

# Transcriptome Analysis of MYB Genes and Patterns of Anthocyanin Accumulation During Seed Development in Wheat

Paulina Calderon Flores<sup>1</sup>, Jin Seok Yoon<sup>2</sup>, Dae Yeon Kim<sup>3</sup> and Yong Weon Seo<sup>1</sup> 

<sup>1</sup>Department of Plant Biotechnology, Korea University, Seoul, Korea. <sup>2</sup>Ojeong Plant Breeding Research Center, Korea University, Seoul, Korea. <sup>3</sup>Department of Biotechnology, Korea University, Seoul, Korea.

Evolutionary Bioinformatics  
Volume 18: 1–12  
© The Author(s) 2022  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/11769343221093341



**ABSTRACT:** Plants accumulate key metabolites as a response of biotic/abiotic stress conditions. In seed coats, anthocyanins, carotenoids, and chlorophylls can be found. They have been associated as important antioxidants that affect germination. In wheat, anthocyanins can impart the seed coat color which have been recognized as health-promoting nutrients. Transcription factors act as master regulators of cellular processes. Transcription complexes such as MYB-bHLH-WD40 (MBW) regulate the expression of multiple target genes in various plant species. In this study, the spatiotemporal accumulation of seed coat pigments in different developmental stages (10, 20, 30, and 40 days after pollination) was analyzed using cryo-cuts. Moreover, the accumulation of phenolic, anthocyanin, and chlorophyll contents was quantified, and the expression of flavonoid biosynthetic genes was evaluated. Finally, transcriptome analysis was performed to analyze putative MYB genes related to seed coat color, followed by further characterization of putative genes. *TaTCL2*, an MYB gene, was cloned and sequenced. It was determined that *TaTCL2* contains a SANT domain, which is often present in proteins participating in the response to anthocyanin accumulation. Moreover, *TaTCL2* transcript levels were shown to be influenced by anthocyanin accumulation during grain development. Interaction network analysis showed interactions with GL2 (HD-ZIP IV), EGL3 (bHLH), and TTG1 (WD40). The findings of this study elucidate the mechanisms underlying color formation in *Triticum aestivum* L. seed coats.

**KEYWORDS:** Colored wheat, MBW complex, MYB, purple wheat seed, transcription factors

**RECEIVED:** November 5, 2021. **ACCEPTED:** March 14, 2022.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a Cooperative Research Program for Agriculture Science and Technology Development (PJ015666) funded by Rural Development Administration and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R111A1A01053494).

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Yong Weon Seo, Department of Plant Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Korea. Email: seoag@korea.ac.kr

## Introduction

Domestication has made, through the selection, to narrow the genetic base and the variance for traits that are important for continued crop improvement.<sup>1</sup> Highlighting the breeding lines, “lost” traits and trait combinations that once were discarded by traditional breeding programs, can find unidentified traits that can, nowadays, be favorable for crop performance.<sup>1–3</sup>

Domestication of cereals made changes in cereal metabolites, such as anthocyanins. In wheat, barley, maize, and rice, most of the genotypes that can be found, have lost the capability of synthesizing anthocyanins in grain. However, due to the aim to improve health, the study of the anthocyanins in grains have gained importance to recreate genotypes that can accumulate anthocyanins and their precursors in the grain of cereals.<sup>4</sup>

On the other hand, seed coat color is an important trait that has been associated with resistance to abiotic and biotic stresses, such as grain dormancy,<sup>5</sup> chilling stress,<sup>6</sup> and preharvest sprouting.<sup>7</sup> Furthermore, deeper seed coat colors have been related to higher antioxidant activity<sup>6,8–11</sup> as well as recognized as health-promoting nutrients.<sup>10,12</sup> In wheat, the antioxidants responsible for seed coat color that are present in the highest levels are anthocyanins and proanthocyanidins.<sup>9,13,14</sup>

Seed coats develop from the integuments that surround the ovule prior to fertilization. Before fertilization, cells of the integuments are relatively undifferentiated. However, specialized cell types are being formed after fertilization due to an extensive cell differentiation.<sup>5</sup> The outermost layer of the endosperm is called the aleurone layer. These parts of the caryopsis are covered by testa (seed coat) fused with the pericarp.<sup>15,16</sup> Cell layers in the seed coats may accumulate large quantities of certain substances, such as mucilage or pigments that can also contribute to overall seed morphology.<sup>5</sup>

During grain development, purple anthocyanins are found only in the seed coat initially but are found in the pericarp later on in development,<sup>17,18</sup> whereas flavonols and PAs (proanthocyanins) are the major pigments in yellow and brown seeds. Flavonols are present in both the seed coat and the embryo. Colorless PAs that accumulate exclusively in the inner integument, are oxidized, and polymerized into brown pigments during seed maturation, causing the formation of brown seed coat color.<sup>19–22</sup>

The flavonoid pathway has been studied elsewhere.<sup>22–25</sup> The genes are divided in early anthocyanin biosynthesis genes (EBGs), and late anthocyanin biosynthesis genes (LBGs). EBGs include *chalcone synthase (CHS)*, *chalcone isomerase (CHI)*,



*flavanone 3-hydroxylase (F3H)*, *flavanone 3'-hydroxylase (F3'H)*, and *flavonol synthase (FLS)* which lead to the production of flavonols, whereas LBGs include *dihydroflavonol reductase (DFR)*, *anthocyanidin synthase/leucoanthocyanidin dioxygenase (ANS/LDOX)*, *UDP-flavonoid glucosyl transferase (UFGT)*, and *anthocyanidin reductase (ANR)* which lead to the production of proanthocyanidins and anthocyanins.<sup>22,26</sup>

Generally, the LBGs for anthocyanin biosynthesis is regulated by a MYB-bHLH-WD40 (MBW) complex.<sup>27</sup> The MBW complex has been reported to regulate the expression of flavonoid biosynthesis genes.<sup>28</sup> MYBs *PAP1* and *PAP2* in *Arabidopsis*,<sup>29</sup> *ZmC1* in maize,<sup>28</sup> and *MYB3* in petunias<sup>30</sup> interact with bHLHs such as *GL3*, *EGL3*, and *TT8*<sup>29</sup> to form a ternary complex with WD40, such as *TTG1* in *Arabidopsis*, thus affecting proanthocyanin accumulation in seeds and anthocyanin accumulation in plant species.<sup>28,29,31</sup>

In this study, we analyzed recombinant inbred lines (RILs) of wheat to get a deeper analysis of the pigmentation accumulation in seed coats of deep purple (DP) and yellow (Ye) seeds. Phenotypic and RNA sequencing was used to determine the transcriptome difference between DP and Ye wheat seeds and isolate a dominant gene for the deep purple grain trait. The findings of this study elucidate the mechanisms underlying color formation in *Triticum aestivum* L. seed coats.

## Materials and Methods

### *Plant materials and growth conditions*

RILs with different seed coat phenotypes, Ye (accession no. 10DS1673, Korea University wheat sub-gene bank) and DP (accession no. 10DS1674), were used.<sup>32</sup> Seeds were germinated on moistened filter paper at room temperature for 24 hours and vernalized at 4°C in a dark chamber for 4 weeks. Each seedling was then transplanted to a pot (5 × 5 × 16 cm) filled with soil (Sunshine mix #1, Sun Gro Horticulture, Vancouver, BC, Canada) in a well-controlled glasshouse at Korea University with a photoperiod of 16:8 hours (day:night) and temperatures between 20°C and 25°C. Spikes were harvested at the flowering stage at 10, 20, 30, and 40 days after pollination (DAP) in green house. Seeds of the same RILs were planted in the field at Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute (Jeongeup, Jeollabuk-do, Republic of Korea), and the spikes were harvested at the flowering stage at 10, 20, 30, and 40 DAP. Samples were immediately immersed in liquid nitrogen and stored at -80°C until further use. All samples were harvested in triplicate biological replicates.

### *Whole caryopses and cryo-cuts*

To observe anthocyanin deposition during seed development in Ye and DP seeds, cryo-cuts (Cryotome CM3050S, Leica, Wetzlar, Germany) were used for sectioning the seeds of the harvested Ye and DP at 10, 20, 30, and 40 DAP. The seeds were

embedded in optimum cutting temperature compound (Sakura® FineTek USA, Inc., Torrance, CA, USA). Cross-sections (0.04 mm) were prepared from the central part of the developing caryopses and observed with a Leica EZ4D microscope (Leica). Photographs were taken using Leica Acquire software (version 3.4.1).

### *Quantification of phenolics and anthocyanins*

Free phenolics (FP) and total phenolics (TP) contents were measured using Folin-Ciocalteu reagent using 20 mg of grain sample ground in liquid nitrogen.<sup>33</sup> Anthocyanin content was determined according to Hong et al<sup>34</sup> using 100 mg from each sample. The absorbance was read at 765 nm for FP and TP and at 530 and 657 nm for anthocyanins using a microplate reader (HIDEX-Sense 425-301, Hidex, Turku, Finland). The results are expressed as the average of 3 biological replicates of phenolics and anthocyanins.

### *Chlorophyll content*

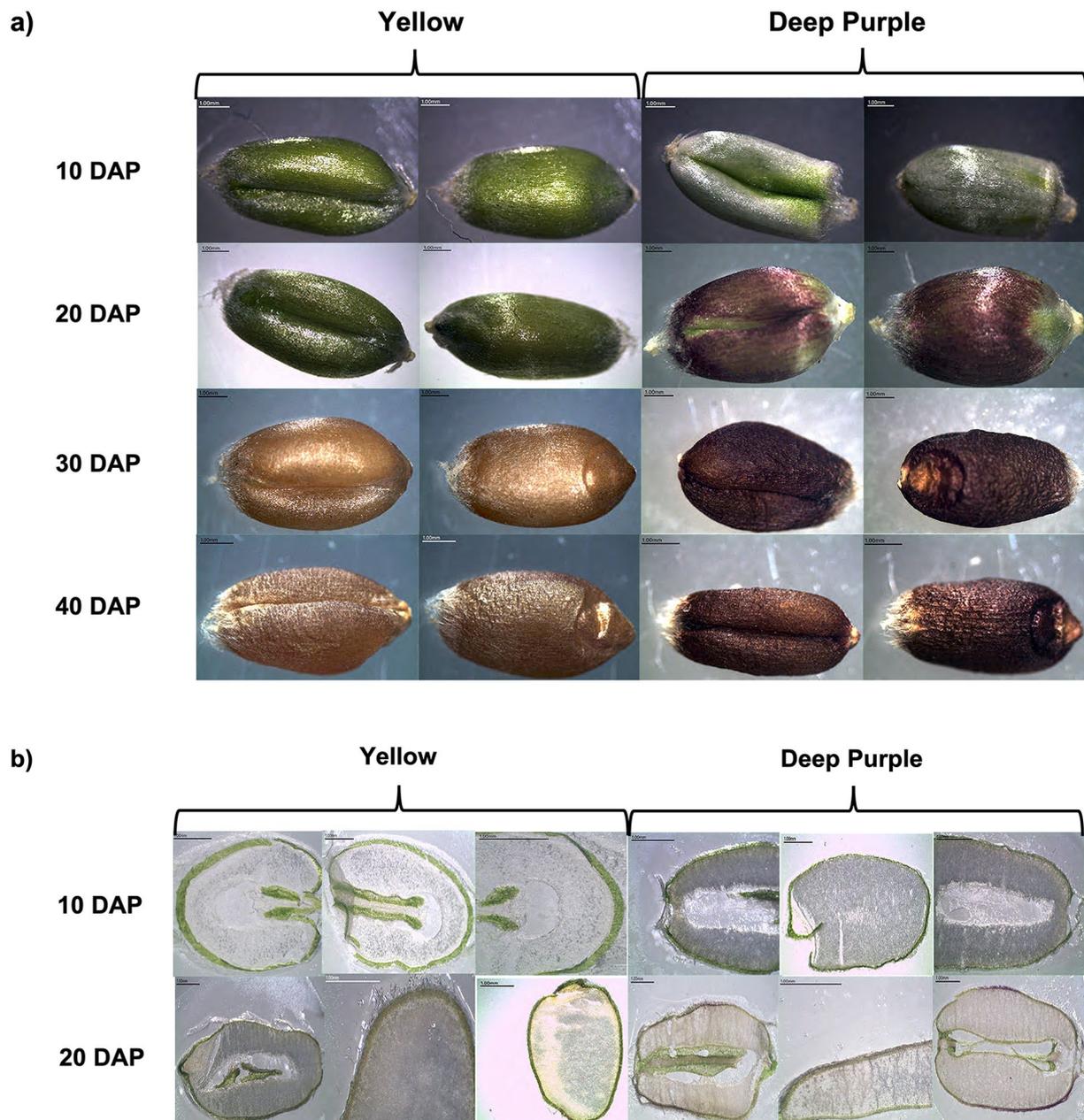
Chlorophyll content was measured using 20 mg of ground grains in liquid nitrogen following the protocol of Warren<sup>35</sup> with 3 biological replicates. The absorbance was measured at 652 and 665 nm for chlorophyll content using a microplate reader (HIDEX-Sense 425-301).

### *Gene expression analysis*

Total RNA was extracted in triplicate from samples taken during the seed development stages described in the sampling section. Seed samples were immediately frozen in liquid nitrogen and stored at -80°C until further use. Samples were ground in liquid nitrogen using a pestle and mortar, and RNA was extracted using the TRIzol method (Invitrogen, Waltham, MA, USA). cDNA was synthesized using the cDNA Takara PrimeScript™ 1st strand cDNA Synthesis Kit (Takara, Tokyo, Japan). Quantitative real-time PCR (qRT-PCR) was performed to analyze the flavonoid pathway genes during seed development of Ye and DP seeds. Gene-specific primers were designed using Primer-BLAST (NCBI, <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) or obtained from previous publications as described in Supplemental Table S1. qRT-PCR was performed in triplicate using EvaGreen 2X qPCR MasterMix (ABM, Milton, ON, Canada) on a CFX-96 RT-PCR machine (Bio-Rad, Hercules, CA, USA). *β-actin* (Accession no. AB181991) was used as an internal control. The 2<sup>-ΔΔCT</sup> method was used to calculate expression levels in fold changes as previously described.<sup>36</sup>

### *cDNA library construction and RNA sequencing*

Total RNA was extracted from whole grains using TRIzol reagent (Invitrogen, Waltham, MA, USA) and then used for



**Figure 1.** Images during grain development of the yellow (Ye; sample 1673) and deep purple (DP; sample 1674) parental seeds: (a) frontal and posterior sides at 10, 20, 30, and 40 days after pollination (DAP) in Ye and DP seeds and (b) cross-sections during grain development of Ye and DP parental seeds at 10 and 20 DAP.

library construction with the Ri-bospinTMII Kit (Geneall Biotechnology, Seoul, Korea) according to the manufacturer's instructions. A total of 9 RNA-seq paired-end libraries at 3 developmental stages (10, 20, and 30 DAP, Supplemental Figure S1) in DP and Ye seeds were constructed from each DAP sample consisting of 3 biological replicates each, using the SMARTer Stranded RNA-Seq Kit (Clontech Laboratories Inc., Mountain View, CA, USA) following the manufacturer's instructions. Each library was then loaded onto the Illumina HiSeq2000 platform, high-throughput sequencing was performed, and paired-end reads were generated. The high-quality reads were obtained after several steps of quality checks, including trimming, removal of adaptor/

primer, and low-quality reads using Trimmomatic v 0.35,<sup>37</sup> and extraction of sequence data with quality scores of  $Q \geq 20$  using SolexaQA.

#### Read mapping and annotation

The hexaploid common wheat (*Triticum aestivum* L.) draft reference genome sequence (IWGSC1+popseq.31) was directly downloaded from EnsemblPlants ([http://plants.ensembl.org/Triticum\\_aestivum/Info/Index](http://plants.ensembl.org/Triticum_aestivum/Info/Index)). To map our reads to the draft genome sequence, an index of the reference genome was constructed using hisat2 (<https://ccb.jhu.edu/software/hisat2/index.shtml>),<sup>38</sup> and paired-end clean reads

were aligned to the reference genome, again using *hsiast2*, with all parameters set to their default values. HTSeq v0.6.1 was used to count the reads mapped to each gene and to the exons of each gene.<sup>39</sup> Differential expression was analyzed for each treatment group using DESeq2,<sup>40</sup> and genes with an adjusted absolute log<sub>2</sub> fold change of <1 were delineated as differentially expressed genes (DEGs). DEGs were annotated by blasting (*blastx*) against the NCBI non-redundant protein database for the Poaceae family and against the UniProt protein sequence databases for rice, *Arabidopsis*, and *Brachypodium* (<http://www.uniprot.org/>), with an E-value cutoff of 10<sup>-4</sup>.

### Selection of putative genes

For the flavonoid biosynthetic pathway, using the *Arabidopsis*\_unipro annotations, we selected for DEGs related to the flavonoid pathway such as *CHS* (*chalcone synthase*), *CHI* (*chalcone isomerase*), *F3H* (*flavonol 3-hydroxylase*), *F3'H* (*flavonol 3'-hydroxylase*), *FLS* (*flavonol synthase*), *OMT1* (*O-methyltransferase 1*), *UGT* (*UDP-glycosyltransferase*), *DFR* (*dihydroflavonol reductase*), *LDOX* (*leucoanthocyanidin dioxygenase*), *ANR* (*anthocyanidin reductase*), *OMT* (*O-methyltransferase*), and *GT* (*glycosyltransferase*). Similarly, we selected the data to identify MYB TFs expressed during the grain developmental stages using *Poaceae*\_NR, *Rice*\_unipro, *Brachypodium*\_unipro, and *Arabidopsis*\_unipro databases. Heat maps were created using Gene-E (<https://software.broad-institute.org/GENE-E/>). Highly expressed transcripts at 20DAP DP seeds were selected for qRT-PCR analysis, as shown in Supplemental Table S2.

For the isolation of MYB putative genes, highly expressed putative genes specifically at 20DAP in DP seeds were selected from transcriptome analysis (Supplemental Table S2). Primers specific to these genes were designed after referencing the sequences in the Ensembl Plants database ([https://plants.ensembl.org/Triticum\\_aestivum/Info/Index](https://plants.ensembl.org/Triticum_aestivum/Info/Index)) or Plant Transcription Factor Database v5.0 (<http://plantfdb.gao-lab.org>). The specific primers are shown in Supplemental Table S1.

### Cloning and sequencing of putative MYB genes

The PCR fragments of genes were purified and cloned into pLUG-T Prime® TA-Cloning Vector (Intron Biotechnology, Seongnam, Korea). The vector and PCR-amplified product were mixed and ligated overnight at 4°C and transformed into *Escherichia coli* DH5α competent cells (YB Biotech, Taipei, Taiwan) using the manufacturer's instructions. The pLUG plasmids were sequenced from both ends at Bionics (Seoul, South Korea). The sequences were compared with genes in the GenBank database using the BLAST program from the National Center for Biotechnology Information (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

### Evolutionary and bioinformatics analysis

Sequences from the cloned genes were used in ExPasy software (<https://web.expasy.org/translate/>) to translate the coding sequence into amino acids. For the cluster analysis MYB genes, the amino acid sequences of Traes\_3DL\_30CF35BB3 were analyzed with BLASTP from NCBI. Among the putative MYB genes, MYB genes of monocot plants were selected for phylogenetic analysis. The phylogenetic tree was constructed with fast minimum evolution method.<sup>41</sup>

The conserved domains were analyzed with the Conserved Domains tool from NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). A specific interaction network was constructed using STRING 11.0b (<https://string-db.org>), as seen in Zhao et al.<sup>42</sup> with an option value >0.700, and both alignment and phylogenetic tree construction were performed in Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>); accession numbers used are shown in Supplemental Table S3. The phylogenetic tree was arranged using iTOL v6 (<http://itol.embl.de>) using the category of TFs controlling anthocyanin synthesis or experimental interactions with At4G16430, AtTT8, AtMYC2, and AtEGL3.<sup>42,43</sup>

### Statistical analysis and software

All data are presented as means from at least 3 biological replicates. Significant differences were subjected to analysis of variance and *t*-test using IBM® SPSS® Statistics for MAC version 25 (IBM Corp., Armonk, NY, USA). All tests were performed at 95%, 99%, and 99.9% confidence.

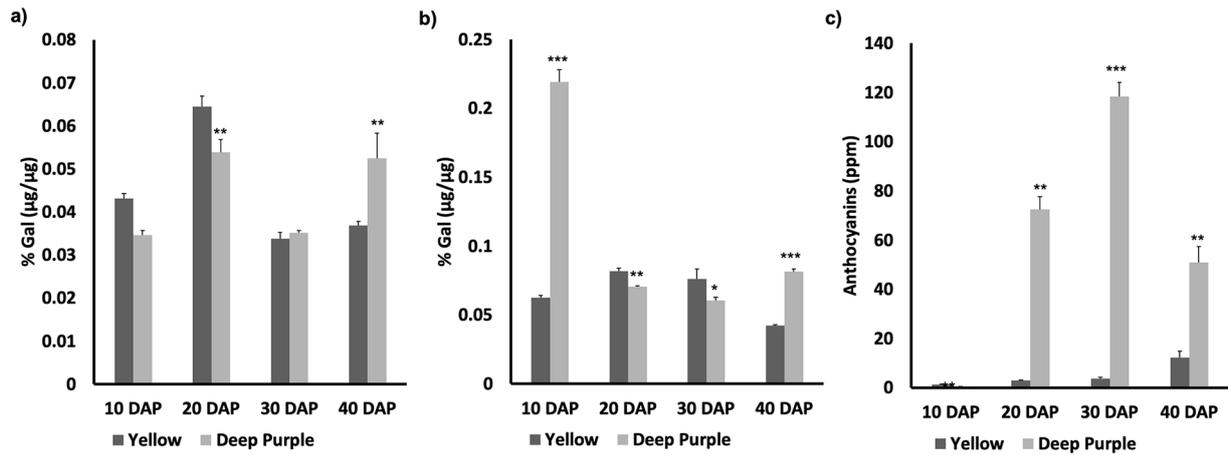
## Results

### Whole caryopses and cryo-cuts color deposition

At 10DAP, Ye and DP seeds were similar in color (Figure 1), indicating that chlorophyll had fully covered both seeds at this stage. Similarly, in the cryo-cuts at 10DAP, Ye and DP seeds were still in an early stage of development, and no color deposition was observed in either type. At 20DAP, Ye seeds exhibited little variation, whereas in DP seeds, a purple color had spread gradually almost throughout the seed coat from the pericarp to the aleurone layer. At 20DAP in Ye seeds, the cryo-cuts showed 2 layers with visible chlorophyll in the seed coat, whereas in DP, the seed coat began to develop a purple color on the top seed coat layer (pericarp) and inside (aleurone layer), and chlorophyll remained under the inner layer (aleurone) or, in some areas, completely disappeared. The seed coat of Ye and DP seeds clearly appeared yellow and purple at 30 and 40DAP, respectively.

### Phenolic, anthocyanin, and chlorophyll quantification

FP and TP contents were quantified during seed development. In both Ye and DP seeds, FP peaked at 20DAP. As seeds



**Figure 2.** Phenolic and anthocyanin contents during seed development: (a) free phenolics, (b) total phenolics, and (c) anthocyanin content in yellow (Ye) and deep purple (DP) seeds in different stages of seed development. DAP, days after pollination. Data are means  $\pm$  SEM of 3 biological replicates. Significant differences, evaluated by *t*-test, are indicated by \* $P \leq .05$ . \*\* $P \leq .01$ . \*\*\* $P \leq .001$  when comparing Ye versus DP within the DAP stage.

**Table 1.** Chlorophyll content during seed development.

	CHLOROPHYLL A ( $\mu\text{G ML}^{-1}$ FW)		CHLOROPHYLL B ( $\mu\text{G ML}^{-1}$ FW)		TOTAL CHLOROPHYLL ( $\mu\text{G ML}^{-1}$ FW)		CHLOROPHYLL A/ CHLOROPHYLL B	
	YE	DP	YE	DP	YE	DP	YE	DP
10 DAP	1.179 $\pm$ 0.057	3.021 $\pm$ 0.683 <sup>a</sup>	0.637 $\pm$ 0.087	1.433 $\pm$ 0.730	1.816 $\pm$ 0.050	4.454 $\pm$ 1.195	1.878 $\pm$ 0.315	2.388 $\pm$ 1.069
20 DAP	1.476 $\pm$ 0.010	0.616 $\pm$ 0.008 <sup>b</sup>	0.631 $\pm$ 0.324	0.478 $\pm$ 0.233	2.106 $\pm$ 0.332	1.094 $\pm$ 0.237	2.894 $\pm$ 1.694	1.642 $\pm$ 1.096
30 DAP	0.073 $\pm$ 0.007	0.074 $\pm$ 0.008	0.121 $\pm$ 0.123	0.223 $\pm$ 0.155	0.194 $\pm$ 0.115	0.297 $\pm$ 0.150	1.195 $\pm$ 0.973	0.727 $\pm$ 0.857
40 DAP	0.325 $\pm$ 0.099	0.166 $\pm$ 0.074	0.656 $\pm$ 0.029	0.370 $\pm$ 0.224	0.981 $\pm$ 0.099	0.536 $\pm$ 0.239	0.496 $\pm$ 0.156	0.583 $\pm$ 0.341

Chlorophyll content in yellow (Ye) and deep purple (DP) seeds in different stages of seed development. DAP, days after pollination. Data are means  $\pm$  SD of 3 biological replicates. Significant differences, evaluated by *t*-test, are indicated as <sup>a</sup> $P \leq .05$ , <sup>b</sup> $P \leq .001$  when comparing Ye versus DP within the DAP stage.

continued to develop, FP decreased in Ye seeds, whereas in DP seeds, there was a significant increase compared to Ye seeds at 40 DAP. TP showed a significant increase in DP seeds compared to that in Ye seeds. In contrast, TP was significantly reduced in DP seeds at 20 and 30 DAP. At 40 DAP, TP was reduced in Ye seeds and significantly increased in DP seeds (Figure 2a and b).

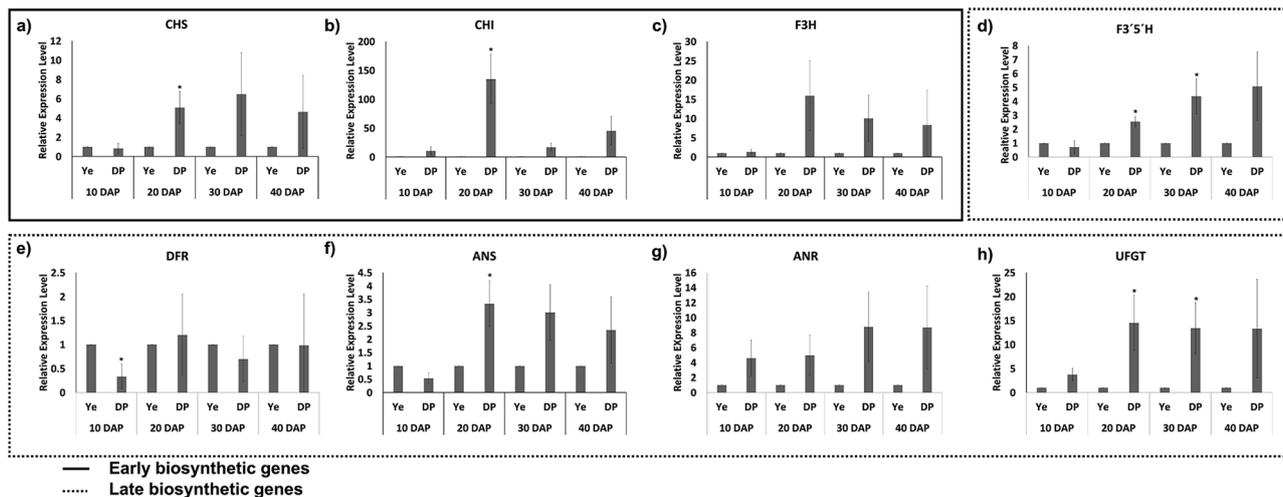
Similarly, the anthocyanin content was lower in Ye seeds than in DP seeds at the same stage. DP seeds had the highest anthocyanin content at 30 DAP and showed a reduction at 40 DAP; however, the content in all stages was significant compared to that in Ye seeds in the corresponding stages (Figure 2c).

We measured the chlorophyll (*Chl*) contents, *Chla*, *Chlb*, *total Chl*, and *Chla/Chlb* ratio during seed development in both Ye and DP seeds. During the early stages (10 and 20 DAP), the *Chl* contents were high in Ye and DP seeds, whereas at 10 DAP, *Chla*, *Chlb*, *total Chl*, and the *Chla/Chlb* ratio were higher in DP seeds than in Ye seeds. At 20 DAP, the Ye seeds had higher contents than DP seeds. As expected, 30 and 40 DAP had the lowest chlorophyll content in both Ye and DP seeds, as displayed in Figure 1, where chlorophyll cannot be seen.

Interestingly at 40 DAP in both Ye and DP seeds, there was an increase in *Chla*, *Chlb*, and *total Chl* contents, whereas the *Chla/Chlb* ratio was reduced (Table 1).

#### Flavonoid biosynthetic gene expression

We analyzed the flavonoid biosynthetic gene expression during color deposition in each seed developmental stage. As shown in Figure 3, the early biosynthetic genes (EBGs; *CHS*, *CHI*, and *F3H*) and late biosynthetic genes (LBGs; *F3'5H*, *DFR*, *ANS*, *ANR*, and *UFGT*) were mostly highly expressed in DP seeds at 20 (*CHI*, *F3H*, *DFR*, *ANS*, and *UFGT*) or 30 (*CHS* and *ANR*) DAP, with the exception of *F3'5H*, which was highly expressed at 40 DAP compared to that in Ye seeds. These results are consistent with our images of color deposition and total anthocyanin quantification (Figure 1). Moreover, *ANR* had similar expression levels during 10 and 20 DAP and during 30 and 40 DAP, whereas *F3'5H* had similar expression from 20 to 40 DAP. Furthermore, based on transcriptome analysis, we created a schematic representation and analyzed expression of flavonoid biosynthetic genes using the *Arabidopsis* unipro annotations in Ye and DP seeds during seed color deposition. The DEG names are listed next to the heat map in Figure 4.



**Figure 3.** Flavonoid-related genes during yellow (Ye) and deep purple (DP) seed development (10, 20, 30, and 40 days after pollination [DAP]). Genes are divided into early and late biosynthetic genes: (a) *CHS*: Chalcone synthase, (b) *CHI*: Chalcone isomerase, (c) *F3H*: Flavonone 3-hydroxylase, (d) *F3'H*: Flavonoid 3'5' hydroxylase, (e) *DFR*: Dihydroflavonol 4-reductase, (f) *ANS*: Anthocyanidin synthase, (g) *ANR*: Anthocyanidin reductase, and (h) *UFGT*: UDP-glucose flavonoid-3-O-glucosyltransferase. Data are means  $\pm$  SD of 3 biological replicates. Asterisks indicate significant differences when comparing Ye DAP in each developmental stage ( $*P \leq .05$ ).

The genes shown are *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, *OMT1UGT*, *DFR*, *LDOX*, *ANR*, *OMT*, and *GT*, which have been shown to be expressed in the flavonoid pathway.<sup>21</sup>

#### Transcriptome analysis and qRT-PCR expression levels

Based on transcriptome analysis (Figure 5a, Supplemental Tables S2, and S4), in general, more MYB TFs showed expression (up or down regulated) during the developmental process from 10 to 30 DAP in DP seeds, resulting in 60 MYB TFs expressed in total. More MYBs (>2 fold change) were detected at 10 and 20 DAP in Ye (DP10 DAP\_Ye10DAP and DP10DAP\_Ye20DAP) than in DP. Moreover, we selected genes for qRT-PCR analysis that showed high expression in DP20DAP\_Ye20DAP: Traes\_3DL\_30CF35BB3 (2.017 fold), Traes\_2BL\_EED456A17 (2.796 fold), Traes\_6BS\_44456AE22 (3.828 fold), and Traes\_2BL\_3041037F1 (4.077 fold). The qRT-PCR results (Figure 5b) revealed that among the genes studied, Traes\_3DL\_30CF35BB3 showed the highest expression in DP seeds at early stages (DP10DAP and DP20DAP) compared to the other genes. Thus, based on these results, we selected Traes\_3DL\_30CF35BB3 for further experiments.

#### Isolation of TaTCL2

Our sequence clone of Traes\_3DL\_30CF35BB3 was cloned using primers shown Supplemental Table S1. After sequencing, the sequence was subjected to BLAST in the NCBI database, which indicated that the PCR product had 99% identity with a predicted sequence of *Aegilops tauschii* subsp. *tauschii* MYB-like TF *TCL2* (LOC109750608), transcript variant X2, mRNA (sequence ID: XM\_020309568.1), which confirmed that the obtained gene corresponded to the MYB TF *TCL2*.

#### Phylogenetic analysis and sequence alignments

To determine the evolutionary relationship of TaTCL2, phylogenetic analyses were performed by fast minimum evolution algorithm. The phylogenetic tree revealed evolutionary distance between TaTCL2 and 28 MYB-like protein from monocot species (Supplemental Figure S2). TaTCL2 and 28 MYB-like protein showed various sequence similarity (>70% amino acid identity). Interestingly, TaTCL2 and HvTCL2 displayed a very high sequence similarity (>92% amino acid identity) (Supplemental Figure S2).

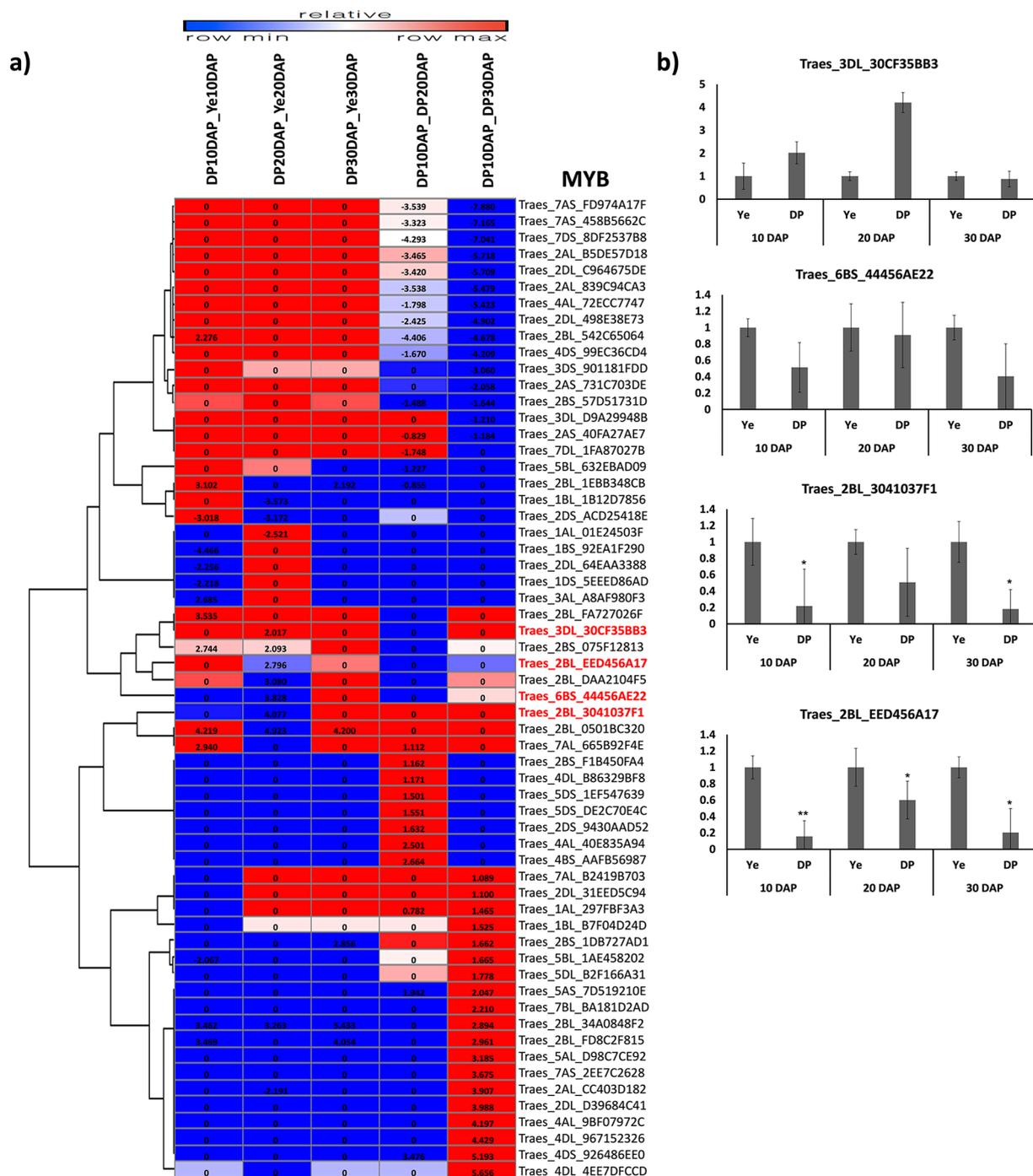
To examine the relationship between TaTCL2 and MYB TFs in other plants that regulated anthocyanin biosynthesis, amino acid sequences or translated products were downloaded from the NCBI database to construct a phylogenetic tree. TaTCL2 was closest to *AtTCL2*, *PhODO1*, *SCAN2like*, *AtMYB*, *FaMYB5*, *PhPH4*, *VvMYB5b*, and *VvMYBCS1* (Figure 6a).

Domain search using Pfam database showed that it was a MYB-like DNA binding domain. This family contains the DNA binding domains from MYB proteins, as well as the SANT domain family.<sup>44</sup> Moreover, our sequence showed the motifs PF00249 and PF13921, both from the SANT domain (a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIB). Likewise, both motifs were identified as MYB-like DNA-binding domains (Supplemental Table S5). Figure 6b shows that the SANT domain is also found in the genes that TaTCL2 was clustered with in our phylogenetic tree.

#### Network interaction analysis of TaTCL2 response to anthocyanin biosynthesis

The above results suggest a high possibility that the TaTCL2 gene is involved in the anthocyanin biosynthetic pathway for





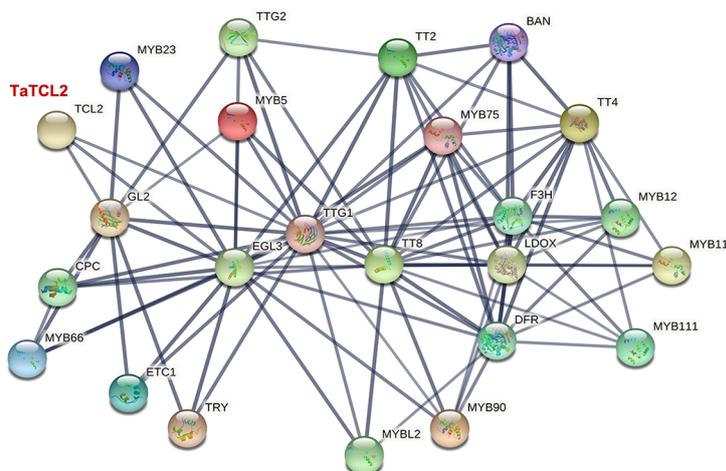
**Figure 5.** Heat map and transcript levels of MYB putative genes: (a) Heat map of the differentially expressed genes of MYB. A positive value indicates upregulation in the late stage; for example, in DP10DAP\_DP20DAP, the plus value of  $\log_2$ FoldChange means upregulation in DP20DAP stage. We used 4 annotation Databases: *Poaceae*\_NR, *Rice*\_unipro, *Brachypodium*\_unipro, and *Arabidopsis*\_unipro. Numbers represent the fold changes. Colors are based using row minimum and maximum values. The mean values were obtained from 3 biological replicates. (b) qRT-PCR transcription analysis of 4 selected MYB putative genes during seed coat color deposition in 3 different stages of deep purple (DP) and yellow (Ye) samples. Ten days after pollination (DAP), 20DAP, and 30DAP (early, middle, and late stages, respectively). Ye is used as an internal control within a stage. Data are means  $\pm$  SD of 3 biological replicates; (b) transcript levels of putative MYB genes. Data are means  $\pm$  SEM of 3 biological replicates. Significant differences, evaluated by *t*-test, are indicated by  $*P \leq .05$ , and  $**P \leq .01$  when comparing Ye versus DP within the DAP stage.

by Trojan et al<sup>15</sup> in purple and blue wheat seeds where in the early stages the chlorophyll pigmentation was covering the grains, whereas at 20 or 30 DAP the pigmentation started to be distributed uniformly.

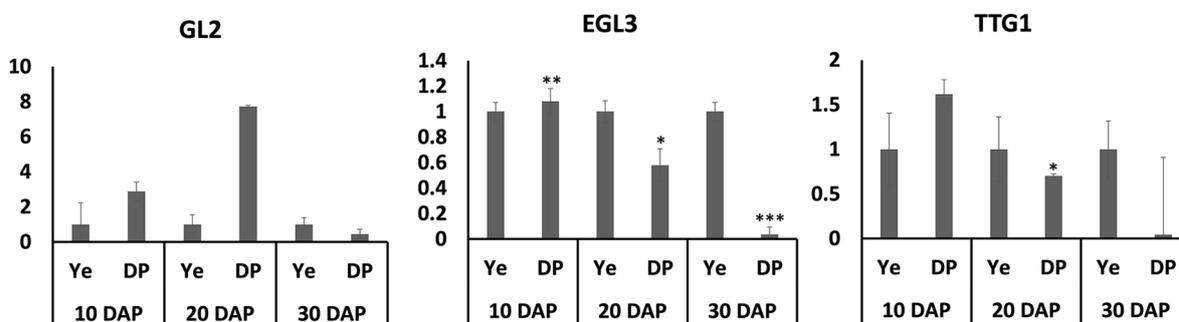
Moreover, in Figure 1b, the purple color was first deposited in the peripheral parts of the pericarp, and the remnants of chlorophyll were detected in the aleurone layer. Chlorophyll disappeared faster in DP seeds than in Ye seeds. Similar



a)



b)



**Figure 7.** Interaction network analysis and gene expression of related interacting genes. (a) Interaction network analysis of bHLH and related genes of *Arabidopsis thaliana*. Line thickness is related to the combined score ( $TaTCL2 > 0.7$ ). The homologous gene in wheat is in red. (b) qRT-PCR of genes showing direct interactions in interaction network analysis. Yellow (Ye) and deep purple (DP) samples. Days after pollination (DAP). Ye in each stage is used as an internal control. Data are means  $\pm$  SD of 3 biological replicates. Asterisks indicate significant differences when comparing Ye DAP in each developmental stage (\* $P \leq .05$ . \*\* $P \leq .01$ . \*\*\* $P \leq .001$ ).

Additionally, *F3'5H* peaked at 40 DAP in DP seeds, in contrast to previously reported results, where it peaked in mid-grain development.<sup>45</sup>

The expression levels of genes involved in pigment biosynthesis were found to vary greatly between the Ye and DP seed lines of *Triticum aestivum* L. However, all these genes were also expressed in the yellow-seeded lines. These data indicate that the absence of pigment synthesis in the yellow-seeded line of *Triticum aestivum* L. involves the downregulation, but not the complete inactivation, of several key genes.

The seed development of DP seeds was accompanied by an increase of anthocyanin content and the expression of *TaTCL2* and genes related to the flavonoid pathway.

#### Phylogenetic tree and domain

Evolutionary studies have revealed that of the genes involved in the anthocyanin pathway, structural genes evolve more slowly than regulatory genes.<sup>47</sup> The evolutionary relationship of *TaTCL2* showed closely relationship from monocot plant in *Poaceae* (Supplemental Figure S2). It has been shown that *TaMpc1* and *HvMpc5* had a high connection between gene

expression and the appearance of anthocyanin pigmentation, and the lineages of wheat and barley diverged 16 to 19 million years ago (MYA).<sup>48</sup> However, a previous study of the anthocyanin biosynthesis pathway showed that the anthocyanin genes are species specific and are not dependent on the class of the plant species.<sup>49</sup>

*TaTCL2* is a functional MYB transcription factor gene regulating anthocyanin biosynthesis. Upon examining the relationship between *TaTCL2* and MYB TFs in other plants, *TaTCL2* was revealed to be closest to *AtTCL2*, *PhODO1*, *SCAN2like*, *AtMYB*, *FaMYB*, *PhPH4*, *VvMYB5b*, and *VvMYBCS1* (Figure 6a). The dendrograms indicated that *TaTCL2* belonged to a distinct cluster of MYB proteins. We searched for domains in our sequence using Pfam database, and the results showed that it contained a SANT (Swi3, Ada2, N-Cor, and TFIIB) which has been reported to be an important domain for the anthocyanin biosynthesis.<sup>43,50</sup> The SANT domain was found in the genes that *TaTCL2* was clustered with in our phylogenetic tree (Figure 6b). Moreover, the transcript level of *TaTCL2* was substantially higher at the early stages (10 and 20 DAP) in DP seed relative to Ye seeds with low anthocyanin content.

### Interaction analysis

Our results showed that our cloned gene, *TaTCL2*, is likely involved in the anthocyanin biosynthesis pathway for wheat seed coat color deposition. R2R3 MYB TFs have been shown to interact closely with bHLH TFs<sup>51,52</sup> of the IIIf subfamily of bHLH for *Arabidopsis*, which have been demonstrated to be involved in flavonoid biosynthesis and trichome formation.<sup>29,53,54</sup> TT8 (Transparent testa), GL3 (Glabra 3), and EGL3 (Enhancer of Glabra 3) from the IIIf subfamily along with TTG1 (Transparent testa glabra 1-WD40) interact to form MBW complexes, which function in regulating the flavonoid biosynthetic genes and influencing seed coat color formation<sup>55</sup> and anthocyanin biosynthesis for *Arabidopsis*, tomato, and strawberry.<sup>56-60</sup> Our network interaction analysis of *TaTCL2* based on the orthologous genes of *Arabidopsis* showed strong interactions with GL2, EGL3, and TTG1, which could form an MBW complex related to anthocyanin biosynthesis (Figure 7a).

We performed qRT-PCR of genes showing direct interactions in interaction network analysis. It is shown that there was a higher expression in *GL2*, *TTG1*, and *EGL3* at the early stages (10 and 20DAP) compared to the late ones (30 and 40DAP), which is in accordance to the color formation in the seeds as seen in Figure 1. These allows us to conclude that there is a high possibility that an MBW complex can be formed and that *TaTCL2* is involved in seed coat color deposition in *Triticum aestivum* L.

### Conclusions

This study analyzed the networks underlying pigment formation, investigated the variation in flavonoid accumulation during seed development in Ye and DP seeds, and monitored the differential expression of the main structural genes and TFs involved in the flavonoid biosynthetic pathway using transcriptome analysis. The findings of this study elucidate the mechanisms underlying color formation in *Triticum aestivum* L. seed coats. Furthermore, by cloning the novel gene *TaTCL2*, we uncovered evidence that this gene is important in seed color formation through an MBW complex that regulates anthocyanin formation in DP seeds. The findings of this study will lay a foundation for understanding the molecular mechanism underlying DP color formation for DP seed breeding in *T. aestivum* L.

### Acknowledgements

PCF acknowledges CONACYT for the scholarship granted.

### Author Contributions

PCF and YWS conceived and designed the experiments. PCF performed experiments, analyzed data, and wrote the manuscript with support from JSY, DYK, and YWS. JSY helped with gene expression analysis. DYK performed transcriptome analysis. YWS contributed with valuable

discussions. All authors have discussed the results and approved the final manuscript.

### ORCID iD

Yong Weon Seo  <https://orcid.org/0000-0002-6052-7491>

### Supplemental material

Supplemental material for this article is available online.

### REFERENCES

1. Rebetzke GJ, Jimenez-Berni J, Fischer RA, Deery DM, Smith DJ. Review: high-throughput phenotyping to enhance the use of crop genetic resources. *Plant Sci.* 2019;282:40-48.
2. Roucou A, Violle C, Fort F, Roumet P, Ecartot M, Vile D. Shifts in plant functional strategies over the course of wheat domestication. *J Appl Ecol.* 2018;55:25-37.
3. Wood SA, Karp DS, DeClerck F, Kremen C, Naeem S, Palm CA. Functional traits in agriculture: agrobiodiversity and ecosystem services. *Trends Ecol Evol.* 2015;30:531-539.
4. Zykin PA, Andreeva EA, Lykholay AN, Tsvetkova NV, Voylovok AV. Anthocyanin composition and content in rye plants with different grain color. *Molecules.* 2018;23:948.
5. Moise JA, Han S, Gudynaite-Savitch L, Johnson DA, Miki BLA. Seed coats: structure, development, composition, and biotechnology. *In Vitro Cell Dev Biol Plant.* 2005;41:620-644.
6. Calderon Flores P, Yoon JS, Kim DY, Seo YW. Effect of chilling acclimation on germination and seedlings response to cold in different seed coat colored wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2021;21:252.
7. Himi E, Noda K. Isolation and location of three homoeologous dihydroflavonol-4-reductase (DFR) genes of wheat and their tissue-dependent expression. *J Exp Bot.* 2004;55:365-375.
8. Deng B, Yang K, Zhang Y, Li Z. The effects of temperature on the germination behavior of white, yellow, red and purple maize plant seeds. *Acta Physiol Plant?* 2015;37:174.
9. Furuta S, Takahashi M, Takahata Y, et al. Radical-scavenging activities of soybean cultivars with black seed coats. *Food Sci Technol Res.* 2003;9:73-75.
10. Takahashi R, Ohmori R, Kiyose C, Momiyama Y, Ohsuzu F, Kondo K. Antioxidant activities of black and yellow soybeans against low density lipoprotein oxidation. *J Agric Food Chem.* 2005;53:4578-4582.
11. Slavin M, Kenworthy W, Yu LL. Antioxidant properties, phytochemical composition, and antiproliferative activity of Maryland-grown soybeans with colored seed coats. *J Agric Food Chem.* 2009;57:11174-11185.
12. Washburn S, Burke GL, Morgan T, Anthony M. Effect of soy protein supplementation on serum lipoproteins, blood pressure, and menopausal symptoms in perimenopausal women. *Menopause.* 1999;6:7-13.
13. Qu C, Fu F, Lu K, et al. Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. *J Exp Bot.* 2013;64:2885-2898.
14. Rahman MM, Lee KE, Kang SG. Studies on the effects of pericarp pigmentation on grain development and yield of black rice. *Indian J Genet Plant Breed.* 2015;75:426-433.
15. Trojan V, Musilová M, Vyhnanek T, Klejduš B, Hanáčková P, Havel L. Chalcone synthase expression and pigments deposition in wheat with purple and blue colored caryopsis. *J Cereal Sci.* 2014;59:48-55.
16. Esau K. *Anatomy of Seed Plants.* John Wiley and Sons; 1977.
17. Knievel DC, Abdel-Aal ES, Rabalski I, Nakamura T, Hucl P. Grain color development and the inheritance of high anthocyanin blue aleurone and purple pericarp in spring wheat (*Triticum aestivum* L.). *J Cereal Sci.* 2009;50:113-120.
18. Liu MS, Wang F, Dong YX, Zhang XS. Expression analysis of dihydroflavonol 4-Reductase genes involved in anthocyanin biosynthesis in purple grains of wheat. *J Integr Plant Biol.* 2005;47:1107-1114.
19. Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science.* 2003;299:396-399.
20. Zhang K, Lu K, Qu C, et al. Gene silencing of BnTT10 family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PLoS One.* 2013;8:e61247.
21. Hong M, Hu K, Tian T, et al. Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. *Front Plant Sci.* 2017;8:1674.
22. Lepiniec L, Debeaujon I, Routaboul JM, et al. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol.* 2006;57:405-430.

23. Saito K, Yonekura-Sakakibara K, Nakabayashi R, et al. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol Biochem*. 2013;72:21-34.
24. Appelhagen I, Thiedig K, Nordholt N, et al. Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta*. 2014;240:955-970.
25. Xu W, Grain D, Bobet S, et al. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol*. 2014;202:132-144.
26. Shin DH, Choi MG, Kang CS, Park CS, Choi SB, Park YI. A wheat R2R3-MYB protein PURPLE PLANT1 (TaPL1) functions as a positive regulator of anthocyanin biosynthesis. *Biochem Biophys Res Commun*. 2016;469:686-691.
27. Ramsay NA, Glover BJ. MYB-bHLH-WDR40 protein complex and the evolution of cellular diversity. *Trends Plant Sci*. 2005;10:63-70.
28. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J*. 1987;6:3553-3558.
29. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J*. 2008;53:814-827.
30. Solano R, Nieto C, Avila J, Cañas L, Diaz I, Paz-Ares J. Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from *Petunia hybrida*. *EMBO J*. 1995;14:1773-1784.
31. Appelhagen I, Nordholt N, Seidel T, et al. TRANSPARENT TESTA 13 is a tonoplast P3A-ATPase required for vacuolar deposition of proanthocyanidins in *Arabidopsis thaliana* seeds. *Plant J*. 2015;82:840-849.
32. Shin OH, Kim DY, Seo YW. Effects of different depth of grain colour on antioxidant capacity during water imbibition in wheat (*Triticum aestivum* L.). *J Sci Food Agric*. 2017;97:2750-2758.
33. Galicia L, Miranda A, Gutiérrez MG, et al. *Laboratorio de calidad nutricional de maíz y análisis de tejido vegetal: protocolos de laboratorio 2012*. CIMMYT; 2012.
34. Hong MJ, Kim DY, Ahn JW, Kang SY, Seo YW, Kim JB. Comparison of radio-sensitivity response to acute and chronic gamma irradiation in colored wheat. *Genet Mol Biol*. 2018;41:611-623.
35. Warren CR. Rapid measurement of chlorophylls with a microplate reader. *J Plant Nutr*. 2008;31:1321-1332.
36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods*. 2001;25:402-408.
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
38. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357-360.
39. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166-169.
40. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with *deseq2*. *Genome Biol*. 2014;15:550.
41. Desper R, Gascuel O. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol Biol Evol*. 2004;21:587-598.
42. Zhao F, Li G, Hu P, et al. Identification of basic/helix-loop-helix transcription factors reveals candidate genes involved in anthocyanin biosynthesis from the strawberry white-flesh mutant. *Sci Rep*. 2018;8:2721.
43. Zong Y, Zhu X, Liu Z, et al. Functional MYB transcription factor encoding gene *AN2* is associated with anthocyanin biosynthesis in *Lycium ruthenicum* Murray. *BMC Plant Biol*. 2019;19:169.
44. Aasland R, Stewart AF, Gibson T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-cor and TFIIIB. *Trends Biochem Sci*. 1996;21:87-88.
45. Yang GH, Li B, Gao JW, et al. Cloning and expression of two chalcone synthase and a flavonoid 3'5'-hydroxylase 3'-end cDNAs from developing seeds of blue-grained wheat involved in anthocyanin biosynthetic pathway. *Acta Bot Sin*. 2004;46:588-594.
46. Yang GH, Zhao XQ, Li B, et al. Molecular cloning and characterization of a DFR from developing seeds of blue-grained wheat in anthocyanin biosynthetic pathway. *Acta Bot Sin*. 2003;45:1329-1338.
47. Rausher MD, Miller RE, Tiffin P. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol Biol Evol*. 1999;16:266-274.
48. Strygina KV, Khlestkina EK. Structural and functional divergence of the Mpc1 genes in wheat and barley. *BMC Evol Biol*. 2019;19:45.
49. Shoeva OY, Glagoleva AY, Khlestkina EK. The factors affecting the evolution of the anthocyanin biosynthesis pathway genes in monocot and dicot plant species. *BMC Plant Biol*. 2017;17:256.
50. Oglesby L, Ananga A, Obuya J, Ochieng J, Cebert E, Tsolova V. Anthocyanin accumulation in muscadine berry skins is influenced by the expression of the MYB transcription factors, MybA1, and MYBCS1. *Antioxidants*. 2016;5:35.
51. Mol J, Jenkins G, Schäfer E, Weiss D, Walbot V. Signal perception, transduction, and gene expression involved in anthocyanin biosynthesis. *CRC Crit Rev Plant Sci*. 1996;15:525-557.
52. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol*. 2001;126:485-493.
53. Toledo-Ortiz G, Huq E, Quail PH. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *Plant Cell*. 2003;15:1749-1770.
54. Pires N, Dolan L. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol Biol Evol*. 2010;27:862-874.
55. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci*. 2015;20:176-185.
56. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J*. 2011;66:94-116.
57. Zhou LL, Shi MZ, Xie DY. Regulation of anthocyanin biosynthesis by nitrogen in TTG1-GL3/TT8-PAP1-programmed red cells of *Arabidopsis thaliana*. *Planta*. 2012;236:825-837.
58. Salvatierra A, Pimentel P, Moya-León MA, Herrera R. Increased accumulation of anthocyanins in *Fragaria chiloensis* fruits by transient suppression of FcMYB1 gene. *Phytochemistry*. 2013;90:25-36.
59. Nemesio-Gorrioz M, Blair PB, Dalman K, et al. Identification of Norway spruce MYB-bHLH-WDR transcription factor complex members linked to regulation of the flavonoid pathway. *Front Plant Sci*. 2017;8:305.
60. Tominaga-Wada R, Ota K, Hayashi N, Yamada K, Sano R, Wada T. Expression and protein localization analyses of *Arabidopsis GLABRA3 (GL3)* in tomato (*Solanum lycopersicum*) root epidermis. *Plant Biotechnol*. 2017;34:115-117.