Genome-wide Identification and Analysis of *CCT* Genes in Wheat (Triticum Aestivum L.)

Hongwei Zhang

College of Agriculture, Hebei Agricultural University; Plant Genetic Engineering Center of Hebei Province, Institute of Genetics and Physiology, Hebei Academy of Agriculture and Forestry Sciences

https://orcid.org/0000-0003-0460-6125

Xinxia Liang

Plant Genetic Engineering Center of Hebei province, Institute of Genetics and Physiology, Hebei Academy of Agriculture and Forestry Sciences

Shuo Zhou (zhoushuobio@163.com)

Plant Genetic Engineering Center of Hebei province, Institute of Genetics and Physiology, Hebei Academy of Agriculture and Forestry Sciences

Haibo Wang

Plant Genetic Engineering Center of Hebei Province, Institute of Genetics and Physiology, Hebei Academy of Agriculture and Forestry Science

---

Research article

Keywords: CCT, Photoperiod Flowering, Vernalization, Triticum aestivum

Posted Date: November 13th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-104841/v1

Abstract

Background: The vernalization, in which the plants must undergo a prolonged winter cold exposure to flower, is mainly controlled by a suppressive MADS-box gene *FLC* in *Arabidopsis.* However, different from *Arabidopsis*, the CCT-domain containing gene *VRN2* is the critical vernalization-related suppressor gene in cereals. Based on this apparent diversity of vernalization in different plants, and involvement of *VRN2* with vernalization in cereals, we conducted a genome-wide analysis of *CCT* genes in wheat, and the relationship between vernalization and these genes were also revealed.

Results: A genome-wide analysis of the *CCT* genes in common wheat was performed by employing a hidden Markov model-based method, and 127 sequences, which assigned to 40 clusters, were obtained in three subgenomes. Specially, two of the gene clusters are duplicated, and distinguishingly located near telomere. Furthermore, these sequences were classified into eight groups by a phylogenetic analysis procedure using the UPGMA method, and this taxonomy is concordant to the classification based on CCT interruptions and domain organization which roughly divided the proteins into four divergently related subfamilies. Moreover, the expression of several *CCT* genes is continually downregulated during and after vernalization, but no continually upregulated *CCT* genes were revealed, as indicated by transcriptome sequencing and real-time quantitative PCR analysis.

Conclusion: This study improves our understanding of the structure and function of *CCT* genes, suggests many vernalization-related *CCT* genes, and may guide future investigations on *CCT* genes and vernalization in wheat.

Background

Flowering time in wheat (*Triticum aestivum*) is mainly controlled by seasonal temperature and photoperiod variations [1]. Winter wheat only flowers after prolonged winter cold exposure and is thus prevented from premature flowering during the former warm autumn. This process, referred to as vernalization, is mainly controlled by *VERNALIZATION 1* (*VRN1*), *VRN2*, and *VRN3* [2–4].

The *VRN1* gene is the major gene responsible for wheat flowering. It is the ortholog of *Arabidopsis thaliana APETALA 1* (*AP1*), and the simultaneous loss-of-function of *VRN1*, *FRUITFULL 2* (*FUL2*) and *FUL3*, leads to incompetent flowering in wheat [2, 5, 6]. This accelerated flowering gene is mainly expressed in the leaves and shoot apical meristem (SAM) and is upregulated by vernalization [2]. VRN1 proteins directly bind to the promoter of another gene *VRN3*, which is the ortholog of *Arabidopsis Flowering Locus T* (*FT*), and thus promote *VRN3* expression in leaf vascular cells under long-day (LD) conditions [4, 7, 8]. The encoded VRN3 proteins are then transported to the SAM through the phloem, which then further accelerate *VRN1* expression along with Flowering Locus D-like 2 (FDL2) and 14-3-3 proteins both in the leaves and SAM [9, 10]. The overexpression of the *VRN3* gene leads to the upregulation of *VRN1* and spring growth habit in winter wheat [9]. Normally, the *VRN3* gene is upregulated by the photoperiod pathway gene *CONSTANS1* (*CO1*)/*CO2* under long-day conditions after vernalization

[11]. Loss-of-function or gain-of-function of the *Photoperiod 1* (*PPD1*) gene in wheat also alters flowering time by affecting *VRN3* and *CO1* expression [12–14]. However, a flowering repressor VRN2 could compete with CO2, forming the VRN2/Heme Activator Protein 3 (HAP3)/HAP5 complex. This triple complex directly binds to the CCAAT motif of the *VRN3* promoter and represses its expression [15]. The *VRN2* gene is the most closely related ortholog of rice gene *Grain number, plant Height and heading Date 7* (*Ghd7*), which is also upregulated by LD conditions and composed of two tandemly arranged *ZCCT* copies [3, 16, 17]. Besides, the *VRN2* gene is also directly suppressed by the binding of the VRN1 protein to its unique CArG motif, and by vernalization independent of *VRN1* [18, 19].

Among this extremely complex process of inducing flowering [20], many genes harbor unique CCT domains, such as *VRN2*, *CO1/2*, and *PPD1*. The CCT domain, which was originally described in the *Arabidopsis* protein CO, CONSTANS-like (COL), and Timing Of CAB expression 1 (TOC1), is about 43 amino acids in length and contains a putative nuclear localization signal within the N-terminal half of the domain [21]. The genes harbor this domain have been assigned as *CCTs* and divided into several subfamilies according to their additional domains. For example, B-box-containing *CCT* genes are also known as *CO/COL* genes, and *CCT Motif Family* (*CMF*) genes are *CCT* genes without any additional domains. Furthermore, an extra Response_reg domain occurs in *Pseudo-Response Regulator* (*PRR*) genes, which occupy another *CCT* gene subgroup [22].

A previous study has suggested that *CCTs* are associated with different biological processes in *Arabidopsis* such as photoperiodic flowering, circadian rhythm regulation, light signaling, and sugar response gene expression [23]. TOC1 is believed to be a component of the central oscillator that interacts with *Late Elongated Hypocotyl* (*LHY*) and *Circadian Clock-Associated 1* (*CCA1*), forming a loop that plays a critical role in clock function in *Arabidopsis* [24, 25]. The mutation of this gene (*toc1-1*) results in shorter circadian rhythm as well as alters photoperiodic flowering responses relative to the wild-type plants [21]. Another CCT protein CO is stabilized by light and degraded in the morning or in darkness, thereby causing a precise daily rhythm in its abundance [26]. This protein activates the MADS-box gene *Suppressor of Over-expression of CONSTANS 1* (*SOC1*) through FT, promoting flowering in *Arabidopsis* under LD conditions, in which the CO protein directly binds to the promoter of *FT* [27, 28]. However, the highest exprssion level of *Heading Date 1* (*Hd1*, the ortholog of *CO*) in rice, is strictly restricted in darkness under its inductive short-day (SD) conditions, indicates a distinct regulatory mechanism [29, 30]. The marked induction of *Arabidopsis Pseudo-Response Regulators 9* (*APRR9*) by red light is severely impaired in the *phyAphyB* mutant [31], and its overexpression leads to shortened hypocotyls under red light [32]. Moreover, *AtCOL7* promotes shoot branching in a high red/far-red light ratio condition, but its overexpression lines develop longer hypocotyls in shade [33]. And *AtCOL4* is involved in ABA and salt stress responses through the ABA-dependent signaling pathway [34]. In addition, *AtCMF8* can induce several endogenous sugar-inducible genes and is also regulated by sugars [35].

The CCT proteins play a critical role in plant flowering. Previous studies have identified CCT proteins in *Arabidopsis*, rice, sorghum, foxtail millet, barley, and *Brachypodium* using whole-genome BLAST analysis [22, 36], but not in wheat. In this study, we performed whole-genome analysis using HMMER3.0 and

hidden Markov model (HMM) profiling of CCT domains, and sequence and structural information, phylogenetic reconstruction, chromatin distribution, and expression patterns under vernalization, were also assessed. Using available data on genome-wide assembly and annotation, this study conducted a comprehensive characterization of wheat *CCT* genes.

Results

Genome-wide scanning of *CCT* genes

Based on the available genome-wide assembly and annotation in wheat Cultivar CS [37], the wheat *CCT* genes were searched by employing the HMM profile using the HMMER3.0 package. After the redundant sequence filtering, Batch CD-Search/FGENESH + verification, and gene complementation by BLAST, a total of 127 candidate sequences were obtained. These genes were subsequently designated based on the available sequences of *Brachypodium*, rice, and wheat [22, 38], and 54 *CO/COLs*, 46 *CMFs*, 15 *PRRs*, and 12 *Zinc-finger protein expressed in Inflorescence Meristem* genes *(ZIMs)*, which belong to 18 *CO/COL* clusters, 13 *CMF* clusters, 5 *PRR* clusters, and 4 *ZIM* clusters, were assigned. Among them, *TaCMF6* and *TaCMF8* consisted of 7 and 6 members, respectively. For the assignment, *TaCMF8-A2* indicates the 2nd copy of the 8th *CMF* gene in the A-genome of *T. aestivum*. The protein product of these genes consisted of 170 (TaCMF14-D) to 763 (TaPRR73-D) amino acids, and lead to a molecular weight ranging from 19.48 kD (TaCMF14-D) to 83.21 kD (TaPRR73-D) and theoretical isoelectric point ranging from 4.32 (TaCO13-B) to 10.22 (TaCMF14-B). Most of the 127 proteins are located in nucleus (111), and only 11 occurred in the chloroplast, 3 were cytoplasmic, and 2 were extracellular, indicating that CCT proteins were predominantly nuclear proteins that had flowering regulatory functions by acting as transcription regulators (Table S1) [15].

The phylogenetic reconstruction of CCTs

Phylogenetic reconstruction using the UPGMA method with 1,000 bootstrap replications and the Jones-Talor-Thornton (JTT) model in wheat, *Arabidopsis*, *Brachypodium*, and rice revealed that the CCT proteins could be classified into 8 groups. Consistent with the findings of previous studies, Groups I to III contain all of the CO/COLs, and the same classification groups were adopted [36]. Among these, 24 wheat, 7 rice, 7 *Brachypodium*, and 6 *Arabidopsis* members were assigned to Group I; 9 wheat, 3 rice, 3 *Brachypodium*, and 4 *Arabidopsis* members to Group II; and 21 wheat, 7 rice, 6 *Brachypodium*, and 7 *Arabidopsis* members to Group III. Although failled to cluster with Group I, Group IV was still adopted the same classification group as previously, even if their members were considered to be CMFs due to the lack of B-box domain [22, 36]. In this group, 9 wheat, 2 rice, 1 *Brachypodium*, and no *Arabidopsis* members were assigned. The rest of the CMFs were assigned to Groups V and VI, except for AtCMF5 and AtCMF7, which was placed in Group VII, together with the PRRs. Group V consisted of 28 wheat, 9 rice, 7 *Brachypodium*, and 9 *Arabidopsis* members; Group VI comprised 9 wheat, 2 rice, 3 *Brachypodium*, and 4 *Arabidopsis* members; and Group VII included 15 wheat, 5 rice, 5 *Brachypodium*, and 7 *Arabidopsis* members. Interestingly, the PRRs were further divided into 3 clades, namely TOC1, PRR3/7, and PRR5/9. Group VIII

contained all of the ZIMs that consists of 12 wheat, 4 rice, 6 *Brachypodium*, and 3 *Arabidopsis* members and is further roughly divided into the ZIM1/ZIM3 and ZIM2/ZIM4 clades (Fig. 1).

The conserved domain and gene structure of *CCTs* in wheat

According to the conserved domains predicted by the Batch CD-Search tool, motifs analyzed by MEME suite, and phylogenetic relationship of the CCTs in wheat, their gene structure coincides with their phylogenetic taxonomic status (Fig. 2). Although most of the CCT domains (corresponding to motif 1 that is predicted by the MEME suite, and motifs 2 to 8 originated from the same prediction) are located in the C-terminal of the sequences except for its loss in TaCO13-B, Group VIII proteins (ZIMs) have their CCT domains in the middle, with the extra tify domain (corresponding to motif 5) at the N-terminal, and the Znf_GATA domain (corresponding to motif 4) at the C-terminal. Group VII proteins (PRRs) harbor their Response_reg domain (corresponding to motif 6 plus motif 3) at the N-terminal, and Groups I to III proteins (CO/COLs) harbor one or two B-box zinc finger domains (corresponding to motif 2 or 7), with all proteins harboring motif 2. Specifically, TaCO13 and TaCO14 harbor two B-boxes, while the rest of the proteins in Group III and all proteins in Group II harbor only one. Both proteins harboring one or two B-boxes are appeared in Group I, and even some obvious differences for TaCO1 and TaCO7 were predicted using B-box prediction in the MEME suite and Batch CD-Search tool (Fig. S1). Furthermore, Group IV to VI proteins (CMF) do not contain any domains other than CCT, except for an unknown motif (motif 8) located at the N-terminal of TaCMF6 in Group V.

The conserved domains of these genes are usually interrupted by introns. For example, the tify domains are usually located in exons 1 and 2, Zn_GATA domains in exons 5 and 6, and Response_reg domains in the first three coding exons, whereas the B-box domains are always located in the first coding exon without interruption. In addition, the CCT domains displayed a distinct interruption mode that coincides with their phylogenetic classification: the *ZIMs* (Group VIII) harbor the CCT domain in the 3rd and 4th coding exons, and its products are interrupted at the 32nd residue of the unique 43-amino acid protein fragment. The CCT domains of Group III gene products are interrupted at the 16th residue and the Group V–VII members just after the 22nd, 37th, and 20th residue, respectively. These genes harbor CCT domains in the last two exons, except for *TaCMF4* and *TaCMF10* in Group V, where these are located at antepenultimate and penultimate exons, and *TaTOC1* in Group VII at the last exon. The rest of the *CCT* genes, i.e., the members of Groups I, II, and IV, harbor the CCT domains in the last exon, without any interruption (Fig. 2 and Fig. 3). Furthermore, the alignment of the CCT domains revealed that these 43-amino acid regions are highly conserved in wheat. Among these, 8 residues (R1, K11, Y23, R26, A30, R35, G38, and F40) are completely identical, whereas R15 and K27 are identical except for TaCMF7 and TaCMF9, respectively. In addition, a PL insertion occurs upstream L17 in TaCMF7 (Fig. 3 and Fig. S2).

Exon distribution in these genes is also divergent, which varies from 1 to 9. Although most of Group I, II, and IV genes embrace 2 exons, *TaCO8* in Group I and *TaCO12* in Group II embrace 2. Group VI genes, together with another Group I gene *TaCO7*, embrace 3 exons in their sequences. Group III genes embrace

4 exons, except for *TaCO13-B*. Groups V, VII, and VIII contain genes with the most divergent exon distribution, which varies from 2 to 5, 6 to 9, and 7 to 8, respectively (Fig. 2, Table S1).

Genome distribution analysis

The 127 *CCT* genes were further analyzed in terms of distribution across 21 chromosomes, whose lengths vary from 474 Mb to 831 Mb. The result indicated that wheat chromosome 3 contain the fewest *CCT* genes, with only one in each subgenome, followed by chromosome 2 with 3; whereas chromosomes 4B, 4D, and 7D contain the highest number of *CCT* genes with 10, and chromosomes 7A and 7B with 9. Interestingly, more than one copy of *TaCMF6* and *TaCMF8* occurs in each sub-genome and is tandemly arranged near telomeres. However, *TaCMF13* did not belong to the same cluster as *TaCMF14*, although these were tandemly arranged in chromosome 5, as it is more closely related to *TaCMF15* in chromosome 4 (Fig. 1, 63.58% identity by pairwise alignment between *TaCMF13* and *TaCMF15*, and 43.37% between *TaCMF13* and *TaCMF14*). Furthermore, the gene rearrangement of chromosome 4A is somewhat different from that of chromosomes 4B and 4D, and a recent recombination between chromosomes 4A and 5A near telomeres is also revealed by the transposition of *TaCMF4* and *TaCMF8* from chromosome 4A to chromosome 5A. Finally, the existence of three *TaPPD1* genes and designation of *TaPPD1-A* and *-D* imply that *TaPPD1-U* is the B-genome gene of *TaPPD1*, and the confusing *TaCO20* arrangement in the genome indicates the possibility of B-genome location of *TaCO20-U* and occurrence of another recombination event (Fig. 4).

The expression pattern of *CCT* genes under vernalization

To clarify the potential functions of *CCT* genes under vernalization in wheat, a transcriptome analysis using leaf tissues of a winter wheat cultivar Shiluan 02 − 1 collected before, during, and after vernalization, was performed. Group II and VI genes exhibited insignificant changes in expression with vernalization, with a continuous low expression in Group VI and high expression of Group II genes, particularly *TaCO10*. *TaCO13*, *TaCO16*, and *TaCO15* showed the highest expression in Group III, wherein *TaCO16* is upregulated and *TaCO15* downregulated under vernalization. Another Group III gene, *TaCO18*, which is barely expressed under normal conditions, was also simultaneously upregulated with vernalization. Group I genes were the most differentially expressed *CCT* genes, in which *TaCO3*, *TaCO4*, and *TaCO6* showed the highest expression levels, with *TaCO6* being upregulated, while *TaCO2* and *TaCO8* had almost no expression, and *TaCO1* was downregulated. Although six copies of *TaCMF8* were observed in Group IV, only *TaCMF8-B* showed slight expression, whereas no expression data was collected for *TaCMF8-B2*, while its paralog gene *TaCMF11* was continuously upregulated under vernalization. Most of the Group V genes exhibited low expression levels except for *TaCMF6*, which was significantly upregulated under low-temperature conditions. Furthermore, the expression of *PRR* genes (Group VII), in which *TaPRR95* and *TaPRR73* upregulated, was sustained at a relatively high level except for *TaPPD1*, and *TaZIM4-A* was the only highly expressed *ZIM* gene (Group VIII), as the expression data of *TaZIM4-B/D* was also not collected (Fig. 5).

To further assess significant changes in the expression of *CCTs* under vernalization, we compared the expression levels of genes during/after to those before vernalization. Among these, 49 genes were upregulated and 31 downregulated at least at one time point. Of these genes, 8 were continuously upregulated, 11 for most of the time, and 10 only for a short time. Among these, *TaCMF6*, *TaCMF11*, *TaCO18*, *TaPRR95*, and *TaCO16* were the most significantly upregulated genes, and remained upregulating even after vernalization. However, only 2 genes were continuously downregulated, while 7 most of time, and 11 only for a while, with *TaCO1* and *TaCO15* showing the most significant downregulation that persisted even after vernalization (Fig. 6, Table S3).

Expression analysis using real-time PCR

To further validate the expression profile of vernalization-related *CCT* genes, real-time PCR analysis of the mostly differentially expressed genes, and the most popularly studied *CCTs* were conducted. Of these genes, *TaCMF6*, *TaCMF11*, *TaCO18*, and *TaPRR95* were significantly upregulated, and *TaCO16* showed slight upregulation under vernalization, which coincided with the results of RNA-sequencing. Surprisingly, their expression levels rapidly decreased to pre-vernalization levels immediately after the temperature increased. However, the remaining genes were downregulated with continuous exposure to low temperature and maintained this low expression even after vernalization. Interestingly, unlike other genes with expression levels that gradually decreased with vernalization, those of *TaCMF8* and *TaCO1* were rapidly reduced to remarkably low levels (Fig. 7).

Discussion

Genome-wide scanning of *CCTs* in common wheat

Plants at different stages during their life cycle are exposed various external stimuli, which in turn influence various processes, such as transition from vegetative growth to reproductive growth, and flowering regulation. This study investigated a group of CCT domain-containing transcription factors in common wheat, obtained 127 distinct sequences by hmmsearch against a unique HMM profile of the CCT domain [39, 40], and designated these based on their orthologs. Although a rice *CMF2* ortholog was identified, none were found in *Brachypodium*, *Setaria italica*, or *Hordeum vulgare* [22], as well in wheat in this study. The existence of almost identical sequences *OsCMF12* and *OsCMF13* (99.86% for DNA identity, and 100% cDNA identity) and their alternative existence in other plants confirm the applicability of designation of *TaCMF13* in wheat. *TaCO9* is thought to be the closest paralog of *TaCMF8*/ *TaVRN2* [41], and no B-box zinc finger motif exists. Thus, it should be re-designated as *TaCMF11*. Furthermore, we did not find any orthologs of the wheat *CCT* gene *TaCMF15* in rice or *Brachypodium* [22], because the previously designated *BdCMF15* is highly homologous to *OsCMF8*/ *Osl*/ *Ghd7* and *TaCMF8*/ *TaVRN2*, which should be re-designated as *BdCMF8*. In addition, we also did not find rice *OsQ* homologs in wheat.

According to the phylogenetic reconstruction and gene structure analysis of CCTs in wheat, 8 well divided groups coincide with grouping based on different domains and specific CCT domain interruptions (Fig. 2

and Fig. 3). We designated these as *CO/COLs* for Groups I to III as previous studies, *CMFs* for Groups IV to VI, *PRRs* for Group VII, and *ZIMs* for Group VIII, and further focused on discussing their distinct structures and possible functions because most of the genes found in plants play distinct functions, even for *ZIMs* that have never been classified as *CCT* genes earlier.

CO/COLs and *CMFs*

There are many extra domains located in the *CCTs*, in which most of these genes (100 out of 127) harboring one or two B-boxes (*CO/COLs*) or without any domains other than CCT (*CMFs*). According to their highly conserved sequences and zinc-binding residue spacing, the B-box domains are divided into two groups. Interestingly, B-box1 occurs in all of the *Arabidopsis* and rice CO/COLs, whereas B-box2 is only present in CO/COL proteins harboring two B-box domains [22, 42, 43]. In wheat, the B-box1 domain was not predicted in TaCO7 by the Batch CD-Search tool, but a multiple alignment was subsequently performed, and the corresponding sequence (motif 2 predicted by MEME suite) coincided with the conserved sequence of the B-box domain (Fig. S1). Furthermore, the B-box2 of TaCO1 was also not predicted by the Batch CD-Search tool, but the alignment revealed that the corresponding sequence (motif 7) lacked the consensus sequence (Fig. S1). These specific B-box domains are generally thought to interact with the coiled coil domain of other proteins to form a functional equivalent complex of RING, B-box, coiled-coil/Tripartite motif (RBCC/TRIM), and thus CO/COLs usually function in processes such as photoperiod flowering, circadian rhythms, seedling photomorphogenesis, light signaling, cold/drought responsible, and hormonal signaling [42].

Phylogenetic reconstruction of CO/COLs and CMFs in wheat roughly coincided with those of *Arabidopsis*, rice, and *Brachypodium* (Fig. 1) [22, 36], illustrating its conservation in plants. This classification scheme is obviously consistent not only with the amount of B-box domain but also with CCT interruptions. The CCT domain in Group I, II, and IV proteins was found to be encoded by a single exon, but the CCT domain of Group III proteins is interrupted at the 16th residue, whereas that of Group V and VI proteins is interrupted after 22th and 37th residue, respectively. A previous study has suggested that CO/COLs have undergone a gradual loss of the B-box domain, from 2 to 1 and then to none [22]. This is somewhat implied by our phylogenetic reconstruction of CCTs, wherein both single and double B-box-containing proteins are found in Groups I and III, and B-box-free Group IV proteins are closely related to these two groups (Fig. 1) [22, 36].

Gene duplication is another critical event that occurred during gene evolution that led to the divergence of gene families. We detected some recent duplications involving *CMF* genes in this study, in which *CMF6* and *CMF8* both have two or three gene copies tandemly arranged in each subgenome in wheat and barley, whereas only one copy occurs in *Brachypodium* and rice [22]. These tandemly arranged genes are both located near telomeres, where recombination rates are relatively high (Fig. 4).

PRRs

We identified five different *PRRs* in a subgenome of wheat, which is concordant with the findings of previous studies involving other plant species such as *Arabidopsis*, rice, and *Brachypodium* [22, 44]. The

subsequent phylogenetic analysis divided them into 3 clades, each representing TOC1, PRR3/7, and PRR5/9 consistent with the findings of a previous study (Fig. 1) [45]. The five *PRR* genes in *Arabidopsis* have been characterized underwent a precise circadian rhythm, with their mRNAs accumulating sequentially in the order of *APRR9*, *APRR7*, *APRR5*, *APRR3*, and *TOC1* after dawn at 2- to 3-h intervals [44], and forming extremely complex regulatory loops with CCA1/LHY and other factors [45]. Indeed, the central oscillator TOC1 represses the expression of *CCA1* and *LHY* by directly binding to their promoters [25], and indirectly promoting *CCA1* expression by interacting with CCA1 Hiking Expedition (CHE), which is the critical repressor of *CCA1* [46]. However, CCA1 and LHY also reduce *TOC1* expression by binding to its promoter, forming the transcriptional feedback loop involving *TOC1*, *CCA1*, and *LHY* [24]. Furthermore, the robust oscillatory expression of *TOC1* is also controlled by its post-translational regulation via E3 ubiquitin ligase SCF^{ZTL}, and the phosphorylation of TOC1 and PRR3 strengthens their interaction at the N-terminal, preventing TOC1 from interacting with SCF^{ZTL} and subsequently proteolysis of TOC1 [47].

The approximate 24-h time oscillation involving the circadian clock allows higher plants to phase its biological activities in response to multiple external cues of the day, thus affecting various processes. The overexpression of *APRR3* in *Arabidopsis* leads to a late flowering phenotype under LD conditions [48], and the mutation of *Ppd-H1* and *Ppd-D1*, the orthologs of *PRR37* in barley and common wheat, also lead to a reduced photoperiod responsible, and late flowering [49, 50]. Shortened petioles and lengthened hypocotyls are also developed in the *prr9prr5prr3* mutant, and the response of *PRRs* to low temperature, again indicates the multiple functions of *PRRs* involved in plant development [51, 52].

ZIMs

The ZIMs are the only proteins predicted to harbor a CCT domain in the middle of the sequences, together with an extra tify domain at the N-terminal and a ZnF_GATA domain at the C-terminal. The ZnF_GATA domain, which comprises a conserved CX_2-$CX_{17\text{-}20}$-CX_2C motif following a highly basic region, is widely distributed in fungal, animals, and plants and plays critical roles in development, differentiation, and control of cell proliferation [53, 54]. The tify domain, which was named after its most conserved TIF[F/Y]XG motif, was first discovered in the *Arabidopsis ZIM* gene and is a major determinant of the plant-specific manner of this gene family [55].

Unlike *PRRs*, the gene copy of *ZIMs* in plants is variable, with four copies in a given wheat subgenome, three in *Arabidopsis*, four in rice, and six in *Brachypodium* (Fig. 1, Table S1, Table S2) [53]. Phylogenetic analysis of these proteins indicates the divergence of ZIM1/ZIM3 and ZIM2/ZIM4, in which the *Arabidopsis* ZIMs are more closely related to the ZIM1/ZIM3 clade (Fig. 1). Furthermore, the *ZIMs* also have distinct functions. For example, *AtZIM* is specifically expressed in flowers and flower buds, and its overexpression leads to a cell elongation in hypocotyls and petioles and leaf upward positioning [56]. Moreover, the overexpression of *TaZIM-A1* also delays flowering by directly binding to the promoters and subsequent suppression of *TaCO-1* and *TaFT-1* [38].

Expression profile of *CCTs* during vernalization

Plant flowering is an extremely complex process in response to various internal and external stimuli. To flower at the most appropriate time, plants need to opportunely receive these signals by employing many extremely complex pathways such as photoperiod and vernalization pathways. For the photoperiod pathway, light signals are integrated with the precise circadian clock system by CO, which is stabilized by light and degraded in the dark [26, 57]. The *PRRs*, especially *TOC1*, are believed to be part of the core circadian oscillator that interacts with *LHY* and *CCA1*, forming the precise feedback loop and regulating clock output processes [24, 25, 45]. However, the vernalization pathway is somewhat different between *Arabidopsis* and cereals; they choose different genes as the core flowering suppressor, i.e., *Flowering Locus C* (*FLC*) in *Arabidopsis* [58], and *VRN2* (*CMF8*) in *wheat* [3].

Our expression analysis performed by RNA-sequencing and real-time PCR revealed many low temperature-responsible *CCTs*, including both upregulated and downregulated genes. Among these, *TaCMF6*, *TaCMF11*, *TaCO18*, *TaPRR95*, and *TaCO16* are the most upregulated genes, but an immediate decrease of their expression after vernalization was also revealed by real-time PCR. *TaCO1* and *TaCO15* are the most continuously downregulated genes even after vernalization. Furthermore, the downregulation of *TaCO2*, *TaCMF8* (*TaVRN2*), and *TaPPD1* was also revealed by real-time PCR and RNA-sequencing even when these are not highly significant. Real-time PCR analysis showed that *TaCMF8* (*TaVRN2*) is the only rapidly downregulated gene except for *TaCO1*, but this was not revealed by our RNA-sequencing analysis. This is probably due to the lack of *TaCMF8-B2* expression data and extremely low expression of the other genes of this cluster.

Conclusions

CCTs are a large group of genes that are closely related to circadian rhythm regulation and photoperiod flowering. In this study, whole-genome scanning of wheat *CCTs* was performed using an HMM-based method, and a total of 127 genes, which were assigned to 40 clusters, were obtained in the three subgenomes. Among these, more than three gene copies were detected in *TaCMF6* and *TaCMF8*, and they distinguishingly located near telomere. Subsequent phylogenetic analysis divided the CCTs into eight groups, which represented four diversely related subfamilies successively corresponding to CO/COLs (Groups I to III), CMFs (Groups IV to VI), PRRs (Group VII), and ZIMs (Group VIII). The conserved domain organization and gene structure of these genes coincide with their phylogenetic relationship. Moreover, transcriptome and real-time PCR analysis did not reveal any remarkably upregulated CCTs, which remained highly expression until the temperature had risen again, but several downregulated genes such as *TaCO1* and *TaCO15* were identified. This finding implied the existence of some vernalization-related *CCT* genes, and provided new insights into wheat vernalization.

Methods

Identification of wheat *CCT* genes

Potential *CCT* genes in wheat were identified following the method of Zhan et al., with some modifications [39]. The HMM profile for the CCT domain (PF06203) was obtained from Pfam (http://pfam.xfam.org/browse), and the available protein sequences and feature information from wheat cultivar Chinese Spring (CS), *Brachypodium*, and *Arabidopsis* (Release 47) were downloaded from Ensembl Plants (http://plants.ensembl.org/info/data/ftp/index.html). The available rice proteins (Release 7) were downloaded from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/). The CCT proteins were obtained using the hmmsearch tool embedded in HMMER3.0 according to the given HMM profile and available sequences. After filtering out redundant sequences, the CCT proteins were confirmed using the Batch CD-Search tool in National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/cdd) and assigned by comparing these to the available sequences from *Brachypodium* and rice. The controversial sequences were reassessed using FGENESH+ (http://linux1.softberry.com/), and genes from different subgenomes were complemented by BLAST.

Phylogenetic analysis

The identified CCT protein sequences were aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), and phylogenetic analysis was performed using Mega6.0 with the UPGMA Method, 1,000 bootstrap replications, and the JTT model [59]. Then, the constructed tree was reconfigured by iTOL v4 (https://itol.embl.de/) [60], or used in gene structure and heat map construction.

Protein properties and sequence analysis

The corresponding theoretical isoelectric point and molecular weight were calculated using online Compute pI/Mw tool in ExPASy (https://web.expasy.org/compute_pi/). Subcellular localization of these proteins was performed using CELLO v.2.5 (http://cello.life.nctu.edu.tw/) [61]. The conserved motifs were predicted by MEME suite (http://meme-suite.org/tools/meme) [62], and conserved domains by Batch CD-Search tool. Multiple alignment of domains was performed using Clustal Omega and shaded by BoxShade (https://embnet.vital-it.ch/software/BOX_form.html). Amino acid representation of the domains was conducted using WebLogo (http://weblogo.berkeley.edu/logo.cgi). Visualization of gene structure and motifs, genome distribution analysis, expression profiling during vernalization, and UpSet plot diagram representation of differentially expressed genes were performed by TBtools [63].

Wheat leaf tissue prepare and RNA extraction

The seeds of wheat cultivar Shiluan02-1 were planted in plots with nutrient soil and vermiculite mixed at a volume ratio of 1:1 at 22 °C under LD conditions (16 h/8 h). After two weeks, leaf tissues were collected for three replicates, with every replicate from 20 individual plants (v0, before vernalization). The remaining plants were vernalized for 6 weeks at 4 °C followed by subsequent growing at 22 °C under LD conditions, and the leaf tissues were continually collected every week (v1–v6 represented vernalization for 1–6 weeks, and pv1 represented post-vernalization for 1 week). The tissues were collected four hours

before darkness to avoid deviations in oscillatory expression of genes and were rapidly frozen by liquid nitrogen before ultra-cryopreservation.

RNA extraction was performed using RNAprep pure Plant Kit (Tiangen, Beijing, China) following the manufacturer's instructions. Subsequently, RNA integrity was assessed by agarose gel electrophoresis, and RNA concentrations were determined by NanoDrop2000 (Thermo Scientific, Shanghai, China).

Transcriptome sequencing

Library construction was performed using NEBNext® Ultra™ RNA Library Prep Kit (NEB, USA) following the manufacturer's instructions. Total RNAs were mixed with 15-min equilibrated mRNA Capture Beads (Vazyme Biotech, Nanjing, China) and denatured for another 5 min. After RNA binding and bead adsorption, the collected beads were washed with Wash Buffer, and the mRNAs were eluted by Tris Buffer at 80 °C. The mRNAs were then recombined to the beads after adding Bead Binding Buffer and eluted again with Tris-HCl after bead adsorption, and the supernatant was removed and the beads were washed. Then, the mRNAs were fragmented, the first-strand cDNA and second-strand cDNA were synthesized, and the cDNA products were mixed with DNA Clean Beads (Vazyme Biotech, Nanjing, China) and purified twice with 80% ethanol. The purified cDNAs were subsequently subjected to end repair, adaptor ligation, and USER enzyme digestion. After fragment selection of the products and PCR reactions, the enriched fragments were subjected to a second purification with 80% ethanol and sequenced in HiSeq 4000 System (Illumina, USA).

The sequenced raw reads were first filtered by removing the adaptor and low-quality reads. The filtered reads were then mapped to the reference genome sequence of CS, and the expression level of the genes was calculated and indicated as Fragments Per Kilobase per Million mapped reads (FPKM).

Real-time PCR analysis

one microgram of total RNAs from different leaf tissues were used as the templates for first-strand cDNA synthesis using PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real-Time) (TaKaRa, Japan). The most differentially-expressed genes discovered by RNA sequencing were reanalyzed by real-time quantitative PCR on ABI 7500 Real-Time PCR System (Thermo Scientific, Shanghai, China). Gene expression was normalized to the housekeeping gene *TaActin*. The primers used for these genes are listed in Table S4, and expression data were analyzed by the $2^{-\Delta\Delta CT}$ method [64].

Abbreviations

VRN
VERNALIZATION; AP1:APETALA 1; FUL:FRUITFULL; SAM:shoot apical meristem; FT:Flowering Locus T; LD:long-day; FDL2:Flowering Locus D-like 2; CO:CONSTANS; PPD1:Photoperiod 1; HAP:Heme Activator Protein; Ghd7:Grain number, plant Height and heading Date 7; COL:CONSTANS-like; TOC1:Timing Of CAB expression 1; CMF:CCT Motif Family; PRR:Pseudo-Response Regulator; LHY:Late Elongated Hypocotyl; CCA1:Circadian Clock-Associated 1; SOC1:Suppressor of Over-expression of CONSTANS 1; Hd1:Heading

Date 1; SD:short-day; APRR:Arabidopsis Pseudo-Response Regulators; HMM:hidden Markov model; ZIM:Zinc-finger protein expressed in Inflorescence Meristem; JTT:Jones-Talor-Thornton; RBCC/TRIM:RING, B-box, coiled-coil/Tripartite motif; CHE:CCA1 Hiking Expedition; FLC:Flowering Locus C; NCBI:National Center for Biotechnology Information

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests

Funding

This research was funded by the HAAFS Agriculture Science and Technology Innovation Project (2019-4-8-06), and the Key Research and Development Program of Hebei (20322912D). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

ZS and WH conceived and designed the study. ZH and ZS carried out the bioinformatic analysis. ZH and LX collected the samples, and performed the molecular experiment. ZH wrote the manuscript. ZS and WH revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Sharma N, Ruelens P, D'hauw M, Maggen T, Dochy N, Torfs S, et al. A flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold regulated in wheat. Plant Physiol.

2017;173(2):1301-15.

2. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci U S A. 2003;100(10):6263-8.
3. Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, et al. The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science. 2004;303(5664):1640-4.
4. Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, et al. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci U S A. 2006;103(51):19581-6.
5. Shitsukawa N, Ikari C, Shimada S, Kitagawa S, Sakamoto K, Saito H, et al. The einkorn wheat (*Triticum monococcum*) mutant, maintained vegetative phase, is caused by a deletion in the *VRN1* gene. Genes Genet Syst. 2007;82(2):167-70.
6. Li C, Lin H, Chen A, Lau M, Jernstedt J, Dubcovsky J. Wheat *VRN1*, *FUL2* and *FUL3* play critical and redundant roles in spikelet development and spike determinacy. Development. 2019;146(14):dev175398.
7. Shimada S, Ogawa T, Kitagawa S, Suzuki T, Ikari C, Shitsukawa N, et al. A genetic network of flowering-time genes in wheat leaves, in which an *APETALA1/ FRUITFULL-like* gene, *VRN1*, is upstream of *FLOWERING LOCUS T*. Plant J. 2009;58(4):668-81.
8. Tanaka C, Itoh T, Iwasaki Y, Mizuno N, Nasuda S, Murai K. Direct interaction between VRN1 protein and the promoter region of the wheat *FT* gene. Genes Genet Syst. 2018;93(1):25-9.
9. Li C, Dubcovsky J. Wheat FT protein regulates *VRN1* transcription through interactions with FDL2. Plant J. 2008;55(4):543-54.
10. Li C, Lin H, Dubcovsky J. Factorial combinations of protein interactions generate a multiplicity of florigen activation complexes in wheat and barley. Plant J. 2015;84(1):70-82.
11. Campoli C, Drosse B, Searle I, Coupland G, von Korff M. Functional characterisation of *HvCO1*, the barley (*Hordeum vulgare*) flowering time ortholog of *CONSTANS*. Plant J. 2012;69(5):868-80.
12. Shaw LM, Turner AS, Laurie DA. The impact of photoperiod insensitive *Ppd-1a* mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum aestivum*). Plant J. 2012;71(1):71-84.
13. Shaw LM, Turner AS, Herry L, Griffiths S, Laurie DA. Mutant alleles of *Photoperiod-1* in wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long days. PLoS One. 2013;8(11):e79459.
14. Wilhelm EP, Turner AS, Laurie DA. Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). Theor Appl Genet. 2009;118(2):285-94.
15. Li C, Distelfeld A, Comis A, Dubcovsky J. Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. Plant J. 2011;67(5):763-73.
16. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat Genet. 2008;40(6):761-7.

17. Distelfeld A, Tranquilli G, Li C, Yan L, Dubcovsky J. Genetic and molecular characterization of the *VRN2* loci in tetraploid wheat. Plant Physiol. 2009;149(1):245-57.
18. Deng W, Casao MC, Wang P, Sato K, Hayes PM, Finnegan EJ, et al. Direct links between the vernalization response and other key traits of cereal crops. Nat Commun. 2015;6:5882.
19. Chen A, Dubcovsky J. Wheat TILLING mutants show that the vernalization gene *VRN1* down-regulates the flowering repressor *VRN2* in leaves but is not essential for flowering. PLoS Genet. 2012;8(12):e1003134.
20. Xu S, Chong K. Remembering winter through vernalisation. Nat Plants. 2018;4(12):997-1009.
21. Strayer C, Oyama T, Schultz TF, Raman R, Somers DE, Más P, et al. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog. Science. 2000;289(5480):768-71.
22. Cockram J, Thiel T, Steuernagel B, Stein N, Taudien S, Bailey PC, et al. Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS One. 2012;7(9):e45307.
23. Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, et al. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. Plant Cell. 2006;18(11):2971-84.
24. Alabadí D, Oyama T, Yanovsky MJ, Harmon FG, Más P, Kay SA. Reciprocal regulation between *TOC1* and *LHY/CCA1* within the *Arabidopsis* circadian clock. Science. 2001;293(5531):880-3.
25. Huang W, Pérez-García P, Pokhilko A, Millar A, Antoshechkin I, Riechmann JL, et al. Mapping the core of the *Arabidopsis* circadian clock defines the network structure of the oscillator. Science. 2012;336(6077):75-9.
26. Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G. Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science. 2004;303(5660):1003-6.
27. Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, et al. *CONSTANS* activates *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* through *FLOWERING LOCUS T* to promote flowering in *Arabidopsis*. Plant Physiol. 2005;139(2):770-8.
28. Tiwari SB, Shen Y, Chang HC, Hou Y, Harris A, Ma SF, et al. The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique cis-element. New Phytol. 2010;187(1):57-66.
29. Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, et al. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. Plant Cell Physiol. 2002;43(10):1096-105.
30. Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, et al. *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. Genes Dev. 2004;18(8):926-36.
31. Ito S, Matsushika A, Yamada H, Sato S, Kato T, Tabata S, et al. Characterization of the APRR9 pseudo-response regulator belonging to the APRR1/TOC1 quintet in *Arabidopsis thaliana*. Plant Cell Physiol. 2003;44(11):1237-45.

32. Sato E, Nakamichi N, Yamashino T, Mizuno T. Aberrant expression of the *Arabidopsis* circadian-regulated *APRR5* gene belonging to the APRR1/TOC1 quintet results in early flowering and hypersensitiveness to light in early photomorphogenesis. Plant Cell Physiol. 2002;43(11):1374-85.
33. Wang H, Zhang Z, Li H, Zhao X, Liu X, Ortiz M, et al. CONSTANS-LIKE 7 regulates branching and shade avoidance response in *Arabidopsis*. J Exp Bot. 2013;64(4):1017-24.
34. Min JH, Chung JS, Lee KH, Kim CS. The CONSTANS-like 4 transcription factor, AtCOL4, positively regulates abiotic stress tolerance through an abscisic acid-dependent manner in *Arabidopsis*. J Integr Plant Biol. 2015;57(3):313-24.
35. Masaki T, Tsukagoshi H, Mitsui N, Nishii T, Hattori T, Morikami A, et al. Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in *Arabidopsis thaliana*. Plant J. 2005;43(1):142-52.
36. Griffiths S, Dunford RP, Coupland G, Laurie DA. The evolution of *CONSTANS-like* gene families in barley, rice, and *Arabidopsis*. Plant Physiol. 2003;131(4):1855-67.
37. IWGSC. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science. 2014;345(6194):1251788.
38. Liu H, Li T, Wang Y, Zheng J, Li H, Hao C, et al. *TaZIM-A1* negatively regulates flowering time in common wheat (*Triticum aestivum* L.). J Integr Plant Biol. 2019;61(3):359-76.
39. Zhan H, Yue H, Zhao X, Wang M, Song W, Nie X. Genome-wide identification and analysis of *MAPK* and *MAPKK* gene families in bread wheat (*Triticum aestivum* L.). Genes (Basel). 2017;8(10):284.
40. Qiao L, Zhang X, Han X, Zhang L, Li X, Zhan H, et al. A genome-wide analysis of the *auxin/indole-3-acetic acid* gene family in hexaploid bread wheat (*Triticum aestivum* L.). Front Plant Sci. 2015;6:770.
41. Woods DP, McKeown MA, Dong Y, Preston JC, Amasino RM. Evolution of *VRN2/Ghd7-like* genes in vernalization-mediated repression of grass flowering. Plant Physiol. 2016;170(4):2124-35.
42. Gangappa SN, Botto JF. The BBX family of plant transcription factors. Trends Plant Sci. 2014;19(7):460-70.
43. Khanna R, Kronmiller B, Maszle DR, Coupland G, Holm M, Mizuno T, et al. The *Arabidopsis* B-box zinc finger family. Plant Cell. 2009;21(11):3416-20.
44. Matsushika A, Makino S, Kojima M, Mizuno T. Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in *Arabidopsis thaliana*: insight into the plant circadian clock. Plant Cell Physiol. 2000;41(9):1002-12.
45. Farré EM, Liu T. The PRR family of transcriptional regulators reflects the complexity and evolution of plant circadian clocks. Curr Opin Plant Biol. 2013;16(5):621-9.
46. Pruneda-Paz JL, Breton G, Para A, Kay SA. A functional genomics approach reveals CHE as a novel component of the *Arabidopsis* circadian clock. Science. 2009;323(5920):1481-5.
47. Fujiwara S, Wang L, Han L, Suh S-S, Salomé PA, McClung CR, et al. Post-translational regulation of the *Arabidopsis* circadian clock through selective proteolysis and phosphorylation of pseudo-response regulator proteins. J Biol Chem. 2008;283(34):23073-83.

48. Murakami M, Yamashino T, Mizuno T. Characterization of circadian-associated APRR3 pseudo-response regulator belonging to the APRR1/TOC1 quintet in *Arabidopsis thaliana*. Plant Cell Physiol. 2004;45(5):645-50.
49. Turner A, Beales J, Faure S, Dunford RP, Laurie DA. The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science. 2005;310(5750):1031-4.
50. Beales J, Turner A, Griffiths S, Snape JW, Laurie DA. A pseudo-response regulator is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet. 2007;115(5):721-33.
51. Niinuma K, Nakamichi N, Miyata K, Mizuno T, Kamada H, Mizoguchi T. Roles of *Arabidopsis PSEUDO-RESPONSE REGULATOR* (*PRR*) genes in the opposite controls of flowering time and organ elongation under long-day and continuous light conditions. Plant Biotechnol. 2008;25(2):165-72.
52. Nakamichi N, Kusano M, Fukushima A, Kita M, Ito S, Yamashino T, et al. Transcript profiling of an *Arabidopsis PSEUDO RESPONSE REGULATOR* arrhythmic triple mutant reveals a role for the circadian clock in cold stress response. Plant Cell Physiol. 2009;50(3):447-62.
53. Reyes JC, Muro-Pastor MI, Florencio FJ. The GATA family of transcription factors in *Arabidopsis* and rice. Plant Physiol. 2004;134(4):1718-32.
54. Lowry JA, Atchley WR. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J Mol Evol. 2000;50(2):103-15.
55. Vanholme B, Grunewald W, Bateman A, Kohchi T, Gheysen G. The tify family previously known as ZIM. Trends Plant Sci. 2007;12(6):239-44.
56. Shikata M, Matsuda Y, Ando K, Nishii A, Takemura M, Yokota A, et al. Characterization of *ArabidopsisZIM*, a member of a novel plant-specific GATA factor gene family. J Exp Bot. 2004;55(397):631-9.
57. Shim JS, Kubota A, Imaizumi T. Circadian clock and photoperiodic flowering in *Arabidopsis*: CONSTANS is a hub for signal integration. Plant Physiol. 2017;173(1):5-15.
58. Michaels SD, Amasino RM. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. Plant Cell. 1999;11(5):949-56.
59. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. 2013;30(12):2725-9.
60. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W9.
61. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Struct Funct Bioinform. 2006;64(3):643-51.
62. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(suppl_2):W202-W8.
63. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194-202.

64. 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(4):402-8.

Figures

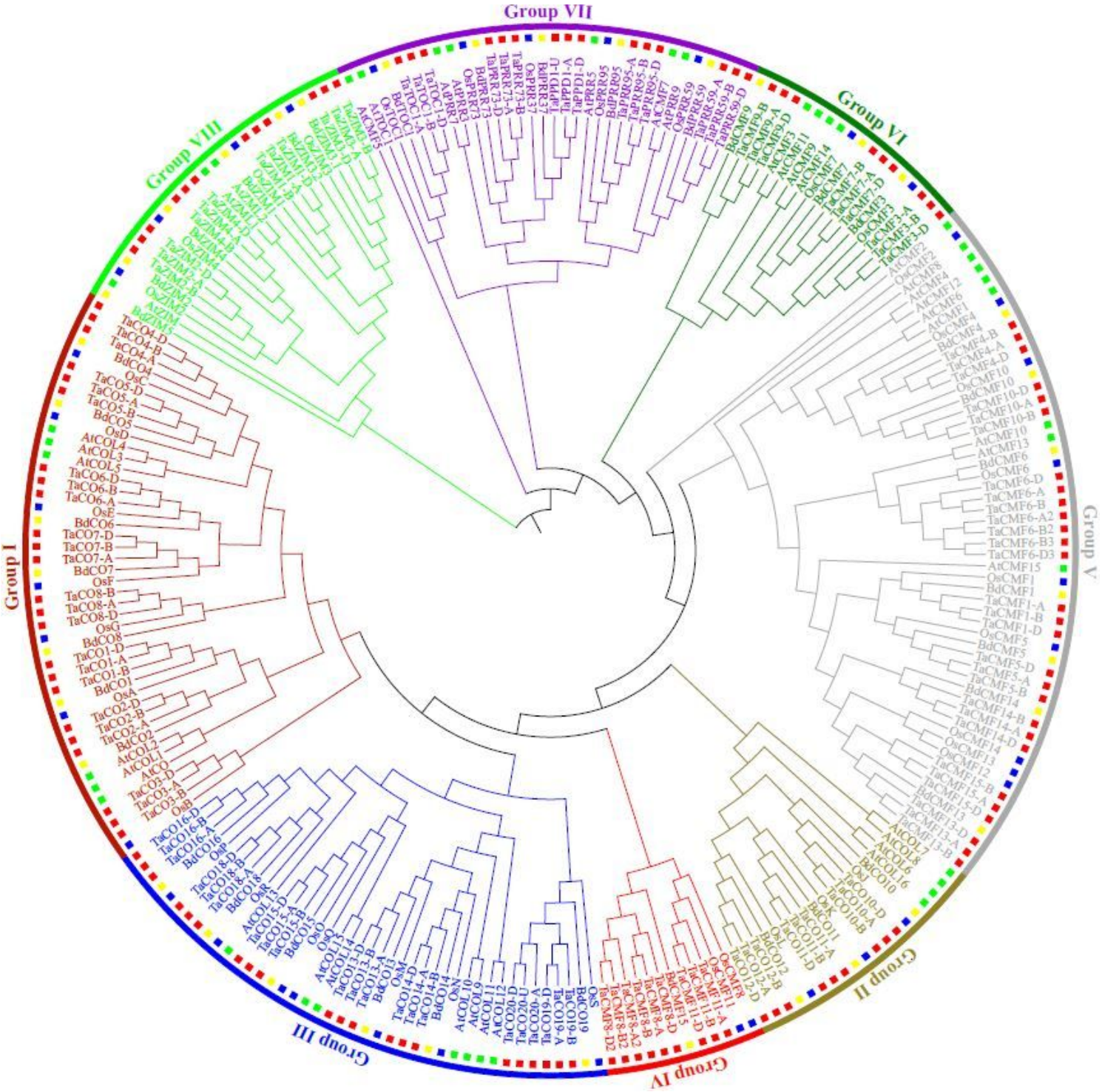

Figure 1

Phylogenetic analysis of CCT proteins in wheat and other plants. The gene ID for each sequence is listed in Table S1 and Table S2. The CCT proteins were classified by Mega 6.0 using the UPGMA method with 1,000 bootstrap replications and the JTT model. The proteins from different plants are represented by different colored squares (wheat, red; Arabidopsis, green; Brachypodium, yellow; rice, blue).

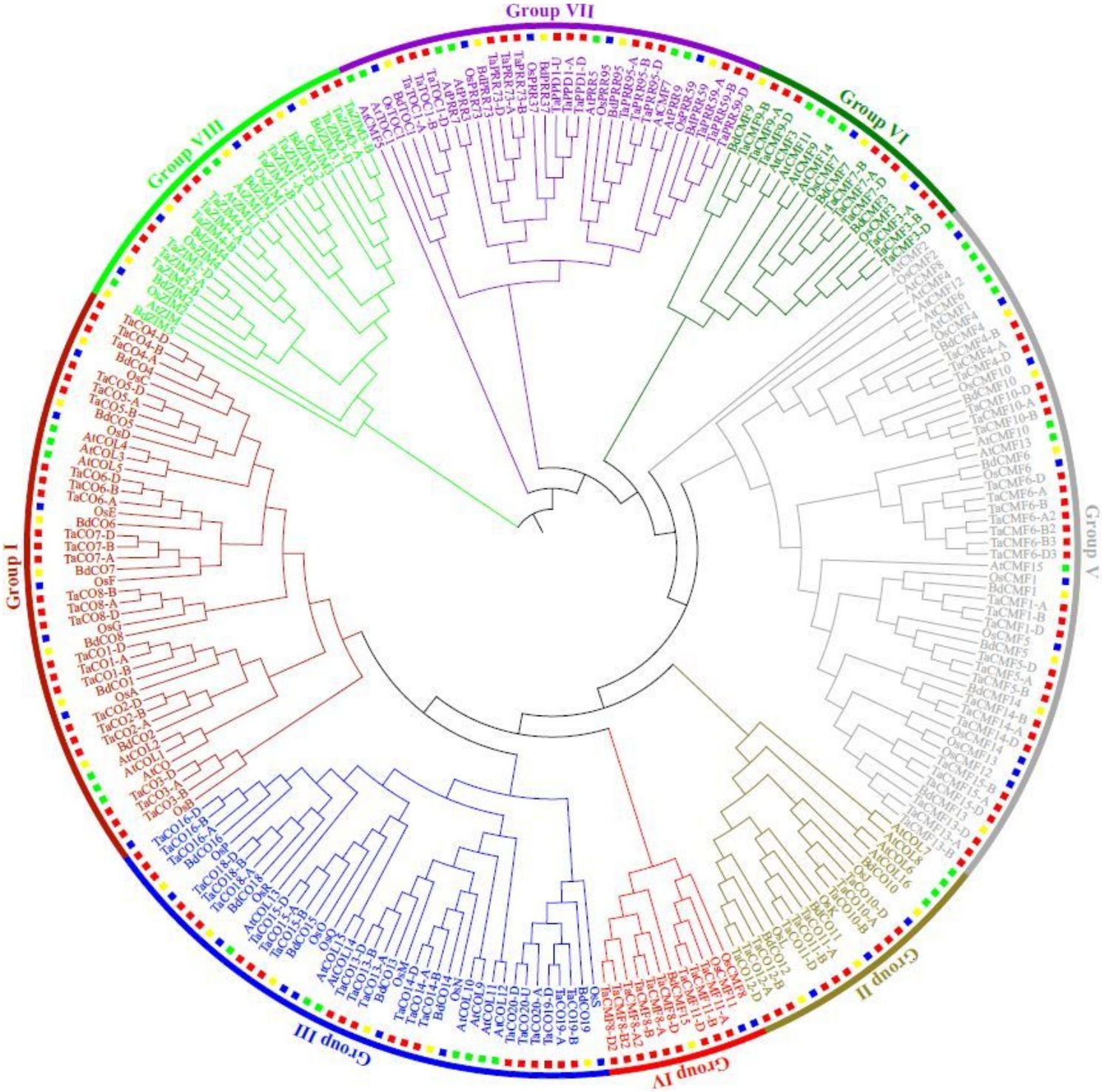

Figure 1

Phylogenetic analysis of CCT proteins in wheat and other plants. The gene ID for each sequence is listed in Table S1 and Table S2. The CCT proteins were classified by Mega 6.0 using the UPGMA method with

1,000 bootstrap replications and the JTT model. The proteins from different plants are represented by different colored squares (wheat, red; Arabidopsis, green; Brachypodium, yellow; rice, blue).

Figure 2

Conserved motif and gene structure analysis of CCT genes in wheat. The motifs and conserved domains were predicted by MEME suite and Batch CD-Search tool, respectively, and displayed by TBtools.

Phylogenetic analysis of the proteins was performed by Mega 6.0 using the UPGMA method, with 1,000 bootstrap replications and the JTT model. The resulting group classification coincided with that in Fig. 1.

Figure 2

Conserved motif and gene structure analysis of CCT genes in wheat. The motifs and conserved domains were predicted by MEME suite and Batch CD-Search tool, respectively, and displayed by TBtools.

Phylogenetic analysis of the proteins was performed by Mega 6.0 using the UPGMA method, with 1,000 bootstrap replications and the JTT model. The resulting group classification coincided with that in Fig. 1.

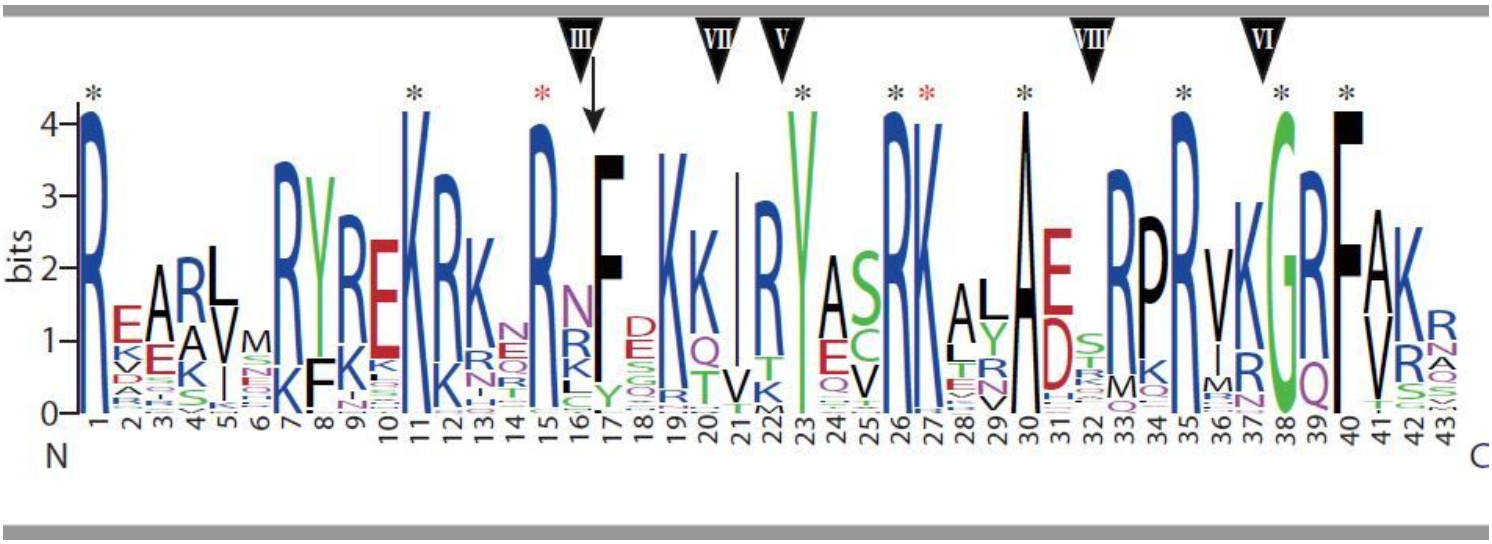

Figure 3

WebLogo representation of the conserved CCT domain. The identical residues are indicated by blank asterisks, and almost identical ones by red. Residue insertion of PL before L17 in TaCMF7 is marked by a black arrow. Intron interruption in the motif is indicated by a numeric triangle that represents the corresponding phylogenetic classification groups.

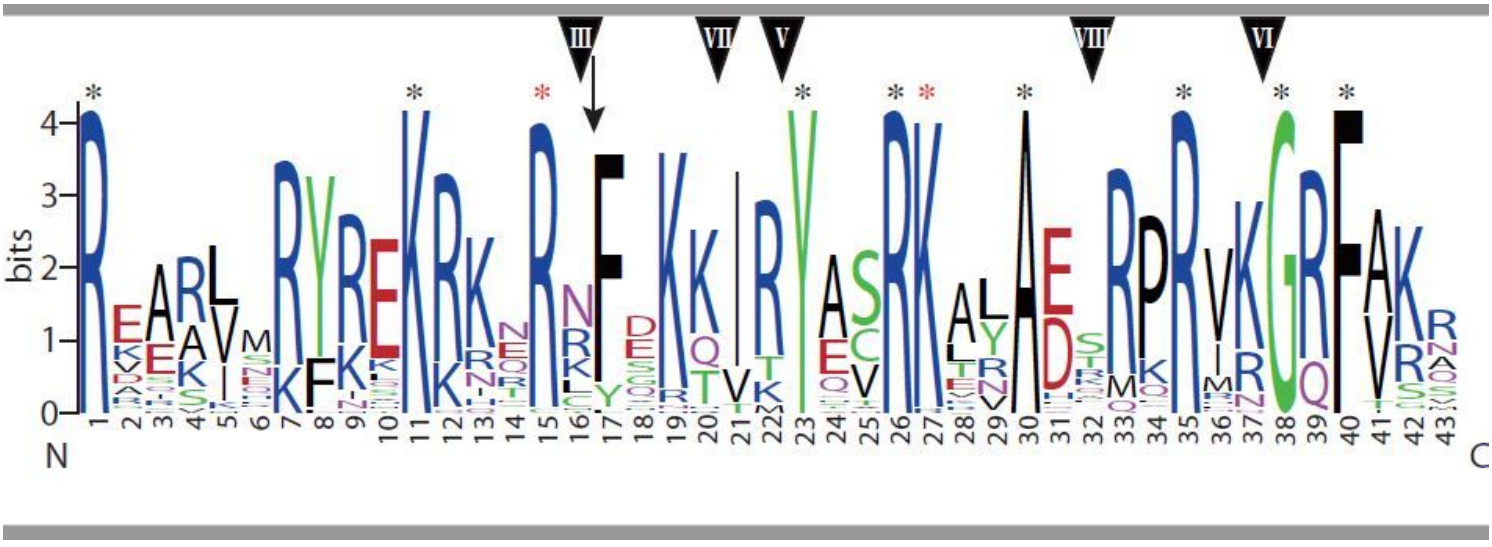

Figure 3

WebLogo representation of the conserved CCT domain. The identical residues are indicated by blank asterisks, and almost identical ones by red. Residue insertion of PL before L17 in TaCMF7 is marked by a black arrow. Intron interruption in the motif is indicated by a numeric triangle that represents the corresponding phylogenetic classification groups.

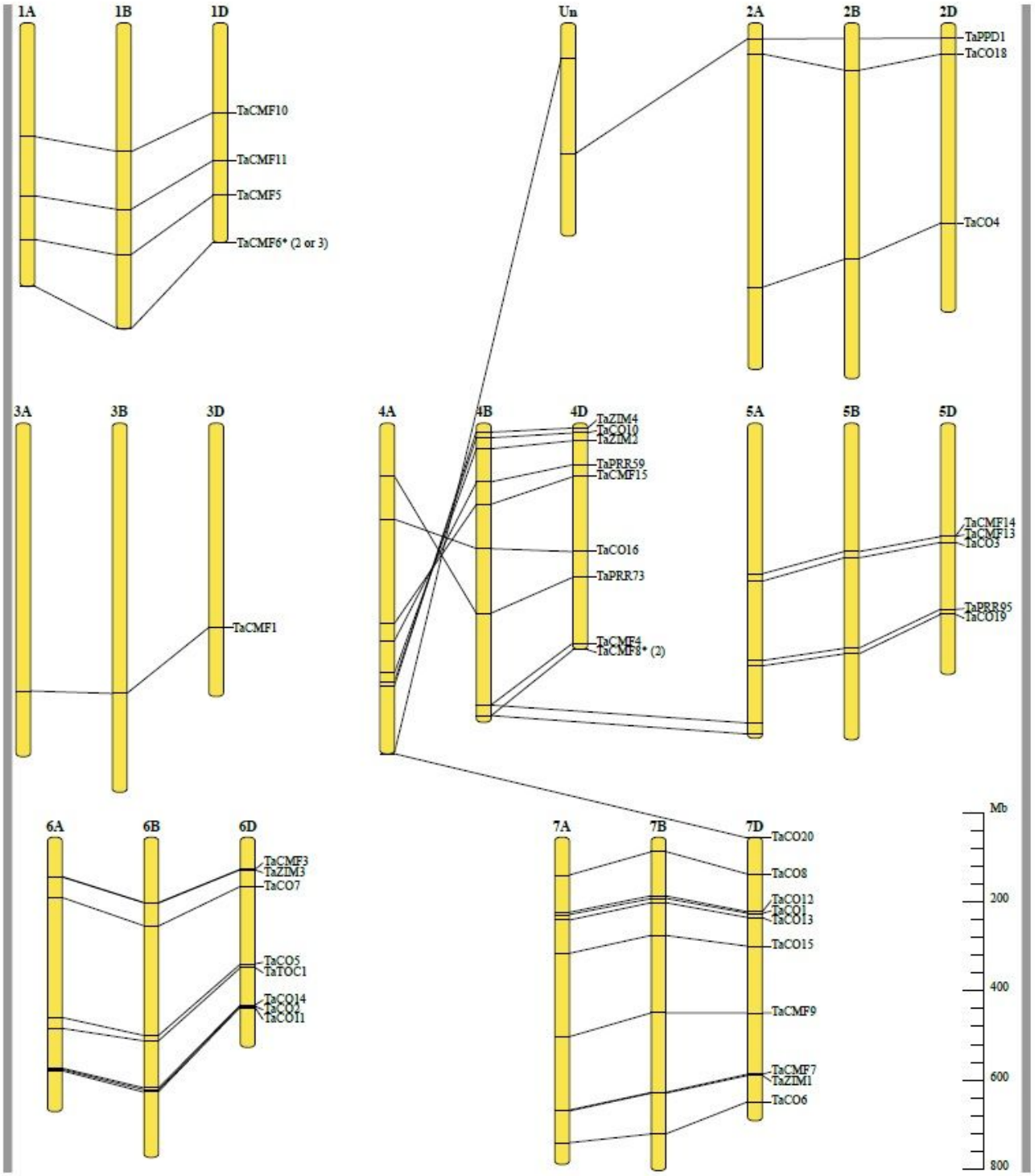

Figure 4

The genome distribution analysis of CCT genes in wheat performed by TBtools. Each gene cluster contained at least three gene members distributed among three genomes, which is marked on the right. The tandemly arranged genes in a cluster are indicated by asterisks, and the copy number in each subgenome is shown in parentheses.

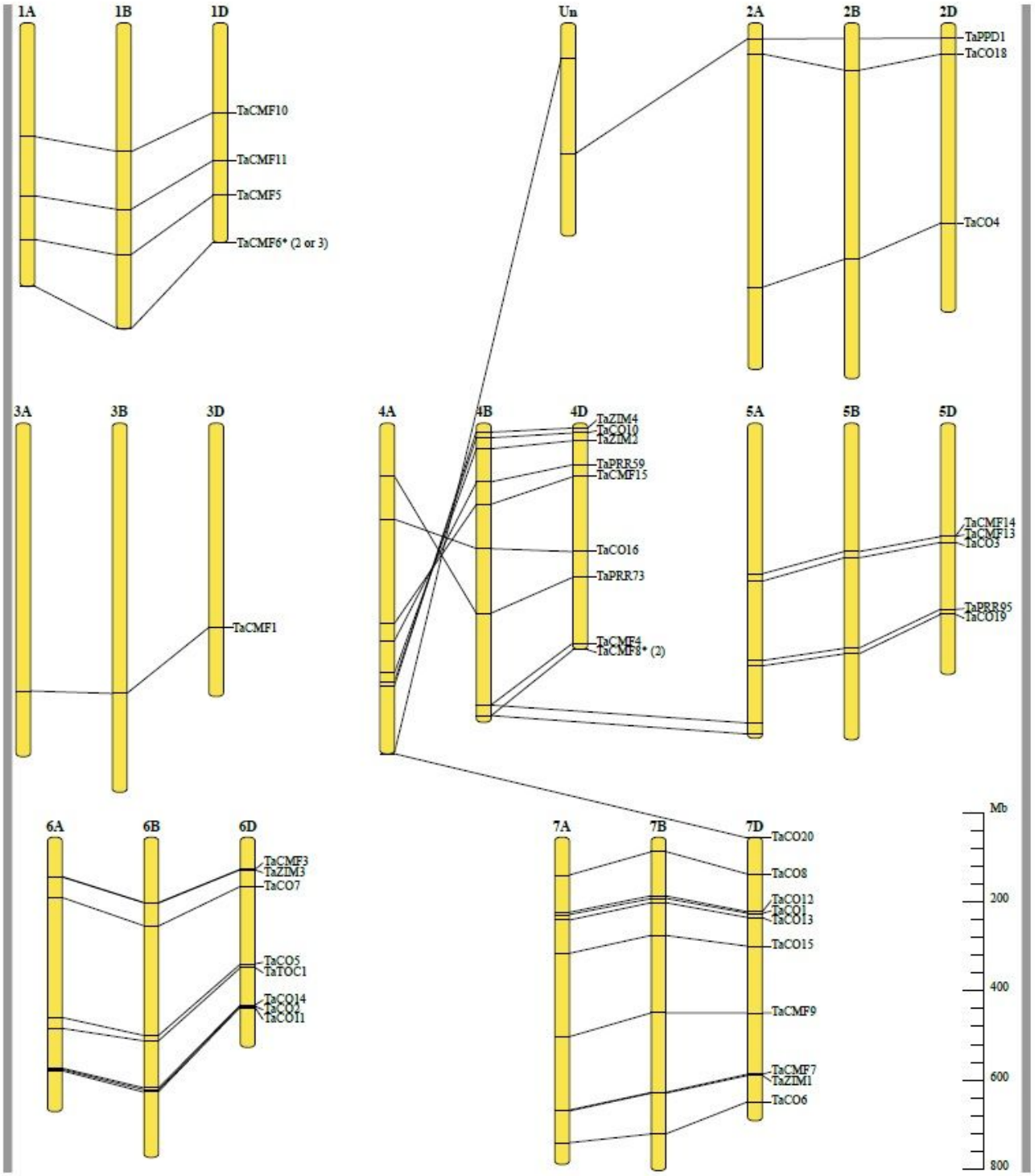

Figure 4

The genome distribution analysis of CCT genes in wheat performed by TBtools. Each gene cluster contained at least three gene members distributed among three genomes, which is marked on the right. The tandemly arranged genes in a cluster are indicated by asterisks, and the copy number in each subgenome is shown in parentheses.

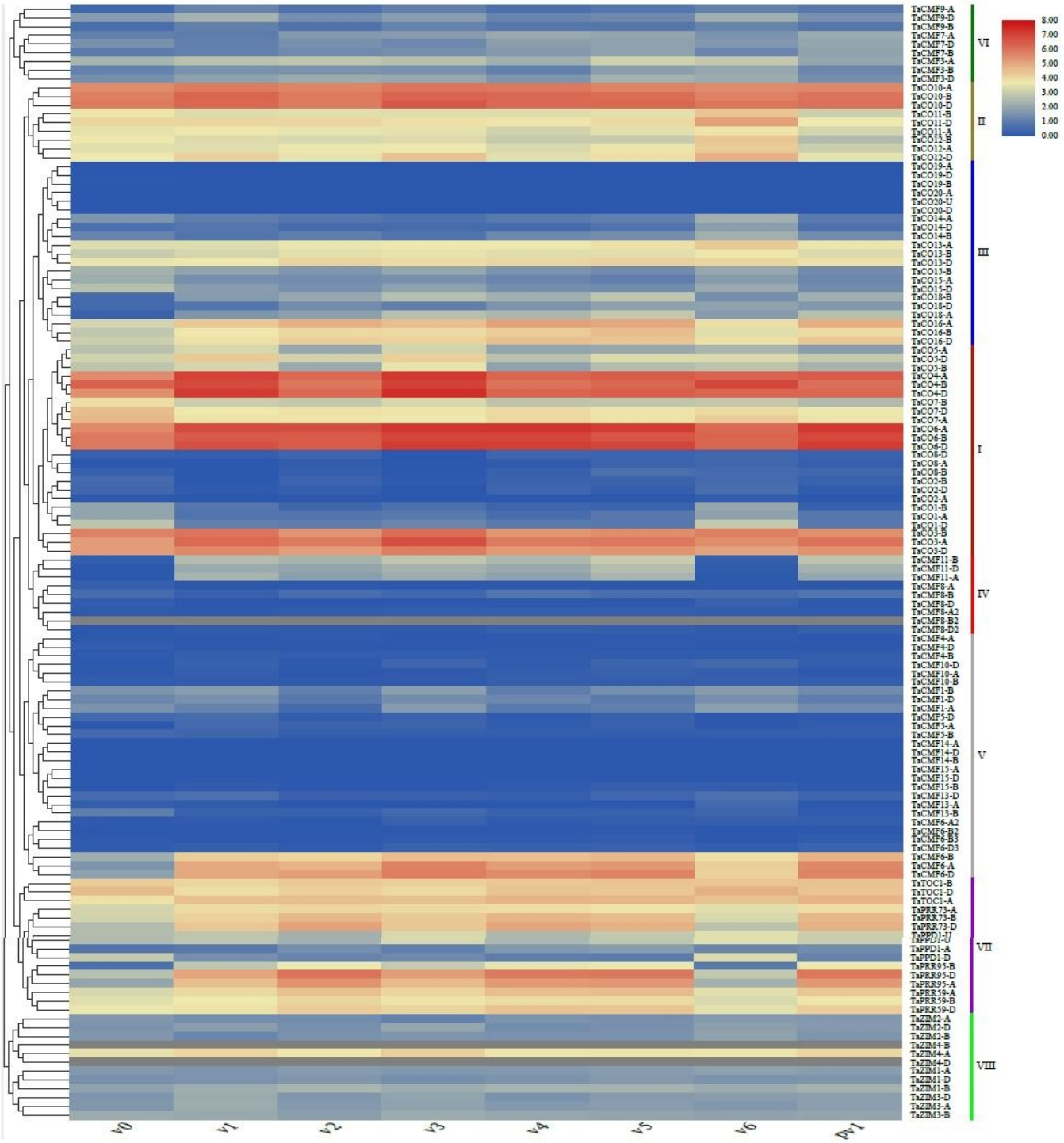

Figure 5

Expression profiling of CCT genes in wheat before, under, and after vernalization by RNA sequencing. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1). Phylogenetic analysis of proteins was performed by Mega 6.0 using the UPGMA method, with 1,000 bootstrap replications and the JTT model. The resulting group classification coincided with that of Fig. 1.

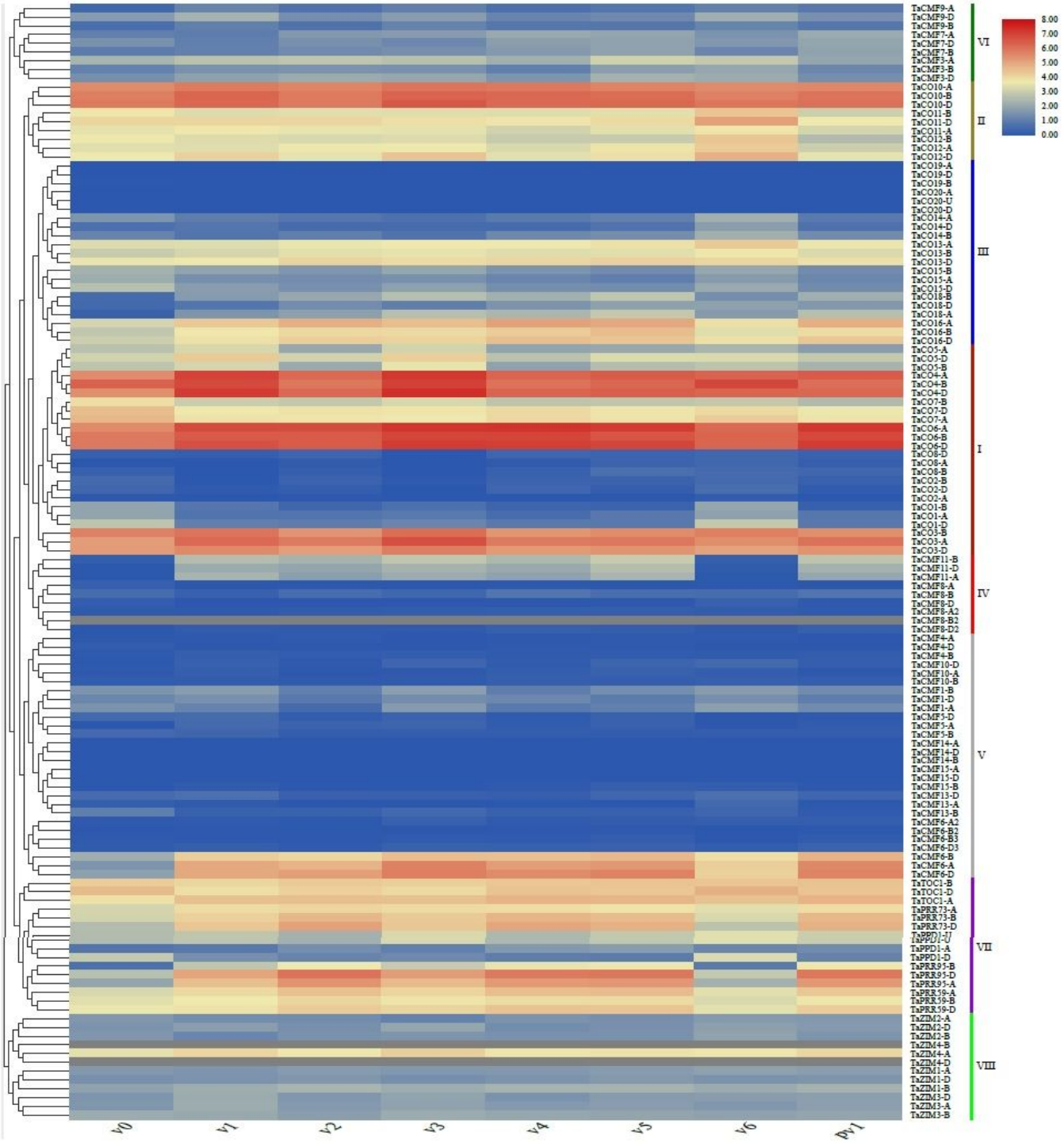

Figure 5

Expression profiling of CCT genes in wheat before, under, and after vernalization by RNA sequencing. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1). Phylogenetic analysis of proteins was performed by Mega 6.0 using the UPGMA method, with 1,000 bootstrap replications and the JTT model. The resulting group classification coincided with that of Fig. 1.

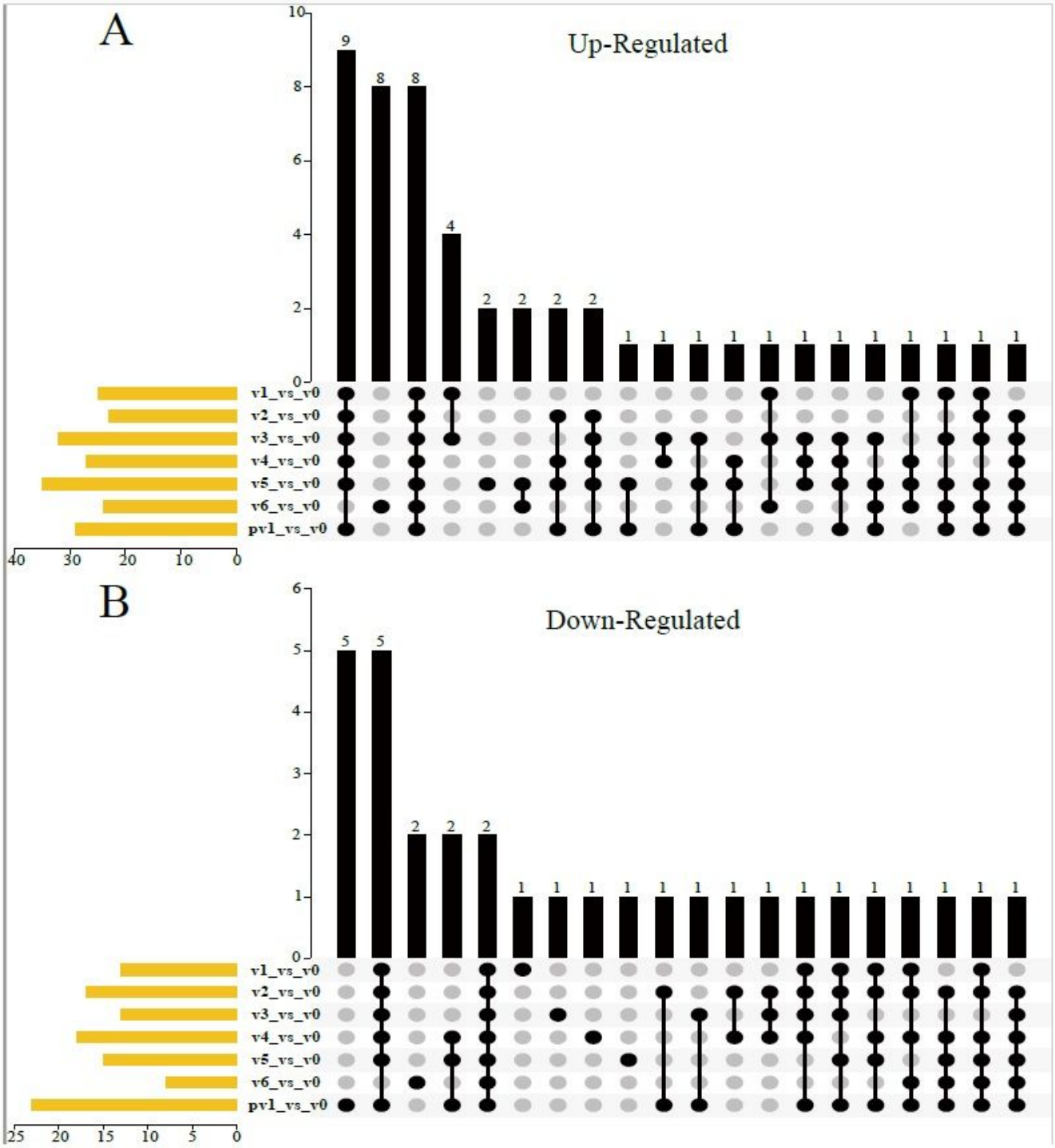

Figure 6

UpSet plot diagram of (A) upregulated and (B) downregulated CCT genes under/after versus pre-vernalization. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1).

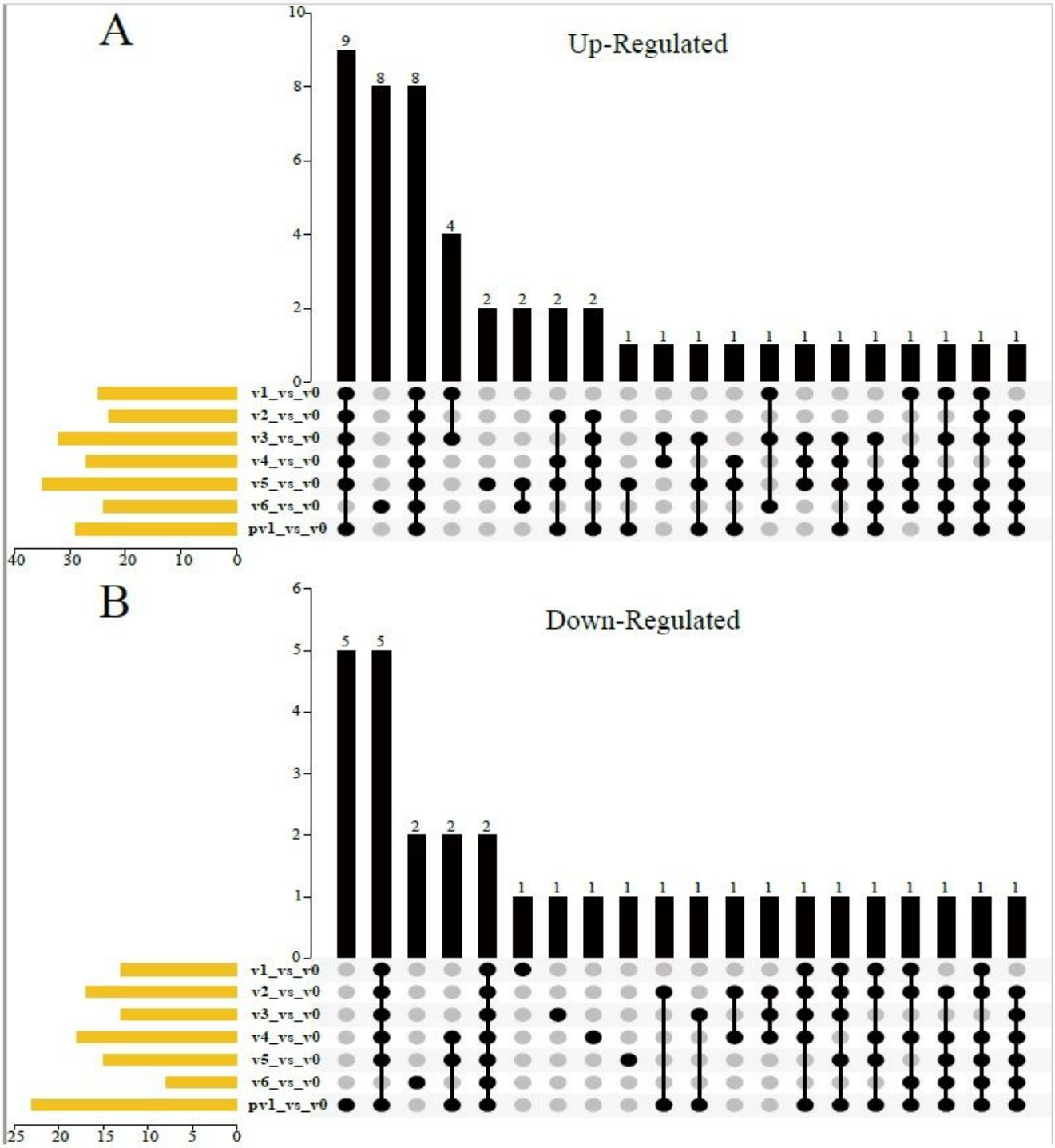

Figure 6

UpSet plot diagram of (A) upregulated and (B) downregulated CCT genes under/after versus pre-vernalization. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1).

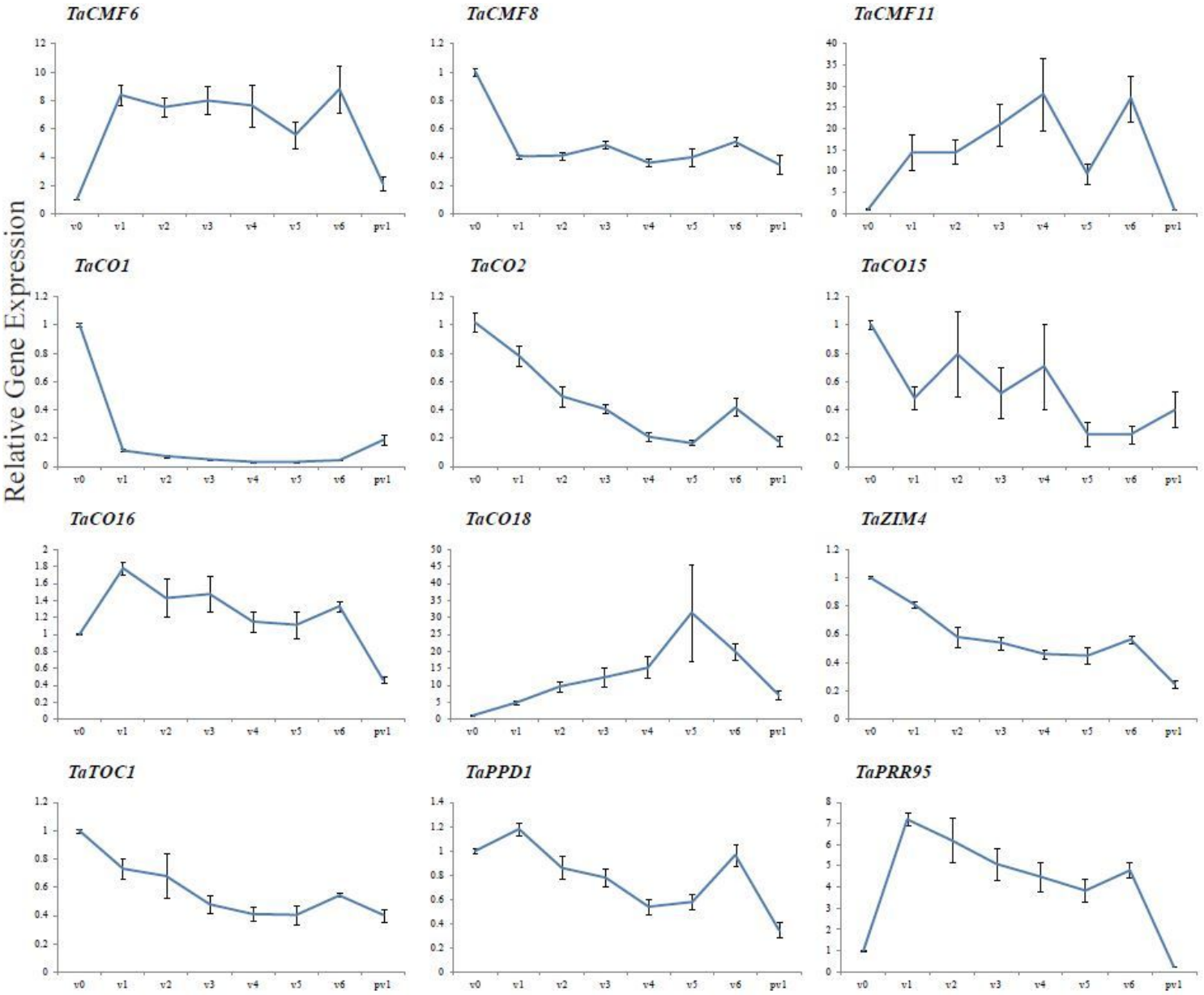

Figure 7

Expression profiling of CCT genes in wheat before, under, and after vernalization using real-time analysis. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1). The expression level of each gene before vernalization was used as standard, and the relative expression of genes under and after vernalization was calculated using the 2-ΔΔCT method [64]. The results shown here were normalized to the expression of the housekeeping gene TaActin. Three biological replicates for each sample were performed, and gene expression deviations are indicated by standard error (SE).

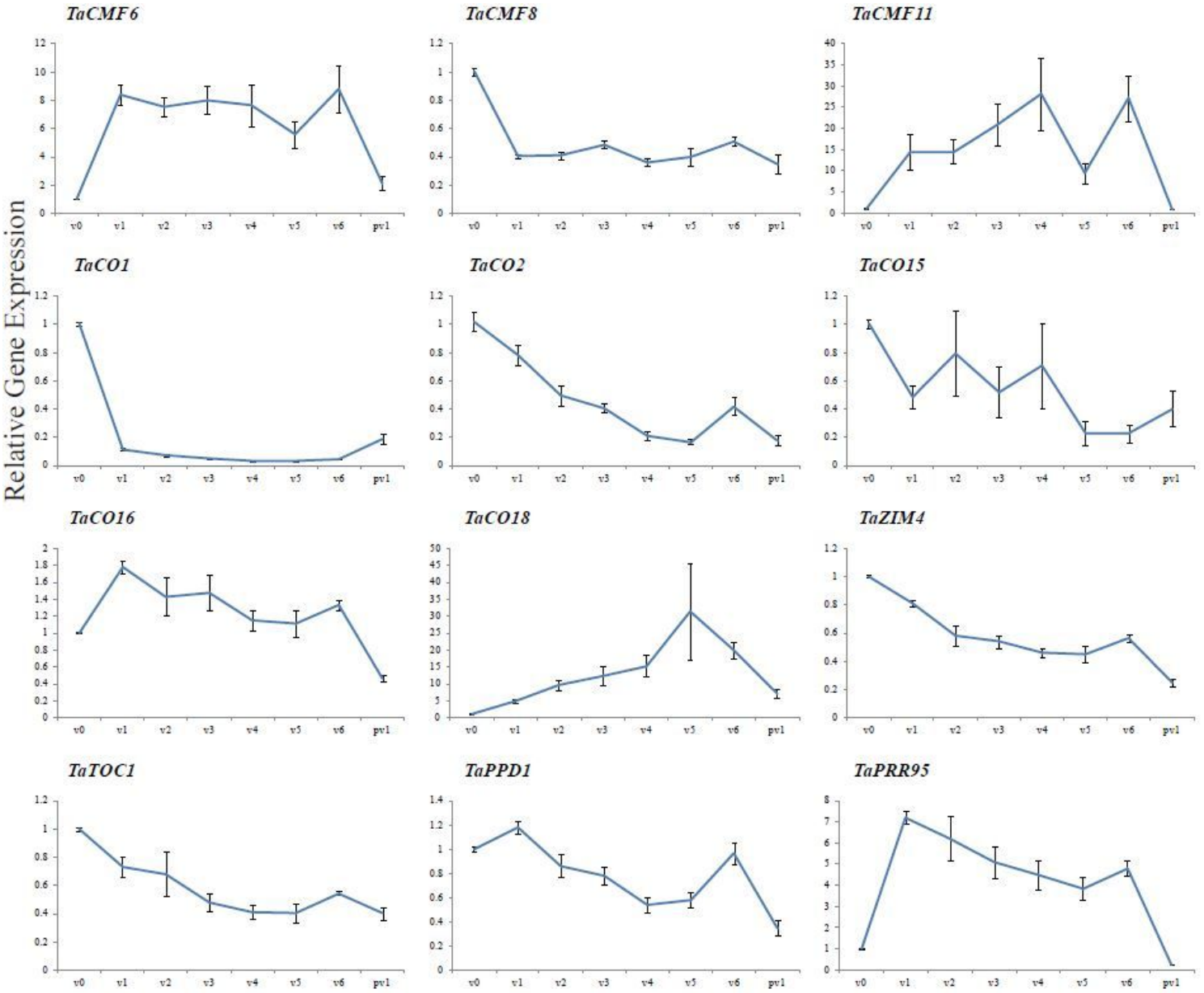

Figure 7

Expression profiling of CCT genes in wheat before, under, and after vernalization using real-time analysis. The samples were harvested before vernalization (v0), under vernalization for 1–6 weeks (v1–v6), and the next week after 6 weeks of vernalization (pv1). The expression level of each gene before vernalization was used as standard, and the relative expression of genes under and after vernalization was calculated using the 2-ΔΔCT method [64]. The results shown here were normalized to the expression of the housekeeping gene TaActin. Three biological replicates for each sample were performed, and gene expression deviations are indicated by standard error (SE).

Supplementary Files

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

- FigureS1.pdf
- FigureS1.pdf
- FigureS2.pdf
- FigureS2.pdf
- TableS1.xlsx
- TableS1.xlsx
- TableS2.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS4.xlsx