerstein and M. Snyder. 2009. RNA-Seq: a revolutionary tool fortranscriptomics. Nature Reviews Genetics. 10:57-63.

Wolf, J.B. 2013. Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Molecular ecology resources. 13:559-572.

Ye, J., L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolund and J. Wang. 2006. WEGO: a web tool for plotting GO annotations. Nucleic acids research. 34: 293-297.

Zeven, A.C. 1991. Wheats with Purple and Blue Grains—a Review. Euphytica. 56:243–258.

Zhang, X., D. Han, Q. Zeng, Y. Duan, F. Yuan, J. Shi, Q. Wang, J. Wu, L. Huang and Z. Kang. 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.