BMC Genomics

RESEARCH Open Access

Genome-wide identification of CaWD40 proteins reveals the involvement of a novel complex (CaAN1-CaDYT1-CaWD40-91) in anthocyanin biosynthesis and genic male sterility in *Capsicum annuum*

Peng Tang[1†], Jingcai Huang[1], Jin Wang[1], Meiqi Wang[1], Qing Huang[1], Luzhao Pan[1,2†] and Feng Liu[1*]

Abstract

Background The WD40 domain, one of the most abundant in eukaryotic genomes, is widely involved in plant growth and development, secondary metabolic biosynthesis, and mediating responses to biotic and abiotic stresses. WD40 repeat (WD40) protein has been systematically studied in several model plants but has not been reported in the *Capsicum annuum* (pepper) genome.

Results Herein, 269, 237, and 257 *CaWD40* genes were identified in the Zunla, CM334, and Zhangshugang genomes, respectively. *CaWD40* sequences from the Zunla genome were selected for subsequent analysis, including chromosomal localization, phylogenetic relationships, sequence characteristics, motif compositions, and expression profiling. CaWD40 proteins were unevenly distributed on 12 chromosomes, encompassing 19 tandem duplicate gene pairs. The 269 *CaWD40s* were divided into six main branches (A to F) with 17 different types of domain distribution. The *CaWD40* gene family exhibited diverse expression patterns, and several genes were specifically expressed in flowers and seeds. Yeast two-hybrid (Y2H) and dual-luciferase assay indicated that CaWD40-91 could interact with CaAN1 and CaDYT1, suggesting its involvement in anthocyanin biosynthesis and male sterility in pepper.

Conclusions In summary, we systematically characterized the phylogeny, classification, structure, and expression of the *CaWD40* gene family in pepper. Our findings provide a valuable foundation for further functional investigations on *WD40* genes in pepper.

Keywords WD40 transcription factors, Pepper, MYB-bHLH-WD40 (MBW), Functional diversification, Classification

†Peng Tang and Luzhao Pan contributed equally to this work.

*Correspondence:
Feng Liu
jwszjx@hunau.edu.cn
[1]Engineering Research Center for Germplasm Innovation and New Varieties Breeding of Horticultural Crops, Key Laboratory for Vegetable Biology of Hunan Province, College of Horticulture, Hunan Agricultural University, Changsha, China
[2]Horticultural Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China

Background

The WD-repeat (WDR) proteins are commonly referred to as the WD40 repeat proteins or WD40 proteins owing to the highly conserved domain of approximately 40 amino acid stretches [1–4]. This conserved domain is characterized by a glycine histidine (Gly-His, GH) dipeptide at the N-terminal (start) and a tryptophan-aspartic acid (Trp-Asp, WD) dipeptide at the C-terminal (end) [1, 4]. Structurally, the WD40 proteins typically have multiple tandem repeated WD motifs, forming a series of four-stranded and antiparallel β sheets [2, 5, 6]. Functionally, the WD40 domain has no catalytic activity of its own and is considered an adaptor protein that recruits other factors to form protein-protein or protein-deoxyribonucleic acid (DNA) complexes [7–9]. WD40 proteins are involved in histone modification [10], DNA damage repair [11], signal transduction [9], and abiotic stress responses [12].

The MYB and basic helix-loop-helix (bHLH) transcription factors and WD40 repeat protein form an MYB-bHLH-WD40 (MBW) complex, which jointly regulates flavonoid biosynthesis, constituting among the most general functions of the *WD40* genes. For example, the TTG1-bHLH-MYB complex promotes flavonoid accumulation and the formation of trichomes in *Arabidopsis* seedlings [13, 14]. Similarly, *TT2* (*AtMYB123*), *TT8* (*AtbHLH42*), and *TTG1* (*WD40*) synergistically regulate the biosynthesis of the proanthocyanidins pigments in the *Arabidopsis* seed coat [15]. The expression of several structural genes is regulated by MBW complex, including dihydroflavonol-4-reductase (DFR), anthocyanin dioxygenase/ anthocyanin synthase (LDOX/ANS), BANYULS (BAN) [15, 16]. In pepper, CaTTG1 forms an MBW complex with CaAN1 and CaGL3 to control anthocyanin synthesis by activating the expression of structural genes in the synthesis pathway of anthocyanin [17].

Some studies have demonstrated the participation of the WD40 in various biological processes involved in reproductive development in *Arabidopsis*. For instance, the XPO1-interacting WD40 protein 1 (XIW1) can interact with ABA insensitive 5 (ABI5) in the nucleus and regulate seed germination and growth [18]. The *JGB* gene with the WD40 domain is a negative regulator of pollen germination [19]. *WDR55* is essential for multiple plant developmental processes, including gametogenesis, seed, and endosperm development [20]. The SWA1 protein is necessary for the normal functioning of the mitotic division cycle by regulating cellular metabolism [21].

The WD40 repeats are widely found in all eukaryotes but are rare in prokaryotic organisms [3, 22–26]. In plants, the WD40 proteins form a superfamily, which are functionally diverse. With the completion of whole genome sequencing of different species, the WD40 protein family has been systematically identified in several [3, 25–29]. The 237 potential WDR proteins from *Arabidopsis* have been identified. These show conservation and divergence in structure and function [3]. In *Triticum aestivum*, 743 WD40 proteins have been grouped into 5 clusters and 11 subfamilies and may be involved in reproductive development and mediating responses to multiple stresses [25]. In *Oryza sativa*, 200 *OsWD40* genes have been identified and are suggested to perform diverse functions through a complex network [26]. Li et al. [27] found 191 *WDRs* in cucumber (*Cucumis sativus* L.). In *Arabidopsis*, *WD40s* show strong conservation during the evolutionary process. Feng et al. [28] identified 220 *WD40s* in the peach genome, and corresponding proteins localized to different subcellular structures. Chen et al. [29] analyzed 164 barley HvWD40 proteins and revealed that the *WD40* gene formed an MBW complex with MYB and bHLH and participated in anthocyanin synthesis. These results suggest that the *WD40* gene family has diverse gene numbers, structures, and functions.

Pepper, a typical spicy vegetable, belongs to the genus Capsicum of the Solanaceae family. It originated in South and Central America [30]. Capsicum species have various morphological characteristics, and several pepper species lack anthocyanins [17, 31, 32]. In recent years, several genes controlling anthocyanin synthesis have been identified in pepper. Zhang et al. [32] reported that a splice acceptor site in *CabHLH1* resulted in a frameshift mutation in a mutant of deficient anthocyanin and fertility reduction of anther (*rpf1*). Two genes, *CaAN3* and *Ca3GT*, controlling anthocyanin biosynthesis in pepper fruit were successively reported [33, 34]. CaHY5 regulated anthocyanin accumulation in pepper hypocotyl by directly binding to the promoter of downstream structural genes in the anthocyanin pathway [35]. A structural gene of anthocyanin synthesis, flavonoid 3',5'-hydroxylase (F3'5'H), controls pigment biosynthesis in stems and anthers in pepper [17].

In this study, 269 *CaWD40* members were identified in the Zunla pepper genome. The physicochemical properties, sequence structural characteristics, phylogenetic construction, and tissue expression of the *CaWD40* gene family were analyzed by bioinformatics methods. Based on yeast two-hybrid (Y2H) technology, we verified the potential involvement of a novel MBW complex (CaAN1-CaDYT1-CaWD40-91) in anthocyanin synthesis. These results are expected to facilitate the understanding of biological functions and molecular mechanisms of WD40 proteins and provide a basis for studying the anthocyanin regulatory network in pepper.

Results

The *CaWD40* gene family shows great variations in sequence length and physicochemical properties

The Hidden Markov Model (HMM) profile of the WD40 repeat domain (PF00400) was used as a query sequence to search WD40 proteins in the Zhangshugang, Zunla, and CM334 genome databases. Preliminarily, 267, 276, and 245 *WD40s* were retrieved from the Zhangshugang, Zunla, and CM334 reference genomes, respectively. All examined pepper WD40 members were verified, and partial *WD40* genes with incomplete WD40 domains, considered pseudogenes, were further removed. Finally, 257, 269, and 237 *WD40s* were identified in the Zhangshugang, Zunla, and CM334, respectively. Since the knowledge of WD40 members in the Zunla genome is relatively complete, the information on this genome was subject to subsequent analyses. For convenience, pepper WD40 candidates were named *CaWD40-1*–*CaWD40-269* according to their positions on the 12 chromosomes in the Zunla reference genome. Detailed information on gene name, chromosome location, protein length, molecular weight, isoelectric point (pI), and domain is listed in Table S1.

CaWD40 genes varied greatly in terms of protein length and physicochemical properties. The length of these *CaWD40* fragments ranged from 100 aa (*CaWD40-85*, *CaWD40-212*) to 3,595 aa (*CaWD40-87*). The molecular weights (MWs) ranged from 10.75 kDa (*CaWD40-85*) to 399.13 kDa (*CaWD40-87*). The isoelectric point (pIs) ranged from 4.21 (*CaWD40-267*) to 9.96 (*CaWD40-151*) (Table S1). The prediction results for the subcellular localization of the 269 *CaWD40s* were as follows: 139 in the nucleus, 56 in the chloroplasts, 55 in the cytoplasm, six in the plasma membrane, five in the cytoskeleton, four in the mitochondria, two in the vacuole membrane, one in the endoplasmic reticulum, and one in the peroxisome.

Chromosome localization reveals multiple tandem duplication events

The 245 out of the 269 *CaWD40* genes were randomly distributed across 12 chromosomes in the Zunla genome, and the remaining 24 *CaWD40s* were not anchored to chromosomes (Fig. 1A). The vast majority of *CaWD40* members were concentrated on the distal end of each chromosome, with a few genes in the middle of the chromosome. The number of *CaWD40s* was highest on chromosome 3 (34), followed by chromosome 1 (33); the lowest was on chromosome 5 (8) (Fig. 1B). The distribution of *CaWD40s* indicated the accumulation of some genes on specific chromosomes.

We analyzed tandem duplications in the *CaWD40* gene family in pepper. Nineteen pairs (I to XIX) of genes among 269 *CaWD40s* were identified as tandem duplication events (Table S2), crucial in gene functional differentiation. Tandem duplication events were unevenly distributed across the 10 chromosomes and were usually concentrated in regions with high gene density. Some tandem repeat genes belonged to the same phylogenetic branch, while others belonged to different branches, hinting at functional differentiation and conservation among the *CaWD40* gene family (Fig. 2).

The combination of phylogenetic and structural analysis reveals the diversity of *CaWD40s*

In this study, the protein sequences of the 269 predicted CaWD40s were used to generate an unrooted tree to analyze the classification of the *CaWD40* gene family in pepper (Fig. 2). Phylogenetic analysis showed that the *CaWD40* gene family was categorized into six main branches (branch A to F) with 15, 22, 15, 42, 68, and 107 members, respectively. Among them, branch F was divided from F1 to F4 (Fig. 2). Branches A, B, and C were the earliest clusters in the phylogenetic tree, and genes in the same branch shared similar motif compositions (Fig. 2 and Fig S1). Similarly, most genes in branches D, E, and F followed this rule. Combined with chromosome localization analysis, gene pairs with tandem duplication were mostly distributed in the E and F branches, suggesting that the differentiation of gene function was gradual along with gene duplication (Figs. 1 and 2).

The structural domain of the predicted 269 *CaWD40s* was counted using SMART (Table 1, Table S1) to analyze the relationship between domain composition and phylogenetic classification. The *CaWD40* gene family was categorized into 17 distinct classes based on domain composition (Table 1). Among them, 191 members contained only a single WD40 domain, and the remaining had at least one or more WD40 domains and other domains.

Gene structural characteristics can elucidate the relationship between gene function and evolution. The number of exons and introns in the 269 *CaWD40s* varied greatly (Fig. S1A and B). Twenty-three *CaWD40* genes (8.55%) had only one exon and were intronless, including *CaWD40-80* and *CaWD40-91* (Fig. S1B). A total of 146 *CaWD40s* (54.28%) contained 1–10 exons; 83 *CaWD40s* (30.86%) contained 11–20 exons, and the remaining 17 *CaWD40s* (6.32%) contained more than 20 exons. Specifically, *CaWD40-218* and *CaWD40-181* had two exons; the former belonged to branch A and the latter to branch F (Fig. 2, Fig. S1B). All three genes (*CaWD40-62*, *CaWD40-124*, *CaWD40-160*) belonged to branch A. *CaWD40-124* had four exons. *CaWD40-160* and *CaWD40-62* had nine and 29 exons, respectively. These results suggest that the different exon-intron compositions of 269 *CaWD40s* are randomly distributed across different branches of the phylogenetic tree, implying a

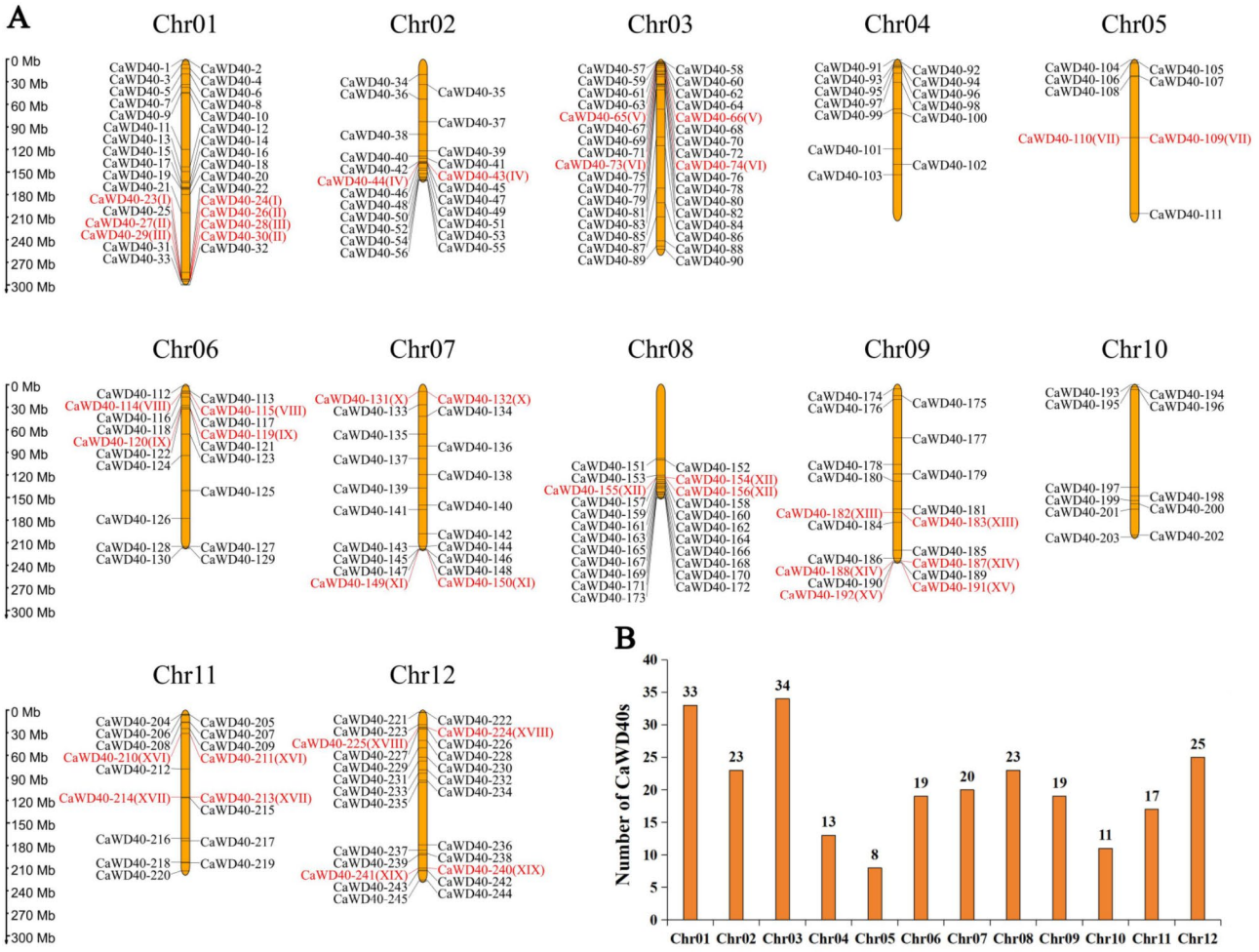

Fig. 1 Chromosome mapping and gene number of CaWD40 proteins in the Zunla genome. (**A**) Chromosome mapping and tandem duplication events of *CaWD40s*. Chromosome numbers are represented on the top, and the scale is shown on the left. Predicted tandem duplicate pairs are indicated in red font, and the categories are marked in parentheses using Roman numerals. (**B**) Statistical analysis for *CaWD40s* on 12 chromosomes

high diversity of the *CaWD40* gene family in the process of evolution.

Using the Multiple Em for Motif Elicitation (MEME) tool, we identified 10 putative conserved motifs (motif 1 to motif 10) in the *CaWD40* gene family (Fig. S1C). No common motifs were observed in 269 *CaWD40s*. Among the 10 motifs, motif 1 was the most widely distributed, found in 264 *CaWD40* genes. Some motifs were present in only a few genes, and motif 9 and motif 10 were present in only 4 (CaWD40-57/117/233/256) and 6 (*CaWD40-86/89/98/147/195/222*) *CaWD40* genes, respectively. Some *CaWD40s* with the same exon–intron structure exhibited a similar motif composition, such as *CaWD40-69* and *CaWD40-266* (Fig. S1C).

Synteny analysis of *CaWD40s* among three genomes

A genome-wide synteny analysis was performed among the *CaWD40s* in the Zunla, Zhangshugang, and CM334 genomes. Zunla and Zhangshugang had 225 gene pairs co-linked and Zunla and CM334 had 178 gene pairs co-linked (Fig. 3, Table S3). Zunla and Zhangshugang showed a close synteny for the *CaWD40* gene, suggesting that these two species may share more similar functions. The relationship of species origin indicates that *Capsicum annuum* cv. Zhangshugang and Zunla are derived from China [31, 56]; the CM334 pepper is from South Korea [57].

CaWD40s exhibit distinct expression profiles in various tissues and stages

By RNA-seq analysis, expression profiles of the putative *CaWD40* genes at various tissues and developmental stages of Capsicum line 6421 were analyzed (Fig. 4). *CaWD40s* showed varied expression patterns, which could be divided into six types (type i–vi) based on a hierarchical clustering of their expressional characteristics (Fig. 4, Table S4). Specifically, type i contained 48 genes, most relatively highly expressed in different tissues

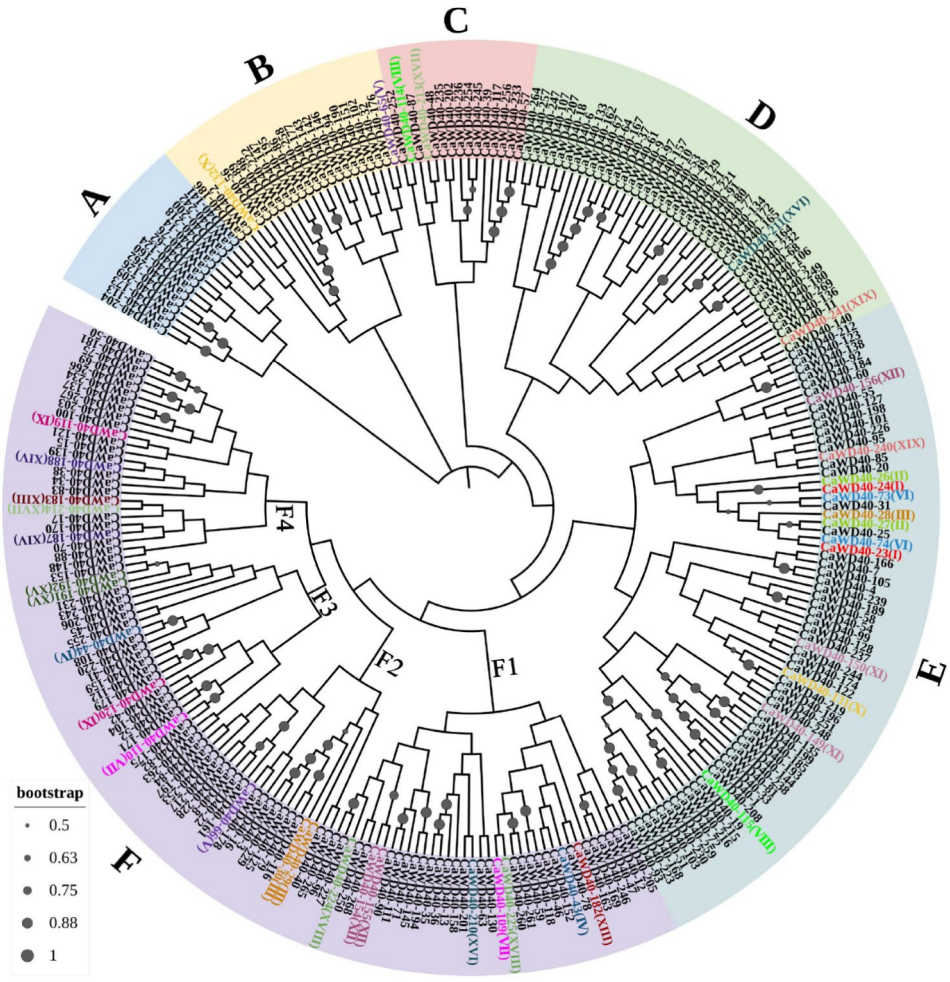

Fig. 2 Phylogenetic relationships among *CaWD40s* in the Zunla genome. Different branches are marked by shading with diverse colors. The gray dots represent the bootstrap, and those with a bootstrap greater than 0.5 are shown. Predicted tandem duplicate pairs are marked with fonts of diverse colors

and periods. Type ii comprised 101 genes with medium or low expression in various tissues or stages. Type iii included 21 *CaWD40s*, highly expressed in almost all tissues. Type iv included 49 genes showing low-to-no expression in most tissues. Type v had 16 members with tissue-specific high expression in specific tissues or stages. For example, four genes (*CaWD40-51*, *CaWD40-135*, *CaWD40-199*, and *CaWD40-223*) were highly expressed in flower and stamen tissues and showed almost no expression in other tissues. *CaWD40-93* and *CaWD40-125* showed high expression levels in fruit or placenta. Group vi contained 34 genes with low expression levels in all tissues.

Verification of expression in the eight *CaWD40s*

We randomly selected eight *CaWD40s* from each hierarchical cluster in the heatmap to determine their expression in different tissues by RT-qPCR to verify the expression of the *CaWD40* gene family. The expression trend identified by RT-qPCR was similar to that of RNA-seq (Figs. 4 and 5). For example, the expression of *CaWD40-26* was not detected in diverse tissues; *CaWD40-69* and *CaWD40-266* were highly expressed in almost all examined tissues and stages (Fig. 5), and *CaWD40-135* and *CaWD40-223* tended to be expressed later during flower development.

A novel MBW complex may be involved in anthocyanin biosynthesis and genic male sterility

Previous reports have shown that WD40 typically acts as part of the MBW complex to regulate anthocyanin biosynthesis [13–15, 17]. A recent report has demonstrated that the MBW complex of CaAN1-CaGL3-CaTTG1 affects pigment accumulation and male gametophyte development in the *Cha1* mutant of pepper [17]. In this study, *CaWD40-91* (*Capana04g000080*) protein showed high phylogenetic and structural similarity with *CaTTG1* (*CaWD40-80/Capana03g001813*) (Figs. 4 and 5, Fig S2, 3). Subcellular localization results also demonstrated CaWD40-91 protein localization in the nucleus

Table 1 Domain composition of the 17 classes of 269 CaWD40s

Classes	Domain composition	Gene number
Class 1	OnlyWD40	191
Class 2	WD40 + LISH + CTLH	8
Class 3	WD40 + Ubox or Fbox	6
Class 4	WD40 + Utp12 or Utp13	5
Class 5	WD40 + NLE	3
Class 6	WD40 + PH_BEACH + Beach or Beach or DUF4704 + PH_BEACH + Beach	3
Class 7	WD40 + RING + TYKc or RING or CLH + RING	3
Class 8	WD40 + Transmembrane region	3
Class 9	WD40 + Coatomer_WDAD + COPI_C or Coatomer _WDAD	3
Class 10	WD40 + Katanin_con80	3
Class 11	WD40 + PFU	2
Class 12	WD40 + STYKc	2
Class 13	WD40 + BCAS3	2
Class 14	WD40 + CAF1C_H4-bd	2
Class 15	WD40 + LRRcap	2
Class 16	WD40 + ZnF_C3H1 or ZnF_C2HC	2
Class 17	Other	29

and cytoplasm (Fig. 6A), consistent with the localization results for CaTTG1 [17]. Therefore, we reasonably speculate that *CaWD40-91* might function similarly to the *CaTTG1* gene.

To confirm whether CaWD40-91 can bind to the CaGL3 and CaAN1 proteins, a Y2H experiment was performed. CaWD40-91 could interact with CaAN1 but not with CaGL3 (Fig. 6B). Considering that CaWD40-91 could indeed bind to CaAN1, we replaced other bHLH proteins to verify the protein interaction. A recent study from our laboratory showed that a candidate bHLH (CaDYT1) activated *CaDFR* expression to control anthocyanin synthesis and male sterility in pepper (unpublished data). Therefore, we verified the interaction between CaWD40-91 and CaDYT1 proteins. The results indicated that CaWD40-91 interacted with the CaDYT1 and CaAN1 proteins, resulting in the formation of a novel MBW complex (CaAN1-CaDYT1-CaWD40-91) (Fig. 6B and C), which may regulate synergistically anthocyanin accumulation and genic male sterility of pepper.

Mutation in *CaDYT1* causes male sterility in pepper

We previously cloned the *CaDYT1* gene by crossing the *dyt1* mutant with wild type (WT) to obtain the F_2 genetic population (unpublished data). The sequencing results are shown in Fig. 7A. The second exon of the *CaDYT1* gene had a 7 bp deletion in the *dyt1* mutant, resulting in a frameshift mutation. Phenotypic analysis showed that the anthers of the *dyt1* mutant were thinner and yellower compared with WT (Fig. 7B). Scanning electron microscopy (SEM) showed a round and full WT anther, with a closely arranged convex surface structure. Several pollen grains are produced following anther dehiscence. In contrast, *dyt1* anthers were small and folded, no pollen grains were observed, and only impurities and starch grains were found (Fig. 7B).

Discussion

Diversity and conservation of CaWD40 proteins in pepper

In plants, WD40 is not directly involved in recognizing target gene promoters, but it plays a role in the MBW complexes with MYB and bHLH proteins [9, 15]. The MBW complex is important in the anthocyanin biosynthesis pathway [17]. In the MBW complex, the WD40 protein is usually located at the center of the triadic structure. It may protect the MBW complex by preventing other transcriptional regulators from binding to MYB or bHLH [15]. The functions of WD40 are diverse. In addition to anthocyanin synthesis, WD40 proteins are critical for pollen germination [36], drought stress [37], RNA processing [38], regulation of plant biological clock [39], skin and fur development [40], and other processes.

In this study, 269 *CaWD40s* were identified in the Zunla pepper genome. Interestingly, 237 and 257 *CaWD40s* were found in CM334 and Zhangshugang genome, respectively. Different pepper cultivars may lead to the difference in *CaWD40* numbers in the three genomes [31, 56, 57]. Finally, the 269 *CaWD40* members

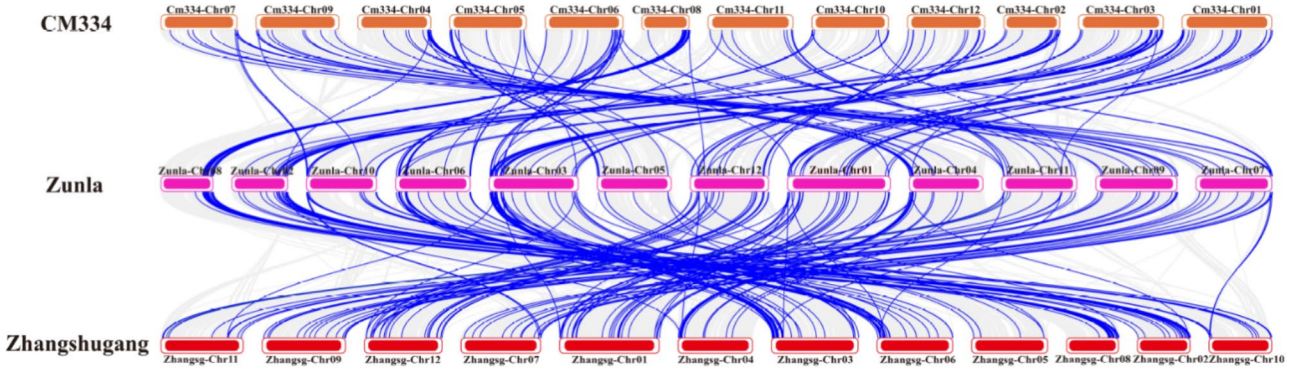

Fig. 3 Synteny analysis for WD40s among different pepper genomes. The identified collinear genes are shown with blue lines

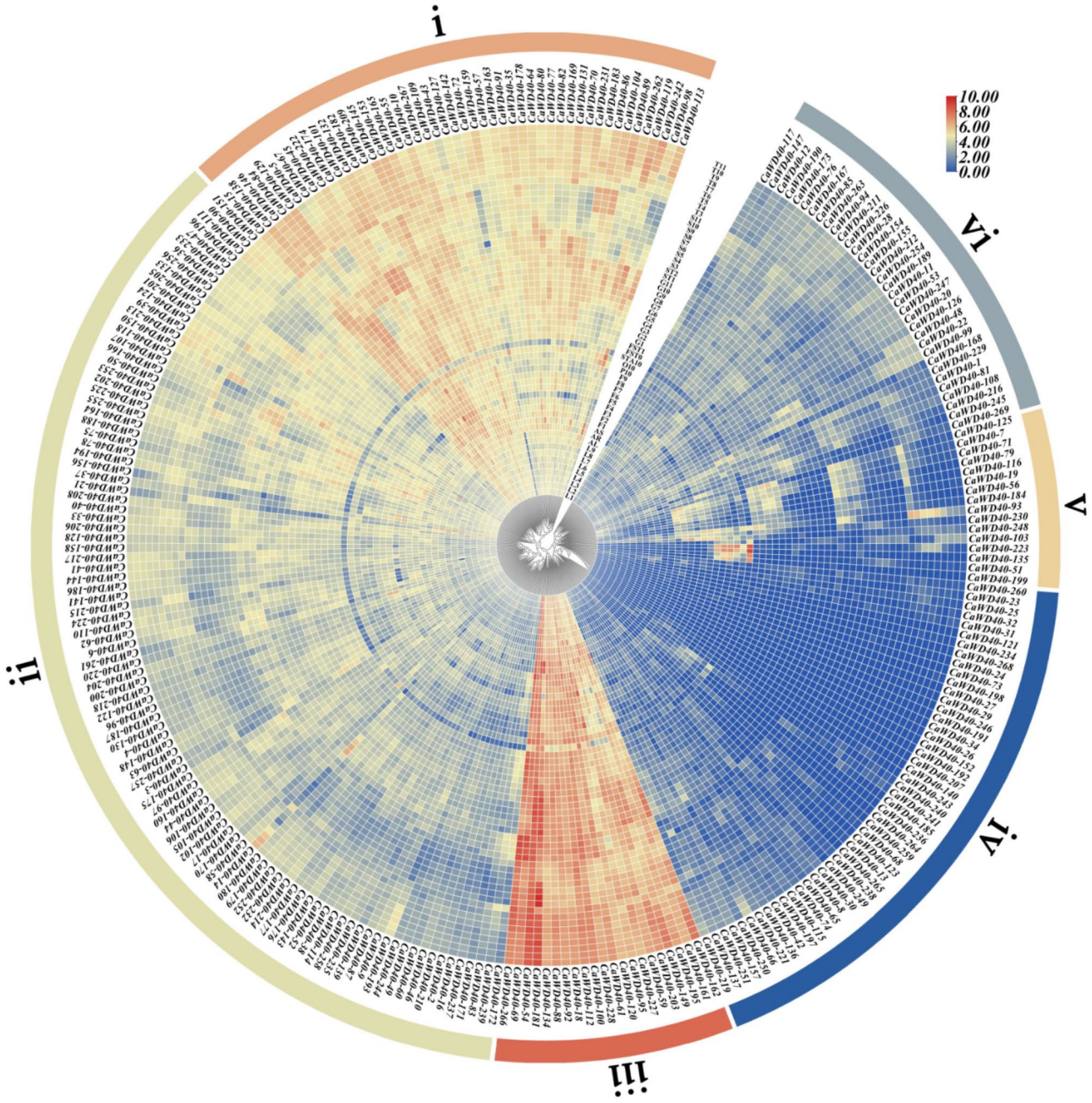

Fig. 4 Expression analysis of *CaWD40* genes in various tissues and stages based on RNA-seq data. Cluster hierarchy based on expression levels. Heat map; blue/yellow/red colors represent low/medium/high expressions, respectively

from the Zunla genome were analyzed. These genes were divided into six main branches based on their gene structure and motif compositions. The CaWD40 members within the branch showed high similarity in domain compositions and sequences, suggesting a conserved gene function.

The *WD40* gene family has been reported in many species. The number of members varies greatly between species, such as 207 in tomato [41], 237 in *Arabidopsis* [3], 200 in rice [26], 191 in cucumber [27], 743 in *Triticum aestivum* [25], 164 in barley [29], and 220 in Peach [28]. Based on published genomic databases, the genome sizes of pepper, tomato, and *Arabidopsis* are predicted to be 3.5 Gb, 0.95 Gb, and 0.125 Gb, respectively. The above results indicate that the number of *WD40s* is not related to the genome size of a species. We found a high polarization in physical and chemical properties: protein length ranges from 100 aa (CaWD40-85, CaWD40-212) to 3,595 aa (CaWD40-87), and molecular weights ranged from 10.75 kDa (CaWD40-85) to 399.13 kDa (CaWD40-87).

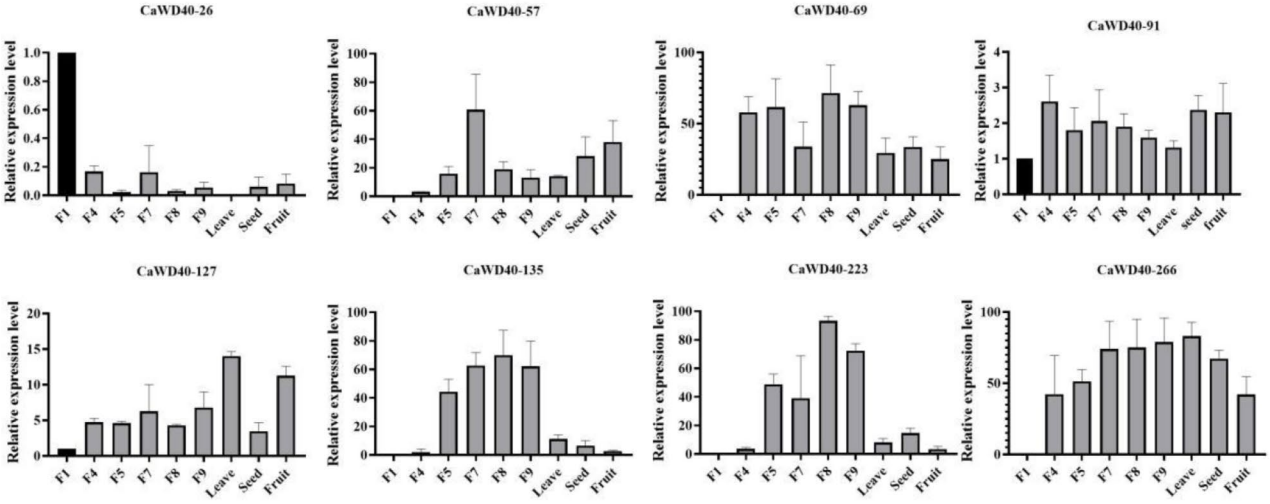

Fig. 5 Relative expressions of *CaWD40-26*, *CaWD40-57*, *CaWD40-69*, *CaWD40-91*, *CaWD40-127*, *CaWD40-135*, *CaWD40-223*, and *CaWD40-266*. F1, F4, F5, F7, F8, and F9 represent different stages of buds in Capsicum line 6421

These results are consistent with previously reported studies [25, 26, 28, 29].

Nineteen pairs (40 *CaWD40s*) were identified as tandem duplication events, and these genes were unevenly distributed in 10 chromosomes (Fig. 1). Tandem repeat genes from a pair also sometimes clustered on the different phylogenetic branches (Fig. 2), suggesting functional divergence in the *CaWD40* gene family. Sequence characteristics revealed that the number of exons and introns in the 269 *CaWD40* gene family varied greatly, that is, from intronless to 30 introns (Fig. S1B and C). However, some *CaWD40s* possessed similar exon–intron and motif compositions (*CaWD40-69* and *CaWD40-266*) (Fig. S1B and C). Overall, CaWD40 proteins were structurally dynamic, which could determine the functional diversity of the *WD40* gene. In humans, *WD40s* are involved in multiple cellular networks, many of which have been implicated in diseases [42]. Two proteins, WDR5 and EED, are involved in chromatin complexes [42]. In *Arabidopsis*, several WD40 proteins are related to gametogenesis, seed, and endosperm development [19, 20]. The WD40 gene family is a superfamily of plant species. Some supergene families, such as WRKY, bHLH, and MYB, are usually accompanied by structural and sequence similarities and differences in the evolution process [19, 43–46].

CaWD40-91 is a potential candidate involved in anthocyanin biosynthesis and pollen fertility

In plants, flavonoids are synthesized through a conserved metabolic pathway, producing several secondary metabolites (SMs), including flavones, anthocyanins, and proanthocyanidins [47]. These SMs have pivotal roles in growth and development as they provide plants with different pigments that scavenge reactive oxide species (ROS) [32], defend against UV damage [48], or mediate plant-microbe interactions [49]. The SM biosynthetic pathway is among the most extensively studied in plants. Several structural genes related to anthocyanin biosynthesis, including *CHS*, *CHI*, *F3H*, *F3'5'H*, *DFR*, *ANS*, and *UFGT*, are conserved in most plants [47, 49, 50]. In recent years, several reports suggest a close association between flavonoids and pollen fertility [32, 48, 51–53].

Wang et al. [17] showed that CaTTG1, CaAN1, and CaGL3 formed an MBW complex, which regulated anthocyanin synthesis. In *Arabidopsis*, TTG1 can bind to AtMYB123 and AtbHLH42 transcription factors to regulate the biosynthesis of the proanthocyanidins pigments in the seed coat [13, 14]. DYSFUNCTIONAL TAPETUM 1 (DYT1), acting as an upstream regulator of tapetum development, is required for pollen formation in *Arabidopsis* and pepper [54, 55]. Our recent study suggests that the *cha1* mutant (CaTTG1) exhibits partially aborted pollen grains (unpublished). In this study, CaWD40-91, a paralog protein of CaTTG1 (named CaWD40-80 in this paper), was confirmed to interact with CaAN1 and CaDYT1. In general, genes that are highly homologous in sequence and structure tend to have similar functions. More interestingly, we recently found that DYT1 could bind to DFR synergistically and regulate anthocyanins and male sterility in pepper (unpublished). Based on the above results, CaWD40-91, CaAN1, and CaDYT1 may form an MBW complex, which can regulate the formation of anthocyanins and pollen fertility.

Conclusions

From bioinformatics analysis, we identified 269, 237, and 257 CaWD40 genes in the Zunla, CM334, and Zhangshugang genomes, respectively. A total of 269 candidates from the Zunla genome were analyzed, and they were named *CaWD40-1* to *CaWD40-269* according to

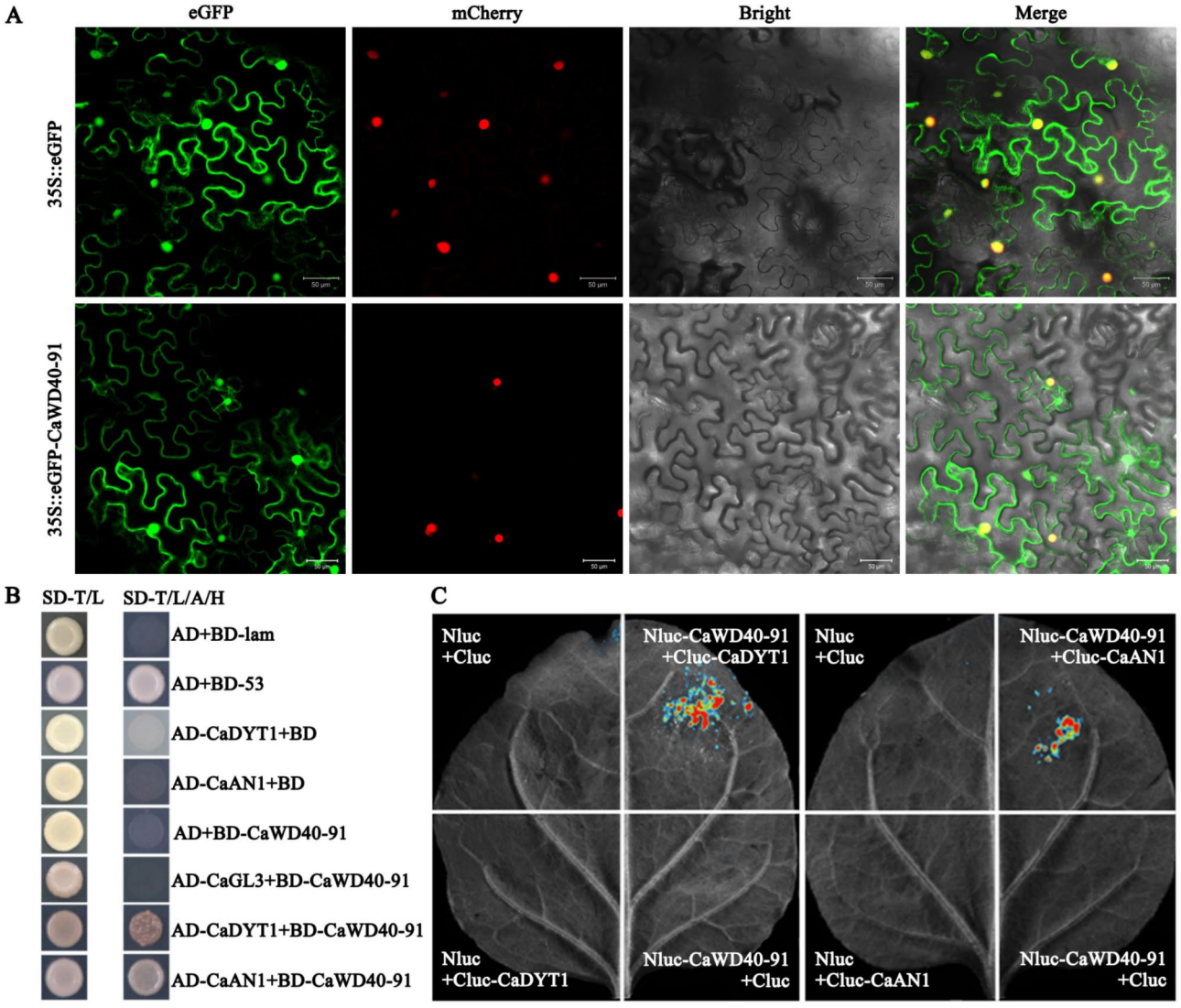

Fig. 6 Subcellular localization and interaction. (**A**) Subcellular localization of CaWD40-91. Scale bars represent 50 μm. (**B**) Interactions of CaWD40-91 with CaAN1 and CaDYT1 via a yeast two-hybrid assay. (**C**) The dual-luciferase reporter gene assay.

their arrangement on the chromosomes. The protein length and physico-chemical properties varied greatly. Protein lengths ranged from 100 to 3,595 aa, and molecular weights ranged from 10.75 kDa to 399.13 kDa. The analysis revealed 245 CaWD40 proteins clustered unevenly along chromosomes, with 24 CaWD40s remaining unlinked to chromosomes. Further analysis revealed 19 tandem duplicate pairs that clustered in different branches of the phylogenetic tree. Sequence analysis indicated that the numbers of exon (1–31) and intron (0–30) varied greatly in the *CaWD40* gene family, and the composition of conserved motifs was diverse. However, several *CaWD40s* showed highly conserved sequence characteristics, expression patterns, and biological function. Finally, we found that *CaWD40-91* showed high homology with *CaTTG1* based on phylogenetic relationship, gene structure, and domain composition, with further speculations that they may share similar functions. A recent report revealed that CaTTG1 could interact with CaGL3 and CaAN1 to form an MBW complex, which could regulate anthocyanin synthesis in pepper [17]. In this study, we found that CaWD40-91 interacted with CaAN1 but not with CaGL3. Interestingly, we observed that CaWD40-91 could interact with CaDYT1 from the Y2H assay. Moreover, CaDYT1 activated CaDFR to synergistically regulate anthocyanin synthesis and male sterility. Therefore, we speculate that CaWD40-91 can interact with CaAN1 and CaDYT1 to control pigment accumulation and sterility in pepper.

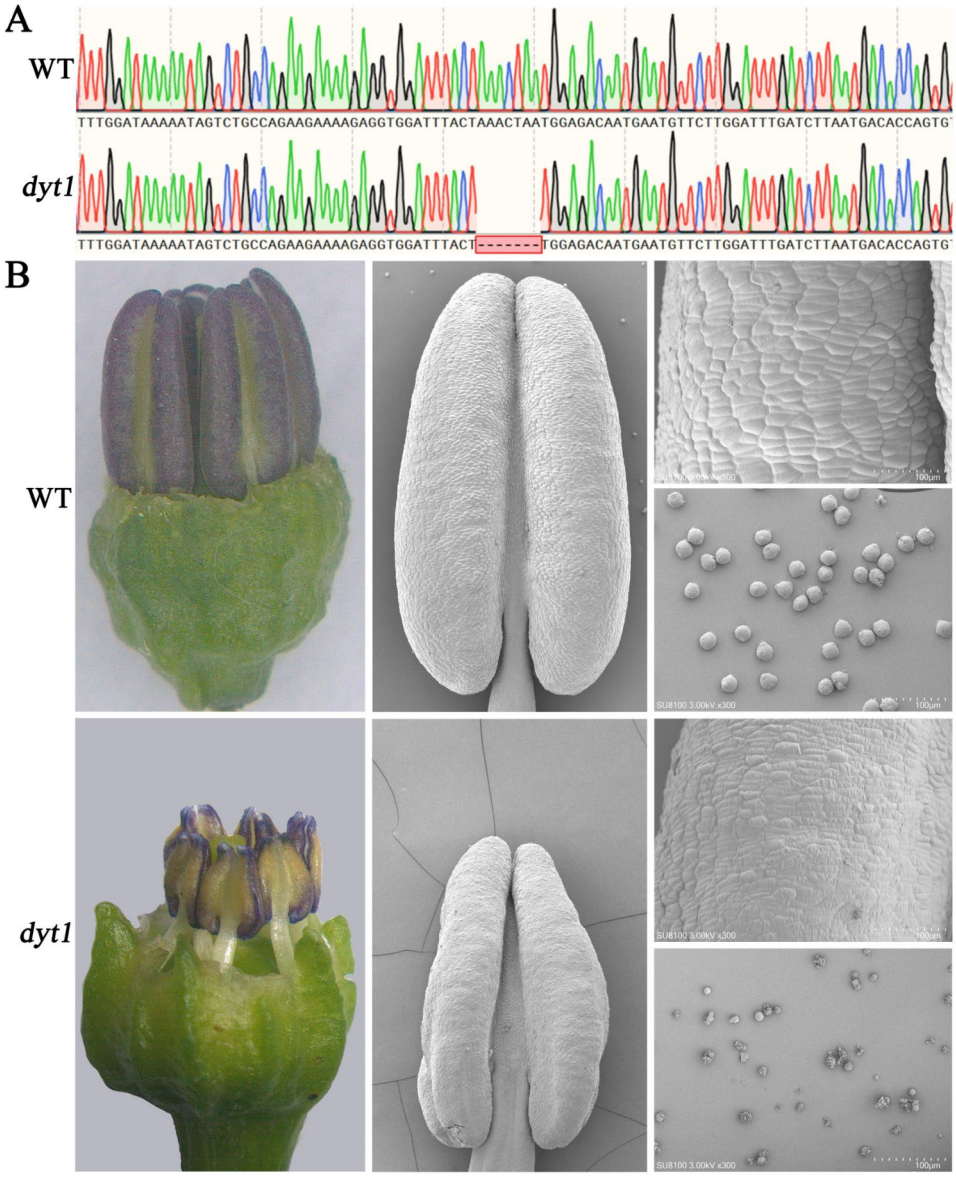

Fig. 7 Phenotypic analysis of the wild type (WT) and *dyt1* mutant. (**A**) Mutation site detection between WT and *dyt1*. (**B**) The phenotype of anther and pollen grain in WT and *dyt1*. Anther phenotype is shown on the left, and SEM analysis of anther and pollen grain is shown on the right

Methods

Retrieval and identification of *WD40* genes in pepper

In this study, the candidate WD40 proteins were retrieved as follows: first, the protein, nucleotide, and genome sequences of the Zunla, CM334, and Zhangshugang genomes were downloaded from the Sol Genomics Network (https://solgenomics.net/organism/Capsicum_annuum/genome) [56, 57] and Pepper Genomics Database (http://ted.bti.cornell.edu/cgi-bin/pepper/index) [31], respectively. Second, the HMM of WD40 protein (PF00400) was downloaded from the Pfam database (http://pfam-legacy.xfam.org/) [58]. Furthermore, the HMMER (v3.0) software [59] was used to search the predicted WD40 proteins using the cut-off value of the default parameter. Finally, the WD40 domain of all potential candidates was examined in InterPro (https://www.ebi.ac.uk/) [60] and SMART database (http://smart.embl-heidelberg.de/smart/batch.pl) [61], and those without the complete WD40 domain were manually deleted.

Sequence analysis and structural characteristics

Protein length, molecular weight, and theoretical pIs of CaWD40 proteins were analyzed using the Expasy tool (https://web.expasy.org/compute_pi/) [62]. The supposed subcellular localization of CaWD40 proteins was predicted using the online tool WOLF PSORT (https://wolfpsort.hgc.jp) [63]. The protein sequences were submitted

to the MEME program (https://meme-suite.org/meme/tools/meme) [64] to assess conserved motifs.

Chromosome localization, tandem duplication, and synteny analysis

Chromosome locations and gene position in pepper were obtained by searching the Sol Genomics Network. Chromosome mapping of the CaWD40 gene family was visualized using MG2C (http://mg2c.iask.in/mg2c_v2.1) [65]. Tandem duplication events were further confirmed using the following criteria: (1) the alignment length had a coverage rate of more than 70% of the full length of the CaWD40 genes; (2) the identity of the aligned region was over 70%, (3) and an array of two or more genes was less than 100 kb distance. The genome sequence was downloaded and annotation files of three different pepper species from NCBI (https://www.ncbi.nlm.nih.gov/assembly/?term=pepper) were obtained. TBtools-II [66] was used to analyze the synteny of *CaWD40* genes among the three pepper genomes.

Phylogenetic analysis

The phylogenetic tree was generated in the following three steps: first, the CaWD40 protein sequences were imported into Clustal X to produce a multiple sequence alignment file. Second, the alignment result was used to build an unrooted tree using MEGA11 with a bootstrap of 1000 replicates and neighbor-joining (NJ) methods [67]. Third, the newly produced phylogenetic tree was visualized using the Interactive Tree of Life online website (https://itol.embl.de/) [68].

RNA-Seq analysis of *CaWD40* genes

Transcriptome sequencing (RNA-seq) data of development was used to explore the distribution of gene expression in pepper (the elite Capsicum line 6421) [69] to gain insight into the expression profiles of the *CaWD40* gene family in different tissues across periods. Expression levels were determined in the following tissues and stages: leaf tissues were sampled at 2, 5, 10, 15, 20, 25, 30, 40, 50 and 60 days after emergence, and marked correspondingly as L1, L2, L3, L4, and L5, L6, and L7, L8, L9, and AL; floral buds were sampled at 0.25, 0.35, 0.5, 0.8, 1.0, 1.2, and 1.7 cm, and marked correspondingly as F1, F2, F3, F4, and F5, F6, and F7, F8, and F9; petals, stamens, and ovaries with stigmas were sampled in fully blossomed flowers, and marked correspondingly as P10, STA10 and O10; fruits were collected at 3, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, and 60 days after flowering (DAF), and marked correspondingly as FST0, FST1, G1, G2, G3, G4, G5, G6, G7, G8, G9, G10, G11; seed samples were collected at 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 and 60 DAF, and marked correspondingly as ST1, ST2, S3, S4, S5, S6, S7, S8, S9, S10, S11; placenta samples were collected at 20, 25, 30, 35, 40, 45, 50, 55 and 60 DAF, and marked correspondingly as T3, T4, T5, T6, T7, T8, T9, T10, and T11. Stems and roots were marked AS and AR, respectively. The treatment methods of all samples were based on those published by Liu et al. [69]. All data of *CaWD40* genes were normalized (log2(FPKM+1)), and a heatmap was drawn using TBtools-II (v2.019) [66].

RNA extraction and RT-qPCR analysis

Capsicum line 6421 was grown at 27 ℃/22 ℃ day/night cycles under a 16/8 h (light/dark) photoperiod. Total RNA was extracted using the TransZol Up (TransGen, China) and reverse-transcribed using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, China). RT-qPCR sample contained 2 × ChamQ Universal SYBR qPCR Master Mix (Vazyme, China) and was performed on a LightCycle® 96 Real-Time PCR System (Roche, Switzerland).

The *UBI-3* gene was used as a reference gene [70]. Gene-specific primers were designed. These are listed in Table S5. Three biological repeats and three technical repeats of RT-qPCR were set, and the relative expression of each gene was calculated using the $2^{-\Delta\Delta Ct}$ method.

Y2H assays

Using specific primer pairs, the coding sequences of CaWD40-91, CaAN1, CaDYT1, and CaGL3 were amplified and inserted into pGBKT7 and pGADT7 vectors, respectively (Table S5). Four plasmids, pGBKT7-CaWD40-91, pGADT7-CaAN1, pGADT7-CaDYT1, and pGADT7-CaGL3, were extracted and co-transformed into Y2HGold receptor cells (WEIDI, China). All transformation products were plated and grown on SD/-Trp/-Leu medium for three days. Protein-protein interactions were detected on the SD/-Trp/-Leu/-His/-Ade selective medium following the instructions in the Matchmaker® Gold Yeast Two-Hybrid System User Manual.

Dual-luciferase reporter assays

The primers for Nluc-CaWD40-91, Cluc-CaAN1, and Cluc-CaDYT1 were designed (Table S5). CaWD40-91 was cloned and inserted into the pCambia1300-Nluc. CaAN1, and CaDYT1 were cloned and inserted into pCambia1300-Cluc linearized by KpnI and SalI. These recombinant plasmids were transferred into *Agrobacterium tumefaciens* GV3101. The plasmids were delivered into tobacco leaves, and a fluorescence signal was captured three days later using the Vilber Fusion FX7 Spectra device system (Vilber Bio Imaging, Paris, France). Nluc-EV and Cluc-EV were used as controls.

Subcellular localization

The full-length ORF sequences of CaWD40-91 without the termination codon were cloned into the

pCAMBIA1300-GFP vector and transformed into *Agrobacterium tumefaciens* GV3101. *Arabidopsis* histone H2B-mCherry was used as a nuclear marker. CaWD40-91 fusion constructs and H2B marker were co-transformed into tobacco (*Nicotiana benthamiana*) leaves in a ratio of 1:1. After three days, the fluorescence signals were observed and captured using a confocal laser scanning microscope (Zeiss LSM510 META, Germany).

SEM analysis

Floral buds were sampled and fixed with a fixative solution. Samples were washed with 0.1 M PB (pH 7.4) thrice for 15 min each. The samples were transferred into 1% OsO_4 in 0.1 M PB (pH 7.4) for 1–2 h at room temperature. Subsequently, samples were washed in 0.1 M PB (pH 7.4) thrice for 15 min each. Then, they were transferred to 30%, 50%, 70%, 80%, 90%, 95%, 100%, and 100% ethanol for 15 min each. Finally, the samples were incubated in isoamyl acetate for 15 min and dried using a Critical Point Dryer. The dried samples were attached to metallic stubs using carbon stickers and sputter-coated with gold for 30 s. Samples were observed and photographs were captured with a scanning electron microscope (SU8100, Hitachi, Japan).

Abbreviations

RT-qPCR	Reverse transcription-Quantitative real-time PCR
HMM	Hidden Markov Model
DAF	Days after flowering
SMs	Secondary metabolites
MWs	Molecular weights
pIs	Isoelectric point
NJ	Neighbor-Joining
Y2H	Yeast two-hybrid
SMART	Simple Modular Architecture Research Tool
MEME	Multiple Em for Motif Elicitation
LUC	Luciferase
SEM	Scanning electron microscopy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10681-9.

Supplementary Material 1

Supplementary Material 2

Author contributions

F.L. & L.P gave the general idea of the paper. J.W & M.W were responsible for processing data and drawing figures. P.T & Q.H finished the experiment. P.T write the paper and J. H revised the context. All authors read and approved the final manuscript.

Funding

This research was supported by grants from the Science and Technology Innovation Program of Hunan Province, China (2021NK1006 and 2021JC0007).

Data availability

All methods using plant material were carried out in accordance with relevant guidelines and regulations in this paper. The data used and/or analyzed during the current study are obtained from the Sol Genomics Network (www.solgenomics.net/organism/Capsicum_annuum/genome) (Accession numbers: GCA_000512255.1, GCA_000710875.1); pepper Genomics Database (www.ted.bti.cornell.edu/cgi-bin/pepper/index) (Accession numbers: GCA_030867735.1).

Declarations

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 13 April 2024 / Accepted: 1 August 2024
Published online: 11 September 2024

References

1. Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. The ancient regulatory-protein family of WD-repeat proteins. Nature. 1994;371:297–300.
2. Smith TF, Gaitatzes C, Saxena K, Neer EJ. The WD repeat: a common architecture for diverse functions. Trends Biochem Sci. 1999;24:181–5.
3. Nocker SV, Ludwig P. The WD-repeat protein superfamily in *Arabidopsis*: conservation and divergence in structure and function. BMC Genomics. 2003;4:50–61.
4. Mishra AK, Puranik S, Prasad M. Structure and regulatory networks of WD40 protein in plants. J Plant Biochem Biotechnol. 2012;21:32–9.
5. Sondek J, Lambright DG, Hamm HE, Sigler PB, Bohm A. Crystal structure of a G-protein beta gamma dimer at 2.1A resolution. Nature. 1996;379:369–74.
6. Xu C, Min J. Structure and function of WD40 domain proteins. Protein Cell. 2011;2:202–14.
7. Haar ET, Musacchio A, Harrison SC, Kirchhausen T. Atomic structure of clathrin: a β propeller terminal domain joins an alpha zigzag linker. Cell. 1998;95:563–73.
8. Chen M, Zhang B, Li C, Kulaveerasingam H, Chew FT, Yu H. TRANSPARENT TESTA GLABRA1 regulates the Accumulation of seed storage reserves in *Arabidopsis*. Plant Physiol. 2015;169:391–402.
9. Jain BP, Pandey S. WD40 repeat proteins: signalling Scaffold with diverse functions. Protein J. 2018;37:391–406.
10. Song JJ, Garlick JD, Kingston RE. Structural basis of histone H4 recognition by p55. Genes Dev. 2008;10:1313–8.
11. Higa LA, Zhang H. Stealing the spotlight: CUL4-DDB1 ubiquitin ligase docks WD40-repeat proteins to destroy. Cell Div. 2007;2:5.
12. Tan L, Salih H, Htet N, Azeem F, Zhan R. Genomic analysis of WD40 protein family in the mango reveals a TTG1 protein enhances root growth and abiotic tolerance in *Arabidopsis*. Sci Rep. 2021;11:2266.
13. 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;53:814–27.
14. Zhao M, Morohashi K, Hatlestad G, Grotewold E, Lloyd A. The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. Development. 2008;135:1991–9.
15. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis* thaliana. Plant J. 2004;39:366–80.
16. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015;20:176–85.
17. Wang J, Dai Y, Pan L, Chen Y, Dai L, Ma Y, Zhou X, Miao W, Hamid MR, Zou X, Liu F, Xiong C. Fine mapping and identification of CaTTG1, a candidate gene that regulates the hypocotyl anthocyanin accumulation in Capsicum annuum L. Hortic Plant J. 2023.
18. Cai J, Huang H, Xu X, Zhu G. An Arabidopsis WD40 repeat-containing protein XIW1 promotes salt inhibition of seed germination. Plant Signal Behav. 2020;15:1712542.
19. Yan J, Ma Z, Xu X, Guo AY. Evolution, functional divergence and conserved exon-intron structure of bHLH/PAS gene family. Mol Genet Genomics. 2014;289:25–36.
20. Bjerkan KN, Jung-Roméo S, Jürgens G, Grini GPE, Arabidopsis. WD REPEAT DOMAIN55 interacts with DNA DAMAGED BINDING PROTEIN1 and is required for apical patterning in the embryo. Plant Cell. 2012;24:1013–33.

21. Shi DQ, Liu J, Xiang YH, Ye D, Sundaresan V, Yang WC. SLOW WALKER1, essential for gametogenesis in *Arabidopsis*, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis. Plant Cell. 2005;17:2340–54.
22. Grigorieva G, Shestakov S. Transformation in the cyanobacterium Synechocystis sp.6803. FEMS Microbiol Lett. 1982;13:367–70.
23. Janda L, Tichý P, Spízek J, Petrícek M. A deduced thermomonospora curvata protein containing serine/threonine protein kinase and WD-repeat domains. J Bacteriol. 1996;178:1487–9.
24. Zou XD, Hu XJ, Ma J, Li T, Ye ZQ, Wu YD. Genome-wide analysis of WD40 protein family in human. Sci Rep. 2016;6:39262.
25. Hu R, Xiao J, Gu T, Yu X, Zhang Y, Chang J, Yang G, He G. Correction to: genome-wide identification and analysis of WD40 proteins in wheat (*Triticum aestivum* L). BMC Genomics. 2018;19:852–65.
26. Ouyang Y, Huang X, Lu Z, Yao J. Genomic survey, expression profile and co-expression network analysis of OsWD40 family in rice. BMC Genomics. 2012;13:100–13.
27. Li Q, Zhao P, Li J, Zhang C, Wang L, Ren Z. Genome-wide analysis of the WD-repeat protein family in cucumber and *Arabidopsis*. Mol Genet Genomics. 2014;289:103–24.
28. Feng R, Zhang C, Ma R, Cai Z, Lin Y, Yu M. Identification and characterization of WD40 superfamily genes in peach. Gene. 2019;710:291–306.
29. Chen L, Cui Y, Yao Y, An L, Bai Y, Li X, Yao X, Wu K. Genome-wide identification of WD40 transcription factors and their regulation of the MYB-bHLH-WD40 (MBW) complex related to anthocyanin synthesis in Qingke (Hordeum vulgare L. var. Nudum hook. F). BMC Genomics. 2023;24:166.
30. Carolina CG, Barfuss MHJ, Sehr EM, Barboza GE, Rosabelle S, Moscone EA, Friedrich E. Phylogenetic relationships, diversification and expansion of Chili peppers (Capsicum, Solanaceae). Ann Botany. 2016;118:35–51.
31. Liu F, Zhao J, Sun H, Xiong C, Sun X, Wang X, Wang Z, Jarret R, Wang J, Tang B, Xu H, Hu B, Suo H, Yang B, Ou L, Li X, Zhou S, Yang S, Liu Z, Yuan F, Pei Z, Ma Y, Dai X, Wu S, Fei Z, Zou X. Genomes of cultivated and wild Capsicum species provide insights into pepper domestication and population differentiation. Nat Commun. 2023;14:5487.
32. Zhang Z, Liu Y, Yuan Q, Xiong C, Xu H, Hu B, Suo H, Yang S, Hou X, Yuan F, et al. The bHLH1-DTX35/*DFR* module regulates pollen fertility by promoting flavonoid biosynthesis in *Capsicum annuum* L. Hortic Res. 2022;9:3458–69.
33. Byun J, Kim TG, Lee JH, Li N, Jung S, Kang BC. Identification of CaAN3 as a fruit-specific regulator of anthocyanin biosynthesis in pepper (Capsicum annuum). Theor Appl Genet. 2022;135:2197–211.
34. Liu J, Ai X, Wang Y, Lu Q, Li T, Wu L, Sun L, Shen H. Fine mapping of the Ca3GT gene controlling anthocyanin biosynthesis in mature unripe fruit of Capsicum annuum L. Theor Appl Genet. 2020;133:2729–42.
35. Chen R, Yang C, Gao H, Shi C, Zhang Z, Lu G, Shen X, Tang Y, Li F, Lu Y, Ouyang B. Induced mutation in ELONGATED HYPOCOTYL5 abolishes anthocyanin accumulation in the hypocotyl of pepper. Theor Appl Genet. 2022;135:3455–68.
36. Kim YJ, Kim MH, Hong WJ, Moon S, Kim EJ, Silva J, Lee J, Lee S, Kim ST, Park SK, et al. GORI, encoding the WD40 domain protein, is required for pollen tube germination and elongation in rice. Plant J. 2021;105:1645–64.
37. Tian G, Wang S, Wu J, et al. Allelic variation of TaWD40-4B.1 contributes to drought tolerance by modulating catalase activity in wheat. Nat Commun. 2023;1200:36901–6.
38. Liu H, Xiu ZH, Yang HH, Ma ZX, Yang DL, Wang HQ, Tan BC. Maize Shrek1 encodes a WD40 protein that regulates pre-rRNA processing in ribosome biogenesis. Plant Cell. 2022;34:4028–44.
39. Airoldi CA, Hearn TJ, Brockington SF, et al. TTG1 proteins regulate circadian activity as well as epidermal cell fate and pigmentation. Nat Plants. 2019;5:1145–53.
40. Wang Z, Yang Z, Li F. Updates on molecular mechanisms in the development of branched trichome in Arabidopsis and nonbranched in cotton. Plant Biotechnol J. 2019;17:1706–22.
41. Yan C, Yang T, Wang B, Yang H, Wang J, Yu Q. Genome-wide identification of the WD40 Gene Family in Tomato (Solanum lycopersicum L). Genes (Basel). 2023;14:1273.
42. Schapira M, Tyers M, Torrent M, Arrowsmith CH. WD40 repeat domain proteins: a novel target class? Nat Rev Drug Discov. 2017;16:773–86.
43. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in *Arabidopsis* thaliana. Curr Opin Plant Biol. 2001;4:447–56.
44. Li X, Xue C, Li J, Qiao X, Li L, Yu L, Huang Y, Wu J. Genome-wide identification, evolution and functional divergence of MYB Transcription Factors in Chinese White Pear (Pyrus Bretschneideri). Plant Cell Physiol. 2016;57:824–47.
45. Liu J, Wang X, Chen Y, Liu Y, Wu Y, Ren S, Li L. Identification, evolution and expression analysis of WRKY gene family in Eucommia ulmoides. Genomics. 2021;113:3294–309.
46. 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.
47. Gou JY, Felippes FF, Liu CJ, Weigel D, Wang JW. Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. Plant Cell. 2011;23:1512–22.
48. Xue JS, Qiu S, Jia XL, Shen SY, Shen CW, Wang S, Xu P, Tong Q, Lou YX, Yang NY, Cao JG, Hu JF, Shen H, Zhu RL, Murray JD, Chen WS, Yang ZN. Stepwise changes in flavonoids in spores/pollen contributed to terrestrial adaptation of plants. Plant Physiol. 2023;193:627–42.
49. Buer CS, Imin N, Djordjevic MA. Flavonoids: new roles for old molecules. J Integr Plant Biol. 2010;52:98–111.
50. Lepiniec LC, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 2006;57:405–30.
51. Grunewald S, Marillonnet S, Hause G, Haferkamp I, Neuhaus HE, Veß A, Hollemann T, Vogt T. The tapetal major facilitator NPF2.8 is required for accumulation of flavonol glycosides on the pollen surface in *Arabidopsis* thaliana. Plant Cell. 2020;32:1727–48.
52. Muhlemann JK, Younts TLB, Muday GK. Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. Proc Natl Acad Sci U S A. 2018;115:E11188–97.
53. Rutley N, Miller G, Wang F, Harper JF, Miller G, Lieberman-Lazarovich M. Enhanced reproductive thermotolerance of the tomato high pigment 2 mutant is associated with increased accumulation of flavonols in pollen. Front Plant Sci. 2021;12:672368.
54. Cui J, You C, Zhu E, Huang Q, Ma H, Chang F. Feedback regulation of DYT1 by interactions with downstream bHLH factors promotes DYT1 Nuclear localization and Anther Development. Plant Cell. 2016;28:1078–93.
55. Cheng Q, Wang P, Liu, Jinqiu, Wu L, Zhang Z, Theoretical. Appl Genet Int J Breed Res Cell Genet. 2018;131:1861–72.
56. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, González-Morales S, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernández S, Leyva-González MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Volume 111. Proc Natl Acad Sci U S A; 2014. pp. 5135–40.
57. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JK, Sørensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet. 2014;46:270–8.
58. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The pfam protein families database in 2019. Nucleic Acids Res. 2019;47:D427–32.
59. Wheeler TJ, Eddy SR. Nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29:2487–9.
60. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A. InterPro in 2022. Nucleic Acids Res. 2023;51:D418–27.
61. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46:D493–6.
62. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40:W597–603.
63. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35:585–7.

64. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28–36.
65. Chao JT, Kong YZ, Wang Q, Sun YH, Gong DP, Lv J, et al. MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. Yi Chuan. 2015;37:91–7.
66. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative Toolkit developed for interactive analyses of big Biological Data. Mol Plant. 2020;13:1194–202.
67. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38:3022–7.
68. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
69. Liu F, Yu H, Deng Y, Zheng J, Liu M, Ou L, Yang B, Dai X, Ma Y, Feng S, He S, Li X, Zhang Z, Chen W, Zhou S, Chen R, Liu M, Yang S, Wei R, Li H, Li F, Ouyang B, Zou X. PepperHub, an Informatics Hub for the Chili Pepper Research Community. Mol Plant. 2017;10:1129–32.
70. Wan H, Yuan W, Ruan M, Ye Q, Wang R, Li Z, Zhou G, Yao Z, Zhao J, Liu S. Identification of reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*Capsicum annuum* L). Biochem Biophys Res Commun. 2011;416:24–30.

Publisher's Note