

MYB Transcription Factor Family in Sweet Cherry (*Prunus avium* L.): Genome-wide Investigation, Evolution, Structure, Characterization and Expression Patterns

Irfan Ali Sabir

Shanghai Jiao Tong University

Muhammad Aamir Manzoor

Anhui Agricultural University

Iftikhar Hussain Shah

Shanghai Jiao Tong University

Xunju Liu

Shanghai Jiao Tong University

Muhammad Salman Zahid

Shanghai Jiao Tong University

Songtao Jiu

Shanghai Jiao Tong University

Jiyuan Wang

Shanghai Jiao Tong University

Muhammad Abdullah

Shanghai Jiao Tong University

Caixi Zhang (✉ acaizh@sjtu.edu.cn)

Shanghai Jiao Tong University

Research Article

Keywords: MYB transcription factors, PavMYB genes, phylogenetic analysis, expression pattern, dormancy

Posted Date: August 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-783177/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on January 3rd, 2022.
See the published version at <https://doi.org/10.1186/s12870-021-03374-y>.

Abstract

Back ground

MYB Transcription factors (TFs) are most imperative and largest gene family in plants, which participate in development, metabolism, defense, differentiation and stress response. The MYB TFs has been studied in various plant species. However, comprehensive studies of MYB gene family in the sweet cherry (*Prunus avium L.*) are still unknown.

Results

In the current study, a total of 69 MYB genes were investigated from sweet cherry genome and classified into 28 subfamilies (C1-C28 based on phylogenetic and structural analysis). Microcollinearity analysis revealed that dispersed duplication (DSD) and whole-genome duplication (WGD) events might play an important role in the expansion of the MYB genes family. Chromosomal localization, the synonymous (Ks) and nonsynonymous (Ka) analysis, molecular characteristics (pI, weight and length of amino acids) and subcellular localization were accomplished using several bioinformatics tools. Moreover, different subfamilies contained different *cis*-acting elements, conserved motifs analysis and introns-exons structures, representing functional divergence in the MYB family. Subsequently, the transcriptomic data exposed that MYB genes might play vital role in bud dormancy. Subsequently, the quantitative real-time qRT-PCR was carried out and the expression pattern indicated that MYB genes significantly expressed in floral bud as compared to flower and fruit.

Conclusion

Our comprehensive findings provide supportive insights into the evolutions, expansion complexity and functionality of *PavMYB* genes. These *PavMYB* genes should be further investigated as they seem to be brilliant candidates for dormancy manipulation in sweet cherry.

Background:

The transcription factors (TFs) (sequence-specific DNA-bindings factor) are the proteins which regulates the rate of transcription of genetic information from DNA to messenger RNA through linkage with a specific DNA sequence. TFs control genes by turning them on and off to ensure that they are expressed in the right cell at the correct time and in the specific amount during the cell's and organism's lives [1, 2]. Transcription factors (TFs) are crucial regulators of gene transcription along with at a DNA-binding domain, nuclear localization signal transactivation domain and oligomerization site.

One of the largest TFs families in the plant kingdom has been classified as the MYB gene family. [3]. MYB proteins characterize the major transcription factor families in the plant kingdom. V-MYB was identified as first MYB TF in the avian myeloblastosis virus [4]. c-MYB-like TF is was the first plant MYB TF which was discovered in *Zea mays* and also justified for anthocyanin biosynthesis, [5]. Three types

MYBTFs (A-MYB, B-MYB and c-MYB) were later discovered in many slime molds, fungi, insects and vertebrates[6].

MYB proteins are distinguished by a strongly conserved MYB DNA-binding domain, which is mainly found at the protein's N-terminus. The MYB domains normally consists of up to four imperfect amino acid sequence repeats (R) of about 50–53 amino acids and forming three alpha–helices. The 2nd and 3rd helices of individual repeat build a helix–turn–helix (HTH) structure with three frequently spaced tryptophan residues, which creates a hydrophobic core [7]. The 3rd helix of each repeat was recognized as the DNA recognition helix that approach directly with DNA [8]. Two MYB repeats are densely compacted in the central groove during DNA interaction and allowing the two recognition helices to attach to the unique DNA recognition sequence motif cooperatively. Based on presence of MYB repeats from 1 to 4 in sequence, MYB is categorized in four groups 1RMYB-, R2R3MYB, 3RMYB- and 4RMYB [9–11]. Gene family identification is now possible at genomic level due to presence of gene sequences. The previous studies indicate that more than 80 R2R3MYB TFs genes are identified in maize genome [12, 13]). The *Oryza sativa* genome contains following number of genes in each category such as, 1R-MYB (62 genes), R2R3-MYB (88 genes) and 3R-MYB (4 genes) [14], while in *Arabidopsis thaliana*, 1R-MYBs (52 genes), R2R3-MYBs (135 genes) and 3R-MYBs (5 genes) were found [15]. MYB proteins are involved in secondary metabolism [16], hormone signal transduction, and response to environmental stress [17], cell differentiation and cell cycle [18]. The functions of MYB proteins are comprehensively studies in wide range of plants such as, *Malus domestica* [19], *Fragaria vesca* [20], *Arabidopsis thaliana* [21], *Pyrus bretschneideri* [22] and *Gossypium raimondii* [23]. In *Dendrobium hybrida*, *DhMYB1* played vital role in the control of floral epidermal cell shape [24]. *SmMYB44* gene may improve bacterial wilt resistance in the eggplant [25]. *MdMYB124* and *MdMYB88* control the cold hardiness, and also improve drought tolerance by controlling cell walls and root vessels in apple [26, 27].

Dormancy (temporary suspension of growth) is a controlled process that regulates plant growth/ and development [28]. Paradormancy, endodormancy, and ecodormancy are three different kinds of dormancy processes. Paradormancy occurs when signals from other plant organs prevent meristem growth (for example, in buds or seeds). The shoot apex, for example, may prevent axillary buds from growing out by exerting apical dominance; but, when the apex is removed, this condition of paradormancy is disrupted, and axillary buds grow out. Endodormancy is a phase in which signals inside the meristem suppress meristem growth. In plants, vegetative buds usually become endodormant in the fall and early winter to overcome the cold stress, and extended phases of chilling (i.e., temperatures slightly above freezing) are needed before growth can resume, even under favorable environmental conditions. Plants remained in ecodormant even after endodormancy has been released due to severe environmental conditions such as cold or drought, which enable cell division and cell elongation [28]. Environmental stress responsive (Short days (SD), cold) and hormonal responsive (ethylene, gibberellin (GA), and abscisic acid (ABA)) are the indicators that have a direct role in growth limitation and bud formation [29, 30]. SD and low night temperatures in the autumn trigger growth suspension and promote vegetative bud set or shoot-tip abscission which is critical initial step of cold adaptation in many trees and other perennial plants [31, 32]. Cold temperatures may cause growth stoppage and induce

endodormancy in certain species like pear, apple etc. [29, 33]. Previous studies proved that MYB genes are involved in dormancy regulation like *comp100540_c0_seq1*, *comp76266_c0_seq2*, *comp62057_c0_seq1* expressed their peak expression in ecodormancy and endodormancy in *Camellia sinensis* [34]. Same phenomenon was also confirmed in *Populus*. During short-day induction or dormancy transition, MYB62 and MYB4 exhibited distinct expression patterns [31, 35] while MYB96 is involved in seed dormancy regulation in *Arabidopsis* [36].

Dormancy is extensively influenced by external temperature, while bud break and flowering time are seriously affected by global warming. In the northern hemisphere, bud break and blooming dates are delayed in apple, cherry, birch and oak [37], while inadequate cold accumulation all through winter may result to inadequate dormancy release. These phenological variations have a direct effect on fruit crop production, potentially resulting in significant economic losses [38]. As a result, in order to combat fruit losses and anticipate future production changes, a better knowledge of bud responses to temperature stimuli in the context of climate change becomes critical. In recent years, numerous studies have used RNA sequencing technologies to evaluate the physiological and molecular processes of bud dormancy transitions in perennials. Sweet cherry (*Prunus avium* L.) is a perennial plant that is very temperature sensitive [33].

To explore the MYB effect on dormancy in sweet cherry, a genome-wide analysis of the MYB family in sweet cherry (*Prunus avium* L.) was carried out in this study, which included database searches using the *PavMYB* gene model and phylogenetic relationships, gene structures, chromosomal positions, syntenic analysis, gene duplication events (TD, PD, TRD, WGD and DSD) and other structural characteristics. In addition, insilico analysis and qRT-PCR expression patterns indicated that *PavMYBs* member may play a vital role in bud dormancy. Furthermore, identifying and evaluating the *PavMYB* genes involved in dormancy induction will assist to understand the dormancy processes which ultimately help to enhance the fruit production.

Material Method:

Plant materials

The sweet cherry cultivar 'Royal Lee' was grown at Shanghai Jiao Tong University's experimental farm in Shanghai, China (31.25°N, 121.48°E). Gisela 6 (G6) was used as rootstock for grafting diploid cultivars. Under the same agricultural practices, all trees were planted at a 5-6 m spacing. Until further usage, all experimental materials were frozen in liquid nitrogen and kept at -80°C.

Identification, database search and Characteristics of the *MYB* Genes in *P. avium*

The whole-genome sequence of *P. avium* (v1.0) along with GFF3 (General feature format file) were downloaded from GDR database (<https://www.rosaceae.org/>) [39]. Additionally, Arabidopsis MYB amino acid sequences were retrieved from the Arabidopsis Information Resource (TAIR) (<http://www.arabidopsis.org/>) and BLASTed against *P. avium*. The MYB domain's HMM (Hidden Markov Model) profile was collected from the Pfam version 31.0 database [40]. Following that, BlastP (E-value = $1e^{-3}$) resemblance searches of MYB and variations of HMM builds (E-value = $1e^{-3}$) were being used to classify MYB in *P. avium*. The sequence was further analyzed by using Pfam (<http://pfam.xfam.org>) [40], InterPro (<http://www.ebi.ac.uk/interpro/sea>) [41], SMART (<http://smart.emblheidelberg.de>) [42] and PROSITE [43]. Finally, the sequences were eliminated which omitted the central or entire MYB_DNA-binding domain. The MYB protein sequence were further subjected to the WoLF PSORT \boxtimes (<http://www.genscript.com/wolf-psort.html>) and ExPasy software (<https://web.expasy.org/protparam/>) for analyzing the subcellular localization and molecular characteristics (length of amino acids, molecular weights and isoelectric points) correspondingly [44, 45].

Phylogenetic analysis and prediction of paralogous gene pairs

All MYBs subfamily members' amino acid sequences were matched using default parameters (pairwise deletion, 1000 bootstrap) in the multiple sequence alignment tool clustalX software. The phylogenetic tree was built with Molecular Evolutionary analysis (MEGA-X) by utilizing maximum likelihood method [46, 47]. Furthermore, the itols program (<http://itol.embl.de>) was used to visualize phylogenetic trees [48]. The resultant tree was used to determine the evolutionary origins of MYBs. The paralogous gene pairs (PGPs) were characterized from the final tree as proteins that were exist in pairs on the terminal nodes, well defined by their bootstrap values

Duplication of MYB genes, ka/ks analysis and collinearity relationships

The MCScanX programme was used to conduct interspecies and intra-species collinearity analyses at the protein with an E-value of $1e^{-5}$ [49]. The duplicate gene classifier script in the MCScanX program was used to quantify the various forms of duplications (tandem, dispersed, WGD or segmental and proximal duplication) [49]. When the pairs of genes inside two segmental regions have cointegration, WGD or segmental duplication gene pairs were assumed. We referred to gene pairs with two duplicated genes as tandem duplications when they were sequential. Collinearity relationships and gene duplications were envisioned by means of circos software and TBtools [50, 51]. Synonymous mutation rates (ks; mutations/substitutions that outcome in a single amino acid alteration on a specified polypeptide) and the non-synonymous (ka mutations/substitutions that would not lead in an amino acid sequence change) for subsequent duplication gene pairs were easily extracted from the Plant Genome Duplication Database (<http://chibba.agtec.uga.edu/duplication>) [52]. MAFFT software was used for alignments and

ka/ks calculators were used to compute the ka/ks ratio within each duplicated gene pair and multiple alignments (https://github.com/qiaoxin/Scripts_for_GB/tree/master/calculate_Ka_Ks_pipeline)[53].

Intron-exon, conserved motif and domain analysis

For the detection of conserved domains of each gene, we used the HMMER version 3.1 programme (<http://hmmer.org/>) and CDART (Conserved Domain Architecture Retrieval web Tool) (<https://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi>) to search against the Pfam database version 31.0 [54]. The Gene Structure and Display server v2.0 (GSDS; http://gsds.cbi.pku.edu.cn/Gsds_abou.php) was utilized to envision the next major features: location and arrangement of intron-exon, conserved elements and binding sites [55]. The conserved motifs were analyzed through the Motif Elicitation (MEME) software (<http://meme-suite.org>) [22]

Chromosomal localization and gene structure analysis

The starting and ending positions of each of the defined MYB genes were obtained from the Plant Transcription Factor Database and validated using the GFF3 file. CLC sequence viewer version 7 (CLC Bio, QIAGEN A/S, Aarhus, Denmark) was used to verify these locations along chromosomes of the downloaded rosaceae genome (<http://www.rosaceae.org/>). Finally, Mapinspect software (<http://mapinspect.software.com>) [45, 56, 57] was being used to analyze the data.

Transcriptomic analysis

The NCBI SRA database (<http://www.ncbi.nlm.nih.gov>) was used to obtain RNA-seq data of *P. avium* on different dormancy stages (organogenesis, paradormancy, endodormancy, ecodormancy and dormancy release) through following SRR8984342, SRR8984344, SRR8984360, SRR8984359, SRR8984367, SRR8984366, SRR8984382, SRR8984381, SRR8984403 and SRR8984402 accession numbers. Data from the SRA database was first encrypted into the FASTQ format using the SRA toolkit. With default parameters, the Hisat2 programme was used to map each dataset to the reference genome. The expression level was computed by using StringTie programme and standardized to TPM (Transcripts Per Kilobase Million) [58]. Heat maps were generated through TBtools.

RNA isolation and quantitative real time PCR

The expression patterns of *PavMYB's* gene was examined by using qRT-PCR. Total RNA was extracted from bud, flower and fruit by using RNAprep pure plant kit (Tiangen). The RNA (1 µg) was utilized to execute reverse transcription was done through the PrimeScript™ RT reagent Kit with gDNA Eraser (Takara). Specific primers was designed and checked through Beacon designer 7.9 software and Actin was used as reference gene (Table S2). The quantitative real-time qRT-PC was done through SYBR premix

ExTaq II: 10µl, template cDNA: 2µl, forward and reverse primer: 0.8µl and nuclease-free water to 20µl under the following conditions: 50°C for 2 min, 95°C for the 30s, followed by 40 cycles of 95°C for 15s, 60°C for 20s and 72°C for 20s 4°C for ∞. The $2^{-\Delta\Delta ct}$ method was utilized to compute the relative expression levels of *PavMYB* genes [59].

Results:

Identification and Molecular Characterization of MYB Gene Family in Sweet cherry

Genome-wide analysis was performed to identify the R2R3-MYB genes in *P. avium* genome. The MYB HMM profile (Pfam: 00249) was being utilized as a query in a BlastP search against the sweet cherry genome database. Finally, 69 *PavMYBs* proteins in sweet cherry were identified and their sequences were further examined for the presence of MYB_DNA-binding domains by using the Pfam, Interpro tool, and SMART database. Single MYB DNA-binding domain was identified in 51 *PavMYB* genes, 14 *PavMYB* genes contained two domain and two *PavMYB* genes had four domains. Only *PavMYB16* contained three MYB_DNA-binding (Table S1). In meanwhile, we discovered that *PavMYB* genes from the same subfamily having identical domain distributions and compositions elaborated that these subfamilies have had a similar evolutionary history. *PavMYB* proteins had amino acid lengths ranging from 63bp to 1056bp with average of 272bp (Table S1). Moreover, the molecular weight of *PavMYB* genes varied from 7421.6 to 117469.94 kDa with an average of 30452.16 kDa and the approximate Isoelectric point (IP) ranged from 4.09 to 10.44 with an average of 6.78 (Table S1).

Phylogenetic relationship of the Sweet cherry MYB Gene Family

To investigate the evolutionary relationship of the MYB genes in sweet cherry, a maximum likelihood method (ML-M) phylogenetic tree were constructed by using the full-length R2R3-MYB protein sequences from *P. avium* and *Arabidopsis thaliana*. We classified the R2R3-MYB members of sweet cherry into 28 subgroups (C1-C28) according to the sequence similarity and topology with Arabidopsis (Fig.1). The defined clades in Arabidopsis were labeled in the evolutionary tree. Subfamily C9 contained the highest MYB members (18), while the subfamily C3, C12, C14 and C24 had the lowest two MYB members (Fig.1 and Fig.2). MYB protein sequences from sweet cherry contributed in all subfamilies with Arabidopsis except C1 subfamily. Our results proposed that they may have been gene gained or lost events that happened through the evolutionary process. The loss and gain of particular MYB gene members instigated functional divergence [60, 61]. This phenomenon revealed that certain evolutionary events happened within MYBs of different species.

Conserved motifs, domain and intron number evaluation

A total of 69 MYB proteins from *Prunus avium* were checked against the Pfam database to validate that they contain the MYB DNA-binding domain. Each MYB gene's protein domains were determined (Fig.1 and Table S1). At least one MYB DNA-binding domain was found in all MYB genes that encode proteins, while 14 genes encoded proteins with two MYB DNA-binding domains such as *Pav_sc0000480.1_g870.1.mk*, *Pav_sc0000464.1_g250.1.br* and *Pav_sc0000143.1_g310.1.mk*. Two genes including *Pav_sc0000124.1_g210.1.br*, *Pav_sc0000872.1_g190.1.mk* contained four MYB DNA-binding domains while only *Pav_sc0000800.1_g1130.1.mk* had 3 MYB DNA-binding domains (Table S1 and Fig.1). We revealed that MYB genes from the same subfamily had equal domain distributions and compositions, implying that these subfamilies had a same evolutionary history. The position of introns was retained throughout evolution and might be implemented to determine a gene's phylogenetic relationship [62, 63]. We used intron structure analysis to figure out genesis of MYB gene family in sweet cherry. Most genes in the same cluster had common exon/intron structures, especially in terms of the number of introns, like I, II, III, and so on (Fig.6). Nevertheless, there were a few notable exceptions. The members of V and VI, for example, had varying numbers of introns. Intron/exon numbers ranged from 1-11/1-11 in sweet cherry (Fig.6 and S1).

Current findings suggested that the *PavMYB* subfamily has a lot of strongly conserved structures within subfamilies while a lot of sequence diversity between subfamilies was observed. Genomic and CDS sequences were evaluated in order to investigate the structural diversity of MYB genes. The majority of MYB genes were found in clade "I" which contained 1 to 2 introns, while two members were found in clade "II," which has only one intron (Fig.6). Furthermore, the number of introns varied from 1 to 11, the clade "VIII" having the most (1 to 11) accordingly (Fig.6). These findings suggested that intron loss or gain happened during the evolutionary process of the MYB genes family. The conserved motifs were identified by using MEME web server to evaluate the sequence characteristics of MYB members. Finally, each comparison revealed twenty motifs, which were classified into 1 to 20. The majority of MYB genes, as specified in Fig.5, included several motifs (1, 2, 3, 4, and 5). Clade II particularly, had only three motifs (1, 4 and 5), while clade III members contained 2 motifs (1, 4) but other clades had several (Fig.5). The motif compositions of most highly associated identical MYB genes indicating that MYB genes within the same subfamily have some functional similarities. Furthermore, we discovered a subfamily-specific motif that could play a key role in subfamily-specific activities. Moreover, several motifs, such as motifs 1 and 2, were identified in practically all subfamilies, suggesting that these motifs are crucial for MYB gene expression with similar functions. The evolutionary study of the MYB gene family was supported by the homogeneity in the compositions of motifs and intron structure of MYB genes within the same group, whereas differences between distinct groups indicated their special functionalities.

Cis-Acting Element Analysis

Two complementary regulatory components exist in the plant transcriptional mechanism: (1) trans-acting elements (2) *cis-acting* elements. Trans-acting factors are transcription factors as well as other DNA-binding proteins that attach to particular regions in *cis-acting* elements to boost or repress gene

transcription. The DNA sequence in the non-coding or coding regions of the genome are defined as *cis*-acting elements. *Cis*-regulatory elements play an important role in the management of regulatory networks, particularly multi-stimulus responsive genes and defining the stress-responsive expression patterns or tissue-specific expression patterns of genes was deeply linked with *cis*-elements in their promoter regions. The *cis*-acting elements on the promoter regions were classified into three major classes: phytohormones responses, biotic/abiotic stress responses and plant growth/development. Moreover, growth and development, *cis*-acting elements are prevalent in promoter regions, containing the Skn-1_motif, GCN4_motif, MRE, Box-4, CAT-box, O₂-site and circadian. We identified the CAT-box motifs who encompassed the highest proportion (33%), which are useful for meristem expression, followed by the O₂-site (30%) and Box-4 (23%) (Fig.3B) that are responsible for zein metabolic control and plant growth in response to light respectively [64]. The TGA-element is engaged in auxin sensitivity, GARE-motif and P-box for gibberellin response and ERE for ethylene responsive expression. The most prevalent *cis*-element in the second group was CGTCA motifs and a TGACG_motif cis acting elements associated to methyl jasmonic acid (MeJA) responsiveness [65-67]. The phytohormones response related cis-elements like ABRE (23%), P-box (6%), TGACG-motif (26%), TCA-element (15%), GARE motif (4%) and CGTCA-motif (26%) were also revealed (Fig.3B), which are linked with SA, ABA, ethylene and MeJA responses, signifying that these played key role in dormancy [68, 69]. Numerous stress-responsive elements were determined, such as ARE (44%), LTR (14%), MBS (28%), which are associated to light stress, cold and drought stress response respectively. These findings suggested that members of MYB gene family have the ability to improve cold stress response.

Gene duplication events, Syntenic Analysis and Expansion Patterns of the Sweet cherry MYB Genes

Expansion of Gene family and the development of novel functionalities are known to be aided by gene duplication (tandem and segmental) and divergence [67]. Duplicated genes generally mutate to acquire new functions or to divide ancestral gene functions which is essential for plant adaptation [70, 71]. An ancient whole-genome duplication (WGD) event occurred around 140 million years ago in apple [72]. We used MCScanX and circos software to visualize duplications within the genome of sweet cherry genome to evaluate the effects of duplication on the sweet cherry MYB genes family. Furthermore, *PavMYB* genes, we revealed 14 dispersed duplication pairs, 7 WGD duplication pair, 5 transposed duplication, 3 proximal duplication and two tandem duplication pairs (Fig.7 and Fig.2). This indicates that DSDs and WGDs play important role for MYB gene family expansion. These results revealed that duplications events contribute vital role in MYB gene expansion. Moreover, DSDs and WGD duplications happened with higher frequency for expansion.

To further understand the evolutionary relationship of MYB members in different plant species, the syntenic relationship was traced between *PavMYB*'s and homologous in other species including *P.mume*, *M. domestica*, *P. persica*, *P. bretschneideri* and *A. thaliana*. These four plants belong to the Rosaceae

family and shared a similar ancient. Consequently, the collinearity relationships were also analyzed between all MYB genes in four Rosaceae genome (Table S4). Total 359 collinear genes pair events were found between four Rosaceae species along with *A. thaliana*, included 57 orthologous pairs among sweet cherry and Japanese apricot, 106 orthologous pairs among sweet cherry and apple, 65 orthologous pairs among sweet cherry and peach, 91 orthologous pairs among sweet cherry and Chinese pear, and 40 orthologous pairs sweet cherry and *A. thaliana* (Fig.4) suggesting the closer relationship among four Rosaceae genomes (Fig.4b). These results recommend that there are collinearity connections between the sweet cherry and the other Rosaceae genomes, signifying a potential evolutionary process between them.

For better visualizing the evolutionary history, Ka/Ks-values was calculated to estimate evolutionary evidence of duplication occurrences in general (whole genome duplication or segmental duplication). Previous research has also been shown that the apple genome was formed by two phases of ancient whole-genome duplications around 140 million years ago (Ks 1.5–1.8) and a recent whole-genome duplication around 30–45 million years ago (Ks 0.15–0.35) [73]. The Ks values were determined to investigate whole or segmental genome duplication events with in MYB gene family (Table S5). The Ks values of duplicated gene pairs were determined. Ks average value was 0.2813 which revealed that MYB genes duplications might well have resulted from a recent whole-genome duplication. Furthermore, the Ka/Ks ratios are commonly utilized to reflect gene selection pressure and rate of evolution. Positive selection with faster evolution is indicated by $Ka/Ks > 1$, purifying selection with the functional constraint is indicated by $Ka/Ks < 1$, and $Ka/Ks = 1$ shows that the genes are migrating naturally [74]. In sweet cherry, the Ka/Ks ratios of duplicated gene pairs < 1 (*Pav_sc0000625.1_g100.1.mk*, *Pav_sc0000464.1_g250.1.br*, *Pav_sc0002360.1_g910.1.mk*, *Pav_sc0001051.1_g240.1.mk*) suggesting that MYB genes developed beneath severe purifying selection (Table S5). Moreover, in sweet cherry gene pairs like *Pav_sc0000103.1_g410.1.mk* (Ka/Ks ~3.37), *Pav_sc0001938.1_g500.1.mk* (Ka/Ks ~1.13), *Pav_sc0001181.1_g940.1.mk* (Ka/Ks ~1.85) *Pav_sc0000464.1_g130.1.br* (Ka/Ks ~1.28) and *Pav_sc0000069.1_g100.1.mk* (Ka/Ks ~1.63) had higher Ka/Ks ratios implying that this family might have a complex evolutionary history. For sweet cherry, the mean Ka/Ks calculated values for the TD, TRD, WGD, PD, and DSD gene pairs, which were 0.72, 1.05, 1.00, 0.70 and 1.06, correspondingly (Fig.6, Table S5). The DSD and PD duplication events had a higher Ka/Ks ratio as compared to other mode of duplications, representing that that MYB genes have expansion and complicated evolutionary history (Fig.6). Positive selection occurs in certain sections of protein-coding genes at the same time, implying the emergence of novel gene functions. Strongest evolutionary constraints were engaged in the evolution history of the MYB gene which may contribute to gene function stability.

Chromosomal distribution analysis

We construct a map of chromosomal locations based on the genomic information of sweet cherry to illustrate the dispersal of the MYB members throughout the chromosomes (Table. S3). The MYB genes were found in a random pattern on 8 chromosomes, with the majority of them clustered near the tail

intermediate end of a single chromosome. The highest MYB genes (15%) were found on Chr1, while, 14% were found on Chr2 and Chr3 which were organized into gene clusters (Fig. S1). Furthermore, chromosome 5 and 6 have 11.5% MYB genes, while chromosome 4 had 5% of total MYB genes. Moreover, the least MYB genes (2.8%) were found on chromosomes 8. Furthermore, on chromosomes 7, 5% MYB genes was traced and 13% genes were found on scaffold (Fig. S1).

Go annotation analysis

Go ontology annotation analysis was performed to predict subcellular localization and evaluated different functions of MYB protein in sweet cherry. 69 *PavMYB* proteins were categorized into 50 functional groups based on protein sequence similarities and categorized into four main ontologies, namely Biological process, Molecular functions, cellular component and subcellular localization. In the molecular process, we analyzed that more than 38.89% of annotated proteins functions in the activity of DNA binding, followed by nucleic acid binding (35.99%), protein-binding (11.59%) and ion binding (3.38%) (Fig.8). In the biological process, *PavMYB* members involved in the cellular nitrogen compound metabolic processes (11.66%) and biosynthetic processes (11.66%), stress response process (9.92%) followed by anatomical structure development(9.59%), signal transduction (9.36%), cell differentiation (8.98%) and reproduction (8.61%) (Fig.8). Subsequently, predicted *PavMYB* proteins are also annotated with cell cycle (4.99%), immune system process (4.51%), and cell morphogenesis (>1%) in biological process annotation. The cellular component annotations showed that the *PavMYB* proteins annotated with the intracellular, nucleus, organelles and having the same percentage (19.17%) and nucleoplasm (5.50%) (Fig.8). Furthermore, in subcellular localization, these results showed that most of the *PavMYB* proteins were contained in the nuclear, mitochondrial, extracellular, cytoplasmic and plasma membrane (81.10%, 10.80%, 4.10%, 2.70%, 1.40 respectively).

Transcriptomic analysis

RNA-seq datasets were used to assess the expression patterns of the MYB genes in different phases of dormancy under normal growth conditions. The amount of Transcripts Per Kilobase Million (TPM) were counted and compared across the investigated samples to obtain standardized RNA-seq data, which are visualized in form of heat map (Fig.9). Based on the current findings, MYB genes showed significant expression patterns on particular stage as shown in Fig.9. According to RNA-seq data MYB genes showed three types of expression level, each of which had genes that were predominantly expressed during distinct phases of bud dormancy (fig). The expression of nine MYB genes (*Pav_sc0001807.1_g120.1.mk*, *Pav_sc0000659.1_g090.1.mk*, *Pav_sc0001968.1_g010.1.br*, *Pav_sc0000678.1_g150.1.br*, *Pav_sc0000028.1_g520.1.mk*, *Pav_sc0000877.1_g610.1.mk*, *Pav_sc0001181.1_g940.1.mk*, *Pav_sc0000910.1_g810.1.mk* and *Pav_sc0000354.1_g490.1.mk*) were considerably higher across all stages of dormancy. The findings also revealed that MYB genes were seldom expressed in paradormancy than in endo dormancy (Fig.9). In all dormancy phases, majority of genes were showed highest

expression level, this phenomenon revealed that these genes play key role for dormancy regulation. During data analyzing it was also found that some genes (*Pav_sc0000009.1_g210.1.mk*, *Pav_sc0000464.1_g230.1.mk*, *Pav_sc0000308.1_g480.1.mk*, *Pav_sc0000638.1_g100.1.mk*, *Pav_sc0001576.1_g360.1.mk*, *Pav_sc0000886.1_g620.1.mk*, *Pav_sc0000110.1_g160.1.br*, *Pav_sc0000192.1_g040.1.mk*, *Pav_sc0000396.1_g530.1.mk*, *Pav_sc0001756.1_g010.1.mk*) were remain silent during all dormancy stages. This phenomenon demonstrated that these genes are not involved in the dormancy initiation or dormancy elimination. These group of genes might be involve some other important functions in later stages of floral development in sweet cherry. Some genes like *Pav_sc0000009.1_g270.1.mk* remain silent in whole dormancy phases but show the aggression at dormancy release while some other genes like *Pav_sc0000872.1_g190.1.mk* and *Pav_sc0000143.1_g310.1.mk* had indicated peak expression in ecodormancy phase. These results illustrated that *PavMYB* genes also had stage specific expression patterns too.

Expression analysis

Moreover, for further investigate the expressional behavior of *PavMYB*'s in dormancy and in other later developmental stages, we chose 28 genes (one gene from each clade) to explore the expression levels in bud, flower, and fruit. The qRT-PCR results validated the RNA-seq data, indicating that these genes had different expression profiles at the bud, flower and fruit. *Pav_sc0002433.1_g080.1.mk*, *Pav_sc0002360.1_g910.1.mk*, *Pav_sc0012310.1_g010.1.br*, *Pav_sc0001335.1_g540.1.mk*, *Pav_sc0001051.1_g240.1.mk*, *Pav_sc0004890.1_g010.1.mk*, *Pav_sc0000591.1_g010.1.mk*, *Pav_sc0001807.1_g120.1.mk*, *Pav_sc0000124.1_g210.1.br*, *Pav_sc0000254.1_g020.1.mk* and *Pav_sc0000311.1_g1150.1.mk* expressions were considerably upregulated at bud stage, revealing that these genes are engaged in the dormancy. *Pav_sc0001756.1_g010.1.mk*, *Pav_sc0000396.1_g530.1.mk*, and *Pav_sc0001576.1_g360.1.mk* expression levels were identical to RNA-Seq data, their expression being elevated only at flower and fruit. No expression was observed on bud stage. Results demonstrated that these genes (*Pav_sc0001756.1_g010.1.mk*, *Pav_sc0000396.1_g530.1.mk*, and *Pav_sc0001576.1_g360.1.mk*) were remained silent in bud dormancy and show their expression in further stages. Some genes like *Pav_sc0001938.1_g500.1.mk* and *Pav_sc0000800.1_g1130.1.mk* had identical expression profiles all dormancy stages while remaining other genes had variable expression levels.

Discussion:

Across all eukaryotes, MYB genes play a critical role in various functionalities. *Arabidopsis thaliana* [14], *Vitis vinifera* [75], *Oryza sativa* [14], *Eucalyptus grandis* [76], *Zea mays* [13] and *Setaria italic* [77] have all presented comprehensively and systematically study of MYB gene family. The sweet cherry (*P. avium* L.) Genome Sequence Project was accomplished in 2020 [78], hence the sweet cherry genome sequence might be valuable for genome-wide analysis of MYB gene family. *PavMYB* gene family was not extensively studied yet and its majority functions remain unclear so far. In this research, The *PavMYB* family was extensively studied with analysis of phylogeny, gene structures, promoter regions, gene

duplication events, sequence characteristics, chromosomal locations, GO annotations and expression patterns. Sweet cherry has fewer *PavMYB* genes (69) than, *Oryza sativa* (155) [14], *Zea mays* (132) [13], *Setaria italic* (209) [77], *Malus pumila Mill* (229) [79], while greater than *Ginkgo biloba* (45) [80]. These results illustrated that the *R2R3-MYB* gene family expansion occurred as compared to *Ginkgo biloba*. The result also demonstrated that R2R3-MYB genes replication happened in evolutionary history and played a vital role in the regulation of plant specific development [61].

Gene duplication are important events that play vital role in plant evolutionary history and divergence of new functions [81]. The expansion of gene families has been facilitated greatly by gene duplication events [82]. In this investigation, a majority of the *PavMYB* genes were identified to have dispersed and whole genome duplications. The 69 *PavMYB* had total of 31 duplication events, with the majorities involving dispersed duplication (14) and whole genome duplication (7) while few others like transposed (5), proximal (3) and tandem duplication (2). In *Ananas comosus* [83], *Arabidopsis thaliana* [82], *Citrus sinensis* [84] and *Solanum tuberosum* [85] similar patterns of MYB gene duplication pairs were identified. The studies suggest that, DSDs (dispersed duplications) and WGDs (whole-genome duplications) events are play an important role for the development of MYB genes family. Collinearity analysis reveals a striking similarity between syntenic orthologous groups and phylogenetic relationships. The K_a/K_s proportions of 69 pure *PavMYB* replications revealed that most of *PavMYB* gene family has experienced purifying selection, demonstrating that it has been involved in a highly conserved evolution. Moreover, positive selection has also occurred in some *PavMYBs*, which indicated that novel gene functions. Whole-genome duplication in Rosaceae species was happened about 5 billion-year-old which raised the chromosomes number from 9 to 17 [86], possibly resulting in the massive MYB family in sweet cherry. Furthermore, 13% *PavMYBs* could not be mapped to any chromosome, which may be related to poor quality of the sweet cherry genome sequence or excessive heterozygosity.

Subsequently, to understand the *PavMYB* genes' evolutionary relationship, a ML-M tree was constructed using the multiple alignment of Arabidopsis R2R3-MYB amino acid sequences. The R2R3-MYB proteins were divided into C1-C28 groups based on the phylogenetic tree with Arabidopsis. The *PavR2R3-MYB* genes grouped with Arabidopsis orthologs in several branches, suggesting that the R2R3-MYB function was extensively conserved across species. The phylogenetic connections of MYB proteins in sweet cherry and Arabidopsis also revealed that mostly clades had varied numbers of *AtMYB* and *PavMYB* genes, demonstrating that these two species had evolved in a similar evolution. In Arabidopsis, two 4R-MYB genes and five 3R-MYB genes have been discovered so far [87]. Nevertheless, in the phylogenetic tree, several *PavMYB* genes from sweet cherry were not grouped with the *AtMYBs* from *A. thaliana*, signaling that such genes are not retained between sweet cherry and *A. thaliana* and that their roles may be unique. Furthermore, in the subfamilies evaluated, *PavMYBs* and *AtMYBs* were not represented equally in evaluated subfamilies. For instance, there were no *AtMYBs* in subfamily C1. These results indicate that following divergence from the most recent primitive ancestor, species-specific MYBs gene were recruited in sweet cherry or eliminated in Arabidopsis lineages.

Subsequently, our study demonstrated that, most genes in the identical group have the similar position of the exon-intron structure, nevertheless there are fewer irregularities, which may be due to gain, loss, or skimming of introns during the formation of MYB gene family [88]. Identical findings have been observed in apple [79], suggesting that evolutionary processes of MYB members within the same family may also be similar. Several genes in plants might well be eliminated all through the evolution process [89]. In the current study, the genes *Pav_sc0000069.1_g120.1.mk*, *Pav_sc0000354.1_g490.1.mk*, *Pav_sc0000383.1_g280.1.mk*, *Pav_sc0000567.1_g060.1.mk* and *Pav_sc0001003.1_g300.1.mk* have only 1 intron (Fig. 6 and Table S1), indicating the intron loss in evolution process. The previous research illustrated that in plants during evolution, selection pressure is the source of intron gain or loss, and genes evolve themselves into several exon-intron structures to achieve typical functions [90]. The absence of introns in genes would accelerate the evolution of genes through gene copying [91]. The function of the genes without any intron and those with one intron were found to be identical in their evolution and functions. These results exhibited that these genes endure conserved under the evolution process and own highly functional similarities. Moreover, most of the *PavMYB* genes contained similar motif compositions in the identical subfamily. A total of 20 diverse motifs were identified among the MYB genes of the sweet cherry (Fig. 5), each genes containing at least two of them. The arrangement and amount of the 20 distinct motif types in sweet cherry expressed that MYB members were functionally diverse. The existence of an identical motif and intron-exon structure between genes suggested that they may have functional similarities.

The *cis*-regulatory elements analysis demonstrated that, almost all *PavR2R3-MYB* genes contained several *cis*-acting elements like ABRE TGACG-motif and CGTCA-motif, MBS and LTR etc. which are related to abiotic stresses like low temperature response, defense/ response and wound response. Previous studies illustrated that MYB regulate the ABA biosynthesis signaling which help the plants in dormancy phase under unfavorable lower temperature [92, 93]. The *MYB96*-regulate alterations in ABA levels might lead to change GA metabolic activities to confirm a dormancy state in Arabidopsis [36]. Total of 14% genes contained LTR elements and 28% genes contained MBS on the promoter regions. These results also indicated that the *Pav_sc0000124.1_g210.1.br* and *Pav_sc0000872.1_g190.1.mk* contained ABRE elements (Fig. 3B). The expression of these genes obviously differed under cold stress. A total of 26% TGACG-motif and CGTCA-motif, 4% GARE motifs and 23% ABREs were the main *cis*-elements of gene expression under responses to dormancy. Those results indicated the potential function of *PavR2R3-MYB* genes in response to dormancy in sweet cherry. In general, the results showed that most *PavR2R3-MYB* members were involved in various response to hormone and abiotic stress responsive.

Moreover, a unique DNA-binding domain may be found in the N-termini of MYB TFs (Dubos et al., 2010) and domain consists of one to four incomplete repetition units, each of which has 3 helices. The 2nd and 3rd helices (R2 and R3) normally form an HLH configuration for DNA binding. This study discovered a total of 69 *PavR2R3-MYB* proteins. The R2 repeats had three evolutionarily conserved Trp residues, but initial Trp (R3) residue was unpredictable. Substitution of the first Trp residue might result in the identification of new target genes and/or a decrease of DNA-binding ability towards target genes [94]. The N-terminal domain is rather stable, whereas the C-terminal region is more malleable [60, 89]. The

PavMYB genes have one or more C-terminal domains and had the 20 different conserved motifs. MYB TFs containing conserved N-terminal domain and transcriptional activations or suppression C-terminal regions might play a significant role in complicated physiological processes. Moreover, mostly *PavMYBs* may be implicated in the regulation of dormancy. Gene expression patterns may deliver significant evidences for discovering the genes function. In current study, RNA-seq data was used to visualize the expression analysis on different stages of bud dormancy. We analyzed gene expression patterns at five distinct bud dormancy phases to precisely understand the complex gene expression profile of sweet cherry MYB gene family members. The *PavMYB* genes have diverse stage-specific expression patterns, which were further separated into three groups based on the expression patterns in various stages of bud dormancy (Table S7). First group contained 38 MYB members that were extensively expressed in all phases of bud dormancy. That expression pattern demonstrated that these *PavMYB* genes could have an involvement in bud dormancy. Second group contained next 23 genes had comparable reduced expression in these phases of dormancy, whereas some genes had expression that was particular to specific phase of dormancy. Eight *PavMYB* genes (out of 69) were categorized into the 3rd group that were not expressed at any kind of bud dormancy stage (Table S7). The Genes which were remain silent on dormancy stage, it might be possible that they had some others functions.

To justify the RNA-seq results, gene expression patterns can be used to deduce this essential information regarding gene activities. Recent research demonstrated MYB genes have been found to be expressed in pear tissue (bud, flower and fruit) [95, 96]. Moreover, the function of *PavMYBs* in dormancy remains unknown in sweet cherry. In order to qRT-PCR, 28 *PavMYB* genes were selected from each subfamily of phylogenetic tree for relative expression patterns in different tissues (Bud, Flower and fruit) of sweet cherry. In our current finding, most of the *PavMYB* genes were highly expressed in bud dormancy phase and 4 genes showed transcripts abundance in fruits and flowers. This phenomenon explained that MYB genes had a vital role in dormancy regulation in sweet cherry. Expression profiling of *PavMYB* genes could help us for future understand the functional characteristics of this gene family in plant growth /development and in dormancy phase. These results is based on the evidence from whole-genome analysis with bioinformatics tools, qRT-PCR and RNA-seq data revealed that *PavMYB* play vital role in bud dormancy.

Conclusion:

The complete genome of sweet cherry (*Prunus avium* L.) provides the resources for functional genomics of MYB gene family. A comprehensive and systematic analysis has been performed and identified 69 *PavMYB* genes. *PavMYB* genes were grouped into twenty eight subfamilies according to phylogenetic analysis and conserved domain. Gene duplication events (TRD, WGD, TD, DSD and PD) were identified as the driving force for the expansion of the MYB genes family.

Bioinformatics analysis was performed including, gene structure (conserved motif and introns/exons), molecular characteristics (molecular weight, isoelectric point, and amino acid lengths), cis-acting elements, chromosomal localizations, transcriptomic profiling, conserved domain, synonymous and non-

synonymous ratios, and collinearity relationship. Furthermore, the qRT-PCR analysis authenticated the expression profile assessed by RNA seq analysis of sweet cherry at different floral developmental stages. Results verified that MYB genes had a strong role in bud dormancy. Genome-wide study of MYB genes gives insights into the evolutionary history and has set a foundation for genes role and functional features, and molecular mechanism in the plant dormancy process.

Abbreviations

MYB, myeloblastosis; HMM, Hidden Markov Model; GO, gene ontology; NJ, neighbor-joining; GSDS, gene structure display server; MEME, Multiple Em for Motif Elicitation; CDS, coding sequence; MW, molecular weight; pI, isoelectric point; MCSscanX, multiple collinearity scan toolkit; TPM, Transcripts Per Kilobase Million; gDNA, genomic DNA; RNA, ribonucleic acid; qRT-PCR, real-time quantitative reverse transcription PCR.

Declarations

Ethics declaration

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The *Prunus avium* MYB protein sequences were collected from GDR database (<https://www.rosaceae.org/>). The Arabidopsis MYB protein sequences were downloaded from TAIR database (<http://www.arabidopsis.org>). All RNA-Seq data were downloaded from NCBI SRA database under the following accession number, SRR8984342, SRR8984344, SRR8984360, SRR8984359, SRR8984367, SRR8984366, SRR8984382, SRR8984381, SRR8984403 and SRR8984402.

Competing interests

the authors declare that they have no competing interests.

Funding

The study was supported by grants from National Postdoctoral Program for Innovative Talents (Grant No. BX20180199), the National Key Research and Development Plan project (Grant No. 2018YFD0201305), China Postdoctoral Science Foundation (Grant No. 2018M642028).

Authors Contributions

IAS and MAM conceived and designed the experiments; IHS, XL, MSZ, SJ, JW, and MA contributed reagents/materials/analysis tools; CZ provided guidance on the whole manuscript. IAS and MAM wrote the article. All authors read and approved the final manuscript.

Acknowledgements

We appreciate Pro. Dr. Caixi Zhang and Dr. Muhammad Abdullah for his effort in revising and provided guidance on the whole manuscript.

Corresponding authors

Correspondence to Caixi Zhang

References

1. Karin M: **Too many transcription factors: positive and negative interactions.** *The New Biologist* 1990, **2**(2):126–131.
2. de Mendoza A, Sebé-Pedrós A, Šestak MS, Matejčić M, Torruella G, Domazet-Lošo T, Ruiz-Trillo I: **Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages.** *Proceedings of the National Academy of Sciences* 2013, **110**(50):E4858-E4866.
3. Peng X, Liu H, Wang D, Shen S: **Genome-wide identification of the *Jatropha curcas* MYB family and functional analysis of the abiotic stress responsive gene JcMYB2.** *BMC genomics* 2016, **17**(1):1–12.
4. Klempnauer K-H, Gonda TJ, Bishop JM: **Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene.** *Cell* 1982, **31**(2):453–463.
5. Paz-Ares J, Ghosal D, Wienand U, Peterson P, Saedler H: **The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators.** *The EMBO journal* 1987, **6**(12):3553–3558.

6. Weston K: **Myb proteins in life, death and differentiation.** *Current opinion in genetics & development* 1998, **8**(1):76–81.
7. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A: **The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation.** *Nature structural biology* 1996, **3**(2):178–187.
8. Jia L, Clegg MT, Jiang T: **Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes.** *Plant Physiology* 2004, **134**(2):575–585.
9. Roy S: **Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome.** *Plant signaling & behavior* 2016, **11**(1):e1117723.
10. Yuwei C, Leifeng X, Panpan Y, Hua X, Guoren H, Yuchao T, Junfang R, Jun M: **Differential Expression of Three R2R3-MYBs Genes Regulating Anthocyanin Pigmentation Patterns in Lilium spp.** *Acta Horticulturae Sinica* 2019, **46**(5):955.
11. Yifan Y, Bin D, Chengyong F, Liyuan Y, Hongbo Z: **Identification of the R2R3-MYB Family of Osmanthus fragrans and Its Expression in the Process of Flower Opening.** *Acta Horticulturae Sinica* 2020, **47**(10):2027.
12. Chen Y, Yang X, He K, Liu M, Li J, Gao Z, Lin Z, Zhang Y, Wang X, Qiu X: **The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family.** *Plant Mol Biol* 2006, **60**:107–124.
13. Du H, Feng B-R, Yang S-S, Huang Y-B, Tang Y-X: **The R2R3-MYB transcription factor gene family in maize.** *PloS one* 2012, **7**(6):e37463.
14. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC: **Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis.** *BMC genomics* 2012, **13**(1):1–19.
15. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q: **The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family.** *Plant molecular biology* 2006, **60**(1):107–124.
16. Uimari A, Strommer J: **Myb26: a MYB-like protein of pea flowers with affinity for promoters of phenylpropanoid genes.** *The Plant Journal* 1997, **12**(6):1273–1284.
17. Lea US, Sliemstad R, Smedvig P, Lillo C: **Nitrogen deficiency enhances expression of specific MYB and bHLH transcription factors and accumulation of end products in the flavonoid pathway.** *Planta* 2007, **225**(5):1245–1253.
18. Cominelli E, Tonelli C: **A new role for plant R2R3-MYB transcription factors in cell cycle regulation.** *Cell research* 2009, **19**(11):1231–1232.
19. Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR: **Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples.** *Plant physiology* 2006, **142**(3):1216–1232.
20. Hui L, XIONG J-s, JIANG Y-t, Li W, CHENG Z-mM: **Evolution of the R2R3-MYB gene family in six Rosaceae species and expression in woodland strawberry.** *Journal of Integrative Agriculture* 2019,

18(12):2753–2770.

21. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L: **MYB transcription factors in Arabidopsis**. *Trends in plant science* 2010, **15**(10):573–581.
22. Cao Y, Han Y, Li D, Lin Y, Cai Y: **MYB transcription factors in chinese pear (*Pyrus bretschneideri* Rehd.): genome-wide identification, classification, and expression profiling during fruit development**. *Frontiers in plant science* 2016, **7**:577.
23. He Q, Jones DC, Li W, Xie F, Ma J, Sun R, Wang Q, Zhu S, Zhang B: **Genome-wide identification of R2R3-MYB genes and expression analyses during abiotic stress in *Gossypium raimondii***. *Scientific reports* 2016, **6**(1):1–14.
24. Lau S-E, Schwarzacher T, Othman RY, Harikrishna JA: **dsRNA silencing of an R2R3-MYB transcription factor affects flower cell shape in a *Dendrobium* hybrid**. *BMC plant biology* 2015, **15**(1):1–14.
25. Qiu Z, Yan S, Xia B, Jiang J, Yu B, Lei J, Chen C, Chen L, Yang Y, Wang Y: **The eggplant transcription factor MYB44 enhances resistance to bacterial wilt by activating the expression of spermidine synthase**. *Journal of experimental botany* 2019, **70**(19):5343–5354.
26. Geng D, Chen P, Shen X, Zhang Y, Li X, Jiang L, Xie Y, Niu C, Zhang J, Huang X: **MdMYB88 and MdMYB124 enhance drought tolerance by modulating root vessels and cell walls in apple**. *Plant physiology* 2018, **178**(3):1296–1309.
27. Xie Y, Chen P, Yan Y, Bao C, Li X, Wang L, Shen X, Li H, Liu X, Niu C: **An atypical R2R3 MYB transcription factor increases cold hardiness by CBF-dependent and CBF-independent pathways in apple**. *New Phytologist* 2018, **218**(1):201–218.
28. Lang G, Early JD, Martin G, Darnell R: **Endo-, para-, and ecodormancy: physiological terminology and classification for dormancy research**. *HortScience* 1987, **22**(3):371–377.
29. MØLMANN JA, Asante DK, Jensen JB, Krane MN, Ernstsens A, Junttila O, Olsen JE: **Low night temperature and inhibition of gibberellin biosynthesis override phytochrome action and induce bud set and cold acclimation, but not dormancy in PHYA overexpressors and wild-type of hybrid aspen**. *Plant, Cell & Environment* 2005, **28**(12):1579–1588.
30. Ruonala R, Rinne PL, Baghour M, Moritz T, Tuominen H, Kangasjärvi J: **Transitions in the functioning of the shoot apical meristem in birch (*Betula pendula*) involve ethylene**. *The Plant Journal* 2006, **46**(4):628–640.
31. Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm Jr, Bhalerao RP, Boerjan W, Rohde A: **A molecular timetable for apical bud formation and dormancy induction in poplar**. *The Plant Cell* 2007, **19**(8):2370–2390.
32. Rohde A, Bastien C, Boerjan W: **Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar**. *Tree physiology* 2011, **31**(5):472–482.
33. Heide OM: **Interaction of photoperiod and temperature in the control of growth and dormancy of *Prunus* species**. *Scientia Horticulturae* 2008, **115**(3):309–314.
34. Hao X, Yang Y, Yue C, Wang L, Horvath DP, Wang X: **Comprehensive transcriptome analyses reveal differential gene expression profiles of *Camellia sinensis* axillary buds at para-, endo-, ecodormancy,**

- and bud flush stages.** *Frontiers in plant science* 2017, **8**:553.
35. Howe GT, Horvath DP, Dharmawardhana P, Priest HD, Mockler TC, Strauss SH: **Extensive transcriptome changes during natural onset and release of vegetative bud dormancy in Populus.** *Frontiers in plant science* 2015, **6**:989.
36. Lee HG, Lee K, Seo PJ: **The Arabidopsis MYB96 transcription factor plays a role in seed dormancy.** *Plant molecular biology* 2015, **87**(4–5):371–381.
37. Legave J-M, Guédon Y, Malagi G, El Yaacoubi A, Bonhomme M: **Differentiated responses of apple tree floral phenology to global warming in contrasting climatic regions.** *Frontiers in Plant Science* 2015, **6**:1054.
38. Snyder RL, Melo Abreu Jd: **Frost protection: fundamentals, practice and economics:** FAO, Roma (Italia). 2005.
39. Jung S, Ficklin SP, Lee T, Cheng C-H, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S: **The genome database for Rosaceae (GDR): year 10 update.** *Nucleic acids research* 2014, **42**(D1):D1237-D1244.
40. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J: **Pfam: the protein families database.** *Nucleic acids research* 2014, **42**(D1):D222-D230.
41. Zdobnov EM, Apweiler R: **InterProScan—an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**(9):847–848.
42. Letunic I, Doerks T, Bork P: **SMART 7: recent updates to the protein domain annotation resource.** *Nucleic acids research* 2012, **40**(D1):D302-D305.
43. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A: **Protein identification and analysis tools on the ExPASy server.** *The proteomics protocols handbook* 2005:571–607.
44. Abdullah M, Cao Y, Cheng X, Meng D, Chen Y, Shakoor A, Gao J, Cai Y: **The sucrose synthase gene family in Chinese pear (*Pyrus bretschneideri* Rehd.): structure, expression, and evolution.** *Molecules* 2018, **23**(5):1144.
45. Manzoor MA, Cheng X, Li G, Su X, Abdullah M, Cai Y: **Gene structure, evolution and expression analysis of the P-ATPase gene family in Chinese pear (*Pyrus bretschneideri*).** *Computational Biology and Chemistry* 2020, **88**:107346.
46. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Molecular biology and evolution* 1987, **4**(4):406–425.
47. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic acids research* 1997, **25**(24):4876–4882.
48. Letunic I, Bork P: **Interactive Tree Of Life (iTOL) v4: recent updates and new developments.** *Nucleic acids research* 2019, **47**(W1):W256-W259.
49. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.** *Nucleic acids research* 2012, **40**(7):e49-e49.

50. Chen C, Chen H, He Y, Xia R: **TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface.** *BioRxiv* 2018:289660.
51. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome research* 2009, **19**(9):1639–1645.
52. Lee T-H, Tang H, Wang X, Paterson AH: **PGDD: a database of gene and genome duplication in plants.** *Nucleic acids research* 2012, **41**(D1):D1152-D1158.
53. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH: **Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants.** *Genome biology* 2019, **20**(1):1–23.
54. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M: **Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions.** *Nucleic acids research* 2013, **41**(12):e121-e121.
55. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server.** *Bioinformatics* 2015, **31**(8):1296–1297.
56. Voorrips R: **MapChart: software for the graphical presentation of linkage maps and QTLs.** *Journal of heredity* 2002, **93**(1):77–78.
57. Sneddon TP, Li P, Edmunds SC: **GigaDB: announcing the GigaScience database.** *GigaScience* 2012, **1**(1).
58. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL: **Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.** *Nature protocols* 2016, **11**(9):1650.
59. 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):402–408.
60. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L: **MYB transcription factors in Arabidopsis.** *Trends Plant Sci* 2010, **15**(10):573–581.
61. Stracke R, Werber M, Weisshaar B: **The R2R3-MYB gene family in Arabidopsis thaliana.** *Curr Opin Plant Biol* 2001, **4**(5):447–456.
62. Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV: **Prevalence of intron gain over intron loss in the evolution of paralogous gene families.** *Nucleic Acids Res* 2004, **32**(12):3724–3733.
63. Fedorov A, Merican AF, Gilbert W: **Large-scale comparison of intron positions among animal, plant, and fungal genes.** *Proceedings of the National Academy of Sciences* 2002, **99**(25):16128–16133.
64. Li G, Liu X, Liang Y, Zhang Y, Cheng X, Cai Y: **Genome-wide characterization of the cellulose synthase gene superfamily in *Pyrus bretschneideri* and reveal its potential role in stone cell formation.** *Functional & Integrative Genomics* 2020, **20**(5):723–738.
65. Zhang LF, Li WF, Han SY, Yang WH, Qi LW: **cDNA cloning, genomic organization and expression analysis during somatic embryogenesis of the translationally controlled tumor protein (TCTP) gene from Japanese larch (*Larix leptolepis*).** *Gene* 2013, **529**(1):150–158.
66. Kumar D, Patro S, Ghosh J, Das A, Maiti IB, Dey N: **Development of a salicylic acid inducible minimal sub-genomic transcript promoter from Figwort mosaic virus with enhanced root-and leaf-activity using TGACG motif rearrangement.** *Gene* 2012, **503**(1):36–47.

67. Song YH, Yoo CM, Hong AP, Kim SH, Jeong HJ, Shin SY, Kim HJ, Yun D-J, Lim CO, Bahk JD: **DNA-binding study identifies C-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and LONG HYPOCOTYL5 proteins.** *Plant physiology* 2008, **146**(4):1862–1877.
68. Meng K, Hou Y, Han Y, Ban Q, He Y, Suo J, Rao J: **Exploring the functions of 9-lipoxygenase (DkLOX3) in ultrastructural changes and hormonal stress response during persimmon fruit storage.** *International journal of molecular sciences* 2017, **18**(3):589.
69. Zhang Y, Li Q, Xu L, Qiao X, Liu C, Zhang S: **Comparative analysis of the P-type ATPase gene family in seven Rosaceae species and an expression analysis in pear (*Pyrus bretschneideri* Rehd.).** *Genomics* 2020, **112**(3):2550–2563.
70. Wei B, Zhang R-Z, Guo J-J, Liu D-M, Li A-L, Fan R-C, Mao L, Zhang X-Q: **Genome-wide analysis of the MADS-box gene family in *Brachypodium distachyon*.** *PLoS One* 2014, **9**(1):e84781.
71. Linlin X, Xin Q, Mingyue Z, Shaoling Z: **Genome-wide analysis of aluminum-activated malate transporter family genes in six rosaceae species, and expression analysis and functional characterization on malate accumulation in chinese white pear.** *Plant Science* 2018, **274**:451–465.
72. Chen Q, Li Q, Qiao X, Yin H, Zhang S: **Genome-wide identification of lysin motif containing protein family genes in eight rosaceae species, and expression analysis in response to pathogenic fungus *Botryosphaeria dothidea* in Chinese white pear.** *BMC genomics* 2020, **21**(1):1–20.
73. Fawcett JA, Maere S, Van De Peer Y: **Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event.** *Proceedings of the National Academy of Sciences* 2009, **106**(14):5737–5742.
74. Hurst LD: **The Ka/Ks ratio: diagnosing the form of sequence evolution.** *Trends in genetics: TIG* 2002, **18**(9):486–486.
75. Matus JT, Aquea F, Arce-Johnson P: **Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes.** *BMC plant biology* 2008, **8**(1):1–15.
76. Soler M, Camargo ELO, Carocha V, Cassan-Wang H, San Clemente H, Savelli B, Hefer CA, Paiva JAP, Myburg AA, Grima-Pettenati J: **The *Eucalyptus grandis* R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function.** *New Phytologist* 2015, **206**(4):1364–1377.
77. Muthamilarasan M, Khandelwal R, Yadav CB, Bonthala VS, Khan Y, Prasad M: **Identification and molecular characterization of MYB transcription factor superfamily in C 4 model plant foxtail millet (*Setaria italica* L.).** *PloS one* 2014, **9**(10):e109920.
78. Wang J, Liu W, Zhu D, Hong P, Zhang S, Xiao S, Tan Y, Chen X, Xu L, Zong X: **Chromosome-scale genome assembly of sweet cherry (*Prunus avium* L.) cv. Tieton obtained using long-read and Hi-C sequencing.** *Horticulture research* 2020, **7**(1):1–11.
79. Cao Z-H, Zhang S-Z, Wang R-K, Zhang R-F, Hao Y-J: **Genome wide analysis of the apple MYB transcription factor family allows the identification of MdoMYB121 gene conferring abiotic stress tolerance in plants.** *PLoS One* 2013, **8**(7):e69955.

80. Liu X, Yu W, Zhang X, Wang G, Cao F, Cheng H: **Identification and expression analysis under abiotic stress of the R2R3-MYB genes in Ginkgo biloba L.** *Physiology and Molecular Biology of Plants* 2017, **23**(3):503–516.
81. Li W-H, Gu Z, Wang H, Nekrutenko A: **Evolutionary analyses of the human genome.** *Nature* 2001, **409**(6822):847–849.
82. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana.** *BMC plant biology* 2004, **4**(1):1–21.
83. Liu C, Xie T, Chen C, Luan A, Long J, Li C, Ding Y, He Y: **Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus).** *BMC genomics* 2017, **18**(1):1–16.
84. Liu C, Wang X, Xu Y, Deng X, Xu Q: **Genome-wide analysis of the R2R3-MYB transcription factor gene family in sweet orange (Citrus sinensis).** *Molecular biology reports* 2014, **41**(10):6769–6785.
85. Sun W, Ma Z, Chen H, Liu M: **MYB gene family in potato (Solanum tuberosum L.): genome-wide identification of hormone-responsive reveals their potential functions in growth and development.** *International journal of molecular sciences* 2019, **20**(19):4847.
86. Valesco R, Zharkikh A, Affourtit J: **The genome of the domesticated apple (Malus x domestica Borkh.).** *Nat Genet* 2010, **42**:833–839.
87. Hou X-J, Li S-B, Liu S-R, Hu C-G, Zhang J-Z: **Genome-wide classification and evolutionary and expression analyses of citrus MYB transcription factor families in sweet orange.** *PLoS One* 2014, **9**(11):e112375.
88. Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV: **Analysis of evolution of exon-intron structure of eukaryotic genes.** *Briefings in bioinformatics* 2005, **6**(2):118–134.
89. Du H, Yang S-S, Liang Z, Feng B-R, Liu L, Huang Y-B, Tang Y-X: **Genome-wide analysis of the MYB transcription factor superfamily in soybean.** *BMC plant biology* 2012, **12**(1):1–22.
90. Wang M, Yue H, Feng K, Deng P, Song W, Nie X: **Genome-wide identification, phylogeny and expressional profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (Triticum aestivum L.).** *BMC genomics* 2016, **17**(1):1–22.
91. Lecharny A, Boudet N, Gy I, Aubourg S, Kreis M: **Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure.** *Journal of structural and functional genomics* 2003, **3**(1):111–116.
92. Seo PJ, Xiang F, Qiao M, Park J-Y, Lee YN, Kim S-G, Lee Y-H, Park WJ, Park C-M: **The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in Arabidopsis.** *Plant physiology* 2009, **151**(1):275–289.
93. Seo PJ, Park CM: **MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in Arabidopsis.** *New Phytologist* 2010, **186**(2):471–483.
94. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y: **Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative**

recognition helices. *Cell* 1994, **79**(4):639–648.

95. Feng S, Wang Y, Yang S, Xu Y, Chen X: **Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10**. *Planta* 2010, **232**(1):245–255.
96. Xie M, Huang Y, Zhang Y, Wang X, Yang H, Yu O, Dai W, Fang C: **Transcriptome profiling of fruit development and maturation in Chinese white pear (*Pyrus bretschneideri* Rehd)**. *Bmc Genomics* 2013, **14**(1):1–20.

Figures

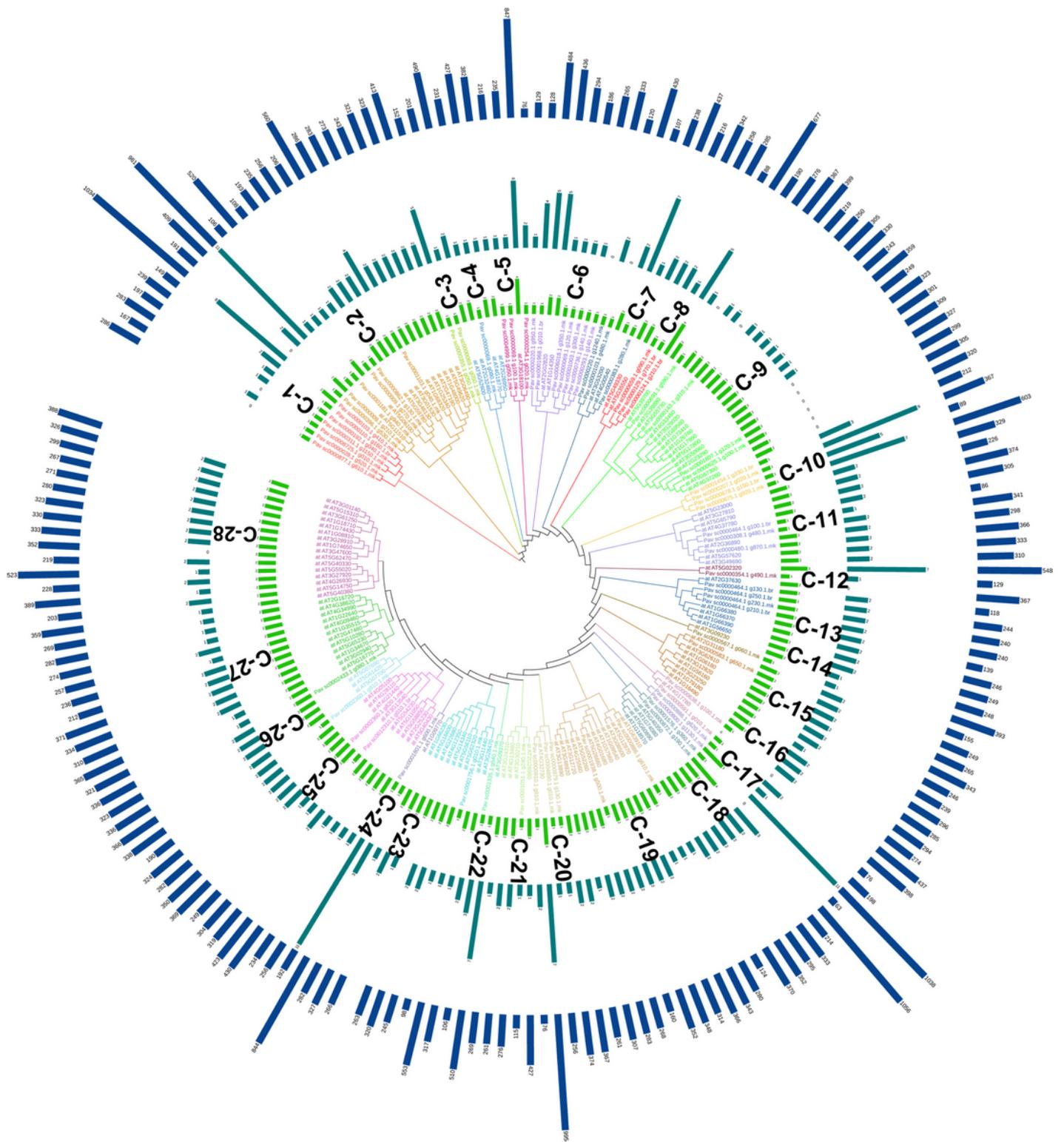


Figure 1

Phylogenetic tree of MYB protein *P. avium*, and *A. thaliana* with each color representing a subfamily (C1-C28) of MYB genes. Green, teal and blue bars indicate domain, number of introns and length of amino acids, respectively.

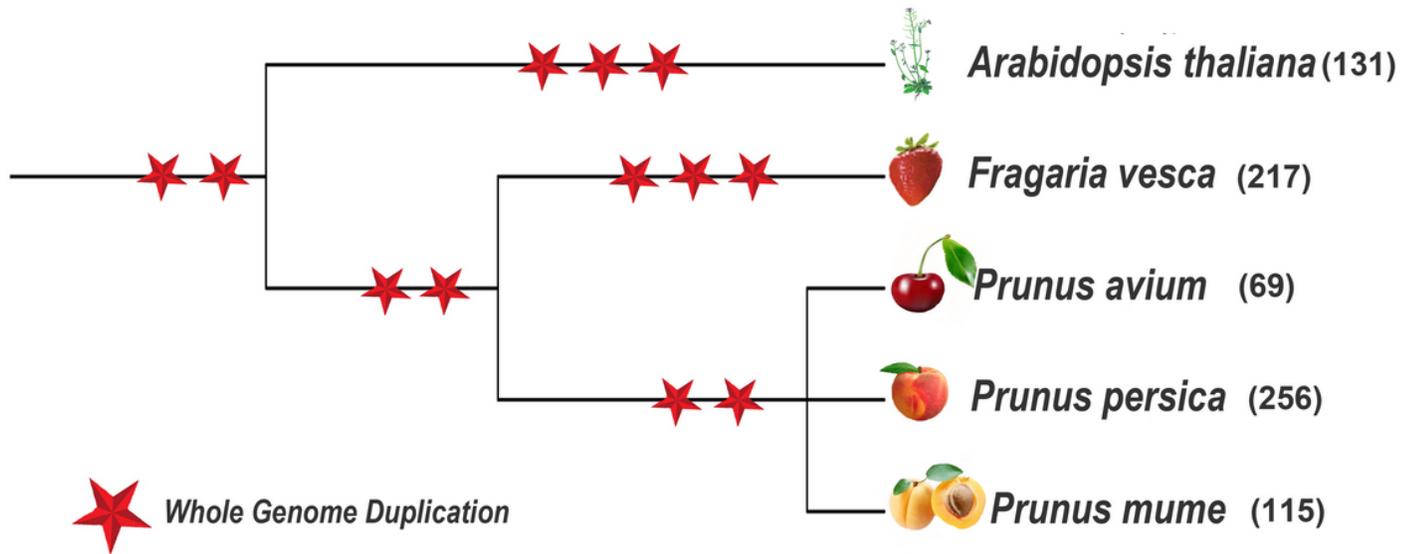


Figure 2

The phylogeny of the *Fragaria vesca*, *Prunus avium*, *Prunus persica*, *Prunus mume* and *Arabidopsis thaliana* genomes. The red star indicated whole genome duplication (WGD) events and the number in the parentheses indicates the total number of MYB members in each species.

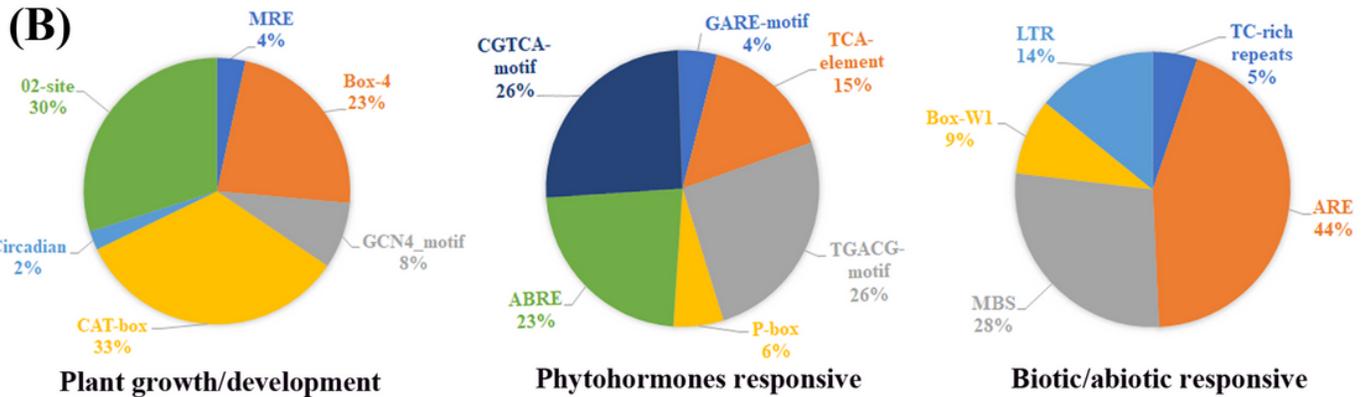
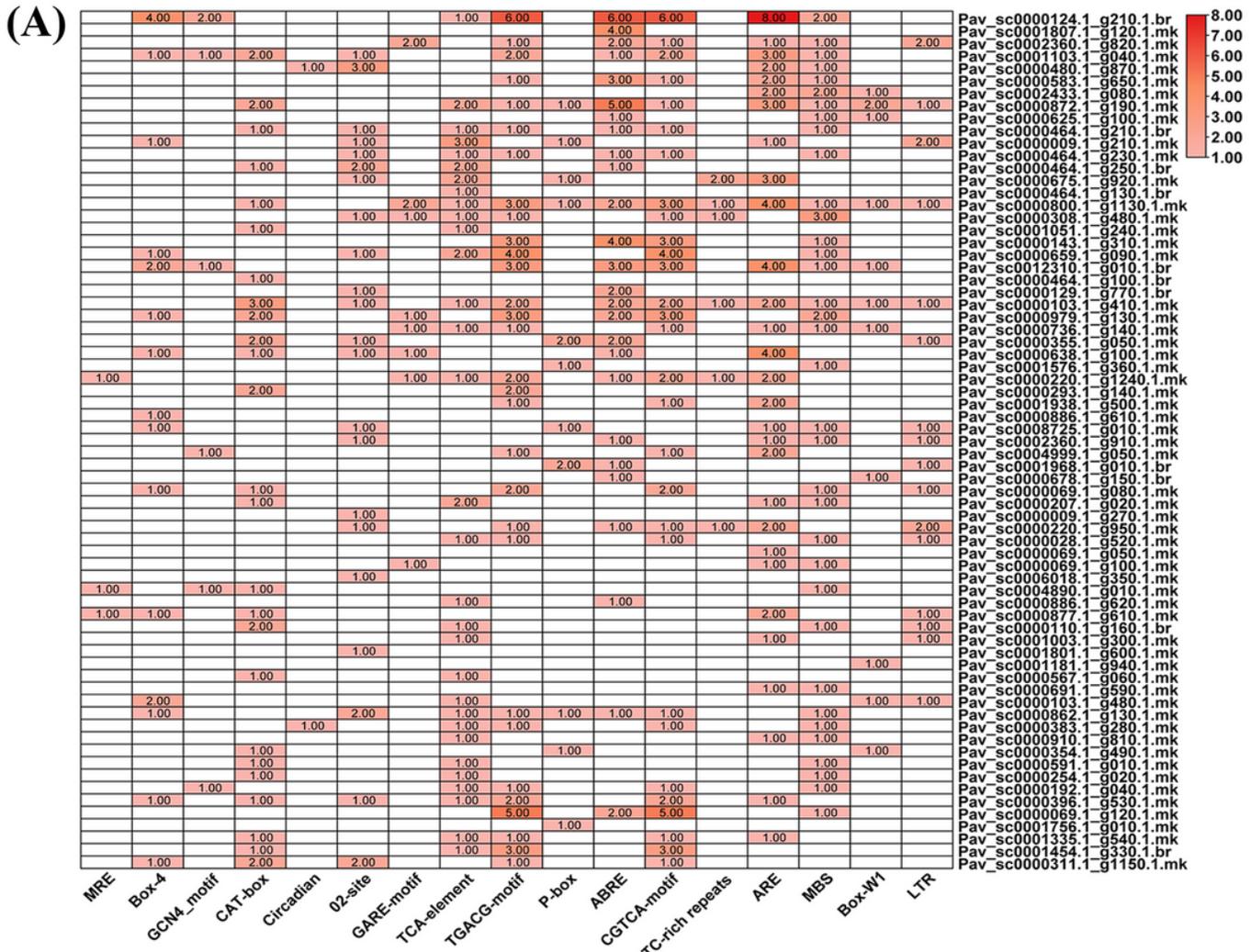


Figure 3

Prediction of cis-acting elements on promoter regions of PavMYB genes and classified into basic three groups (phytohormones responsive, biotic/abiotic stress and plant growth & development). (A) cis-acting elements of different number illustrated with specific colors. (B) The Pie charts size represented the percentage of promoter element in specific category.

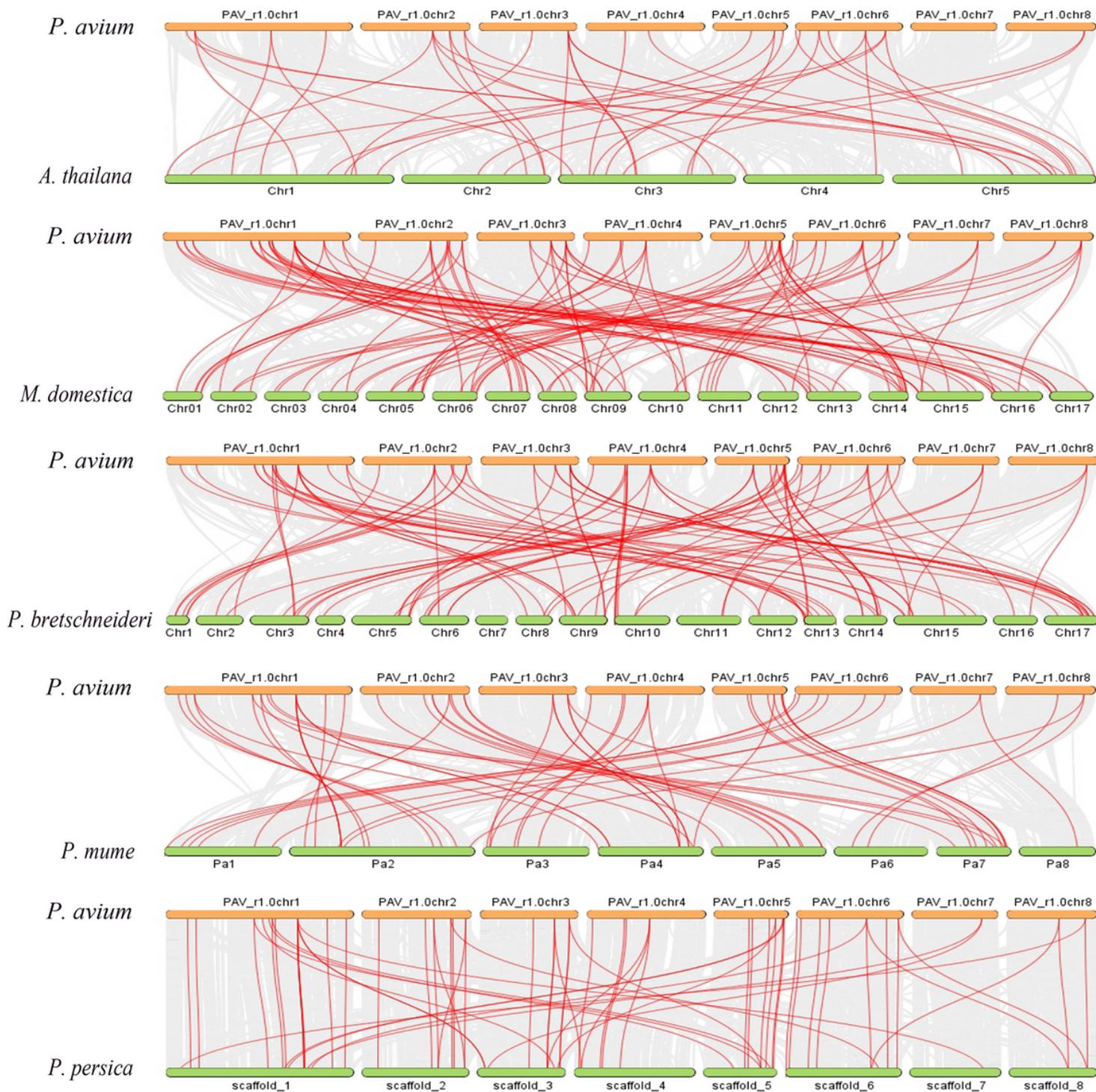


Figure 4

Syntenic relationship of MYB genes in *P. avium* with, *A. thaliana*, *M. domestica*, *P. bretschneideri*, *P. mume* and *P. persica* and syntenic genes pairs are connected with red color lines.

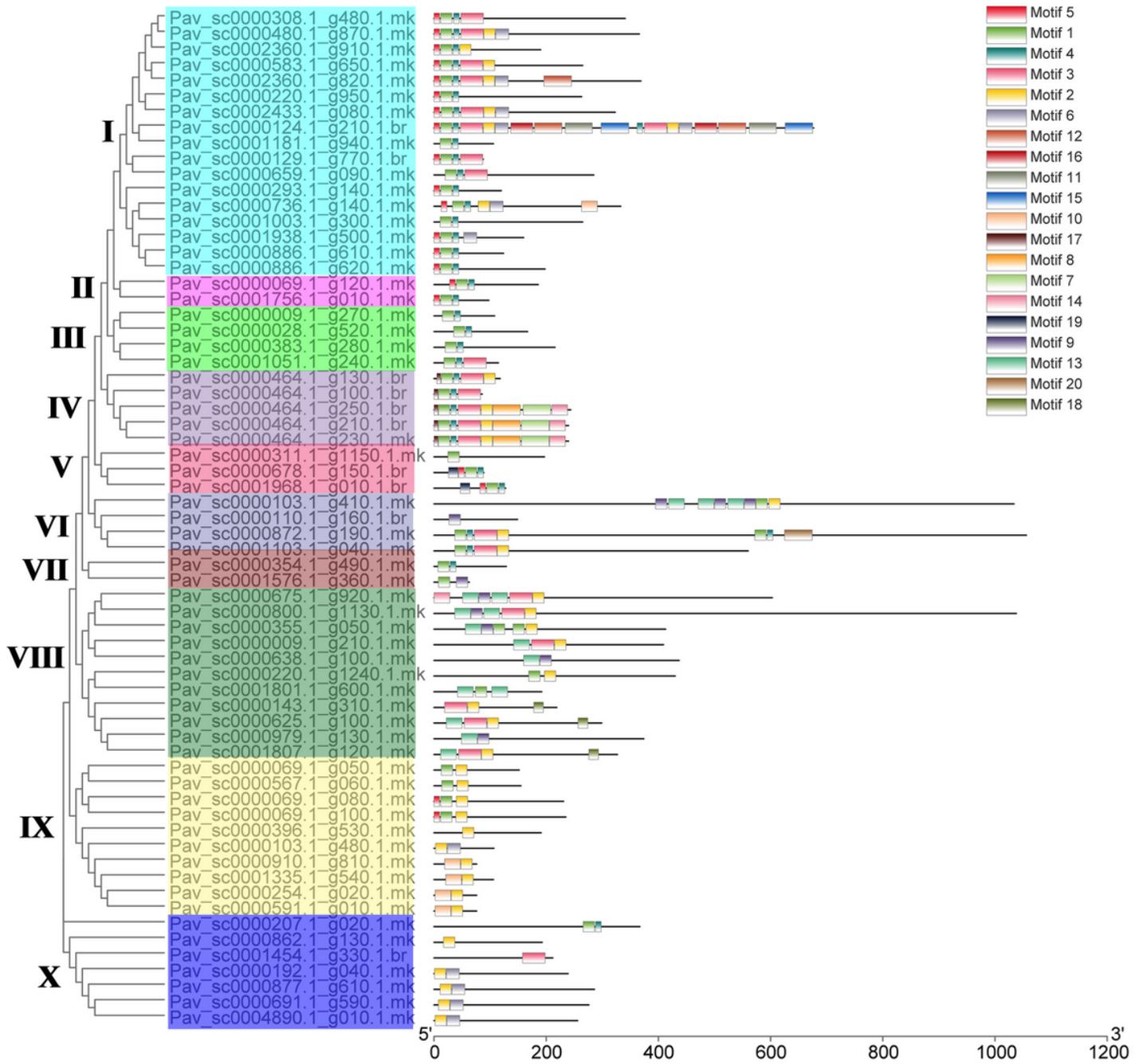


Figure 5

The phylogeny and conserved motifs in the MYB TFs in *Prunus avium* and figure legends are mentioned on the top.

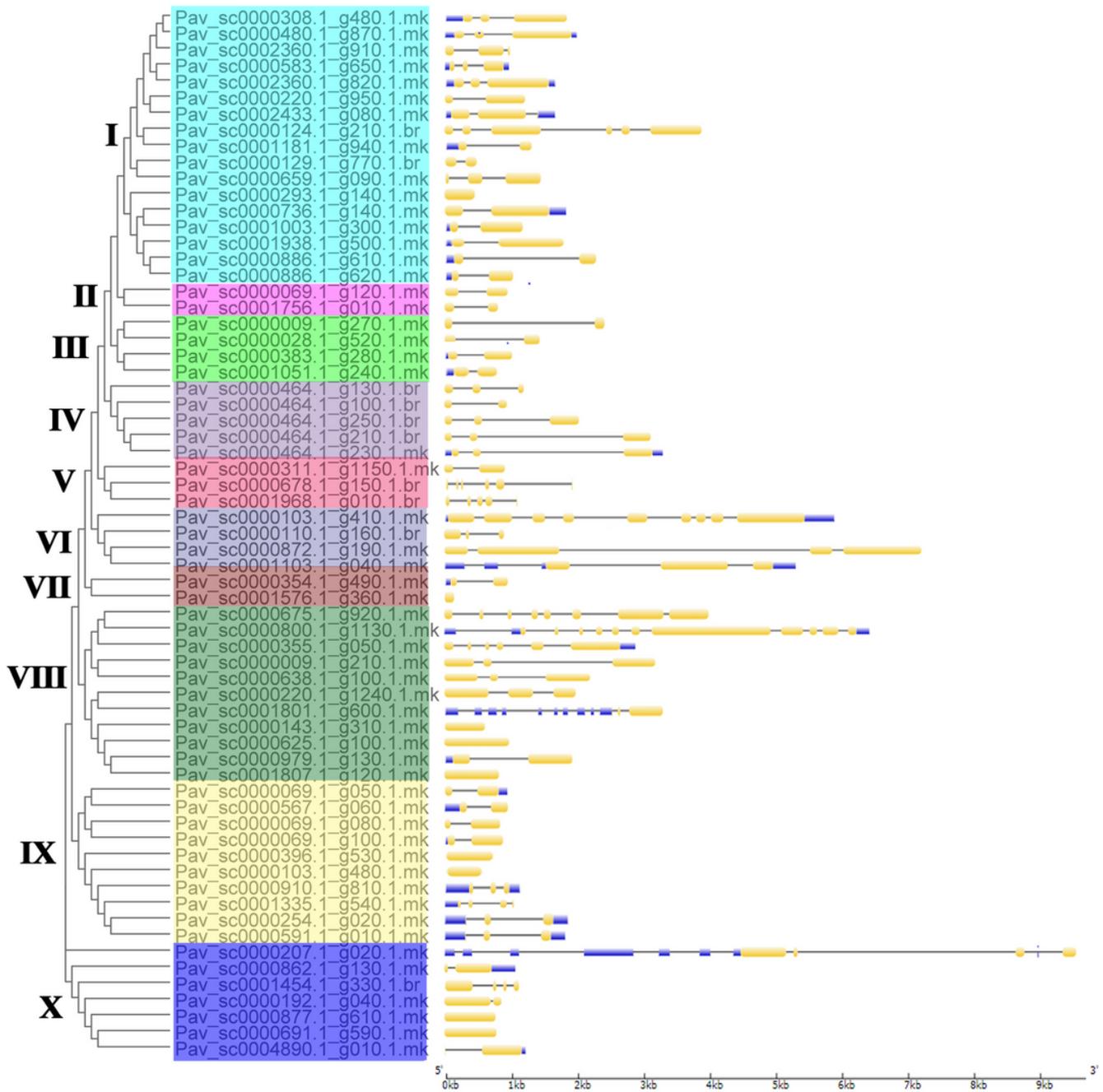


Figure 6

Phylogenetic relationships and gene structures (introns/exons) MYB genes of the sweet cherry. The gray line represented the intron (s) while yellow and blue color illustrated the exon (s) and UTR (s).

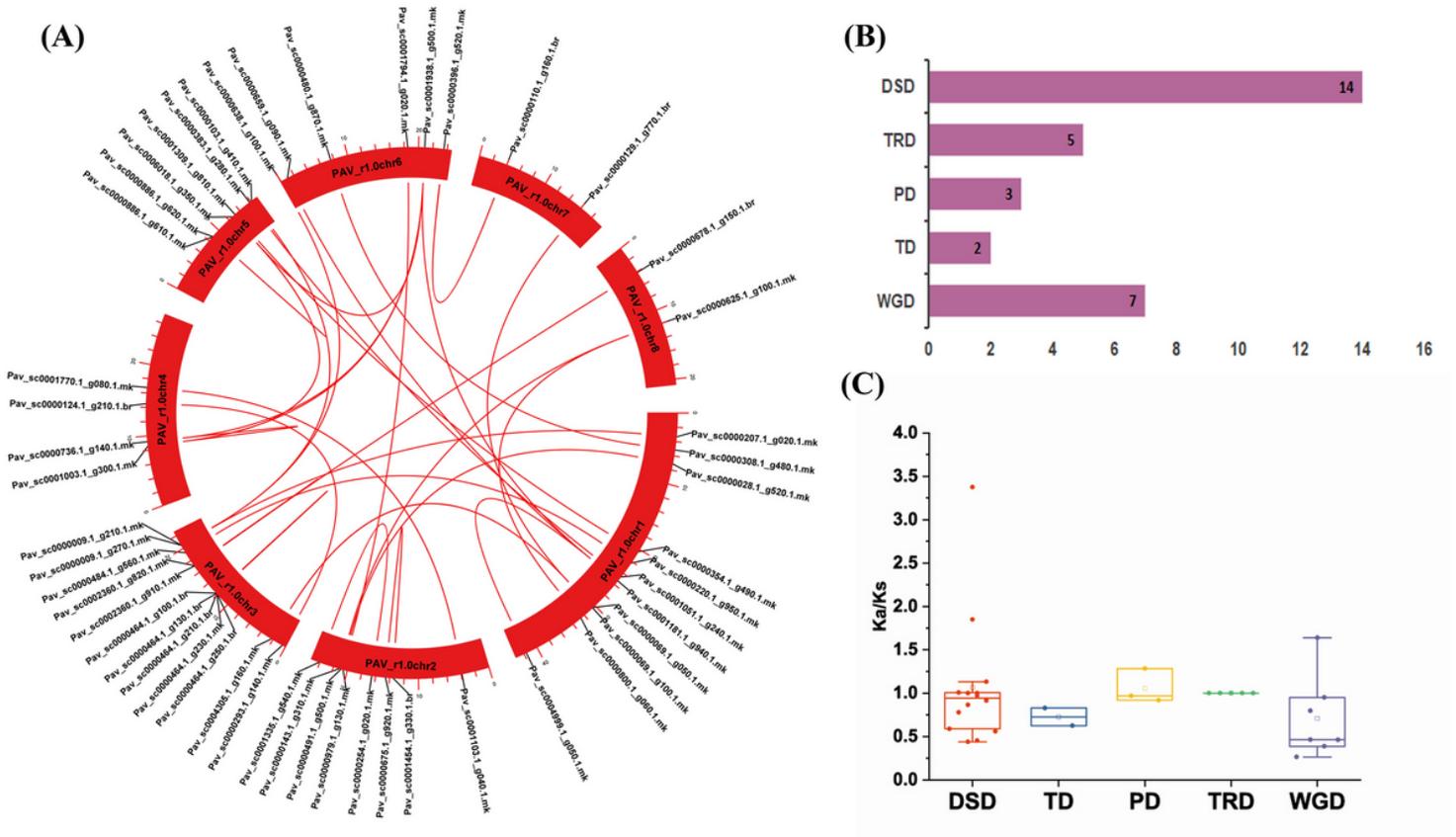


Figure 7

Gene duplication (DSD, WGD, PD, TRD and TD) events and localization of PavMYB genes family in *P. avium*. (A) Duplicated gene pairs placed on the different chromosomes and linked with red lines (B) different numbers represents the number of duplications in each event (C) calculated ka/ks values present the diverse gene duplication (DSD, TD, PD, TRD and WGD) events.

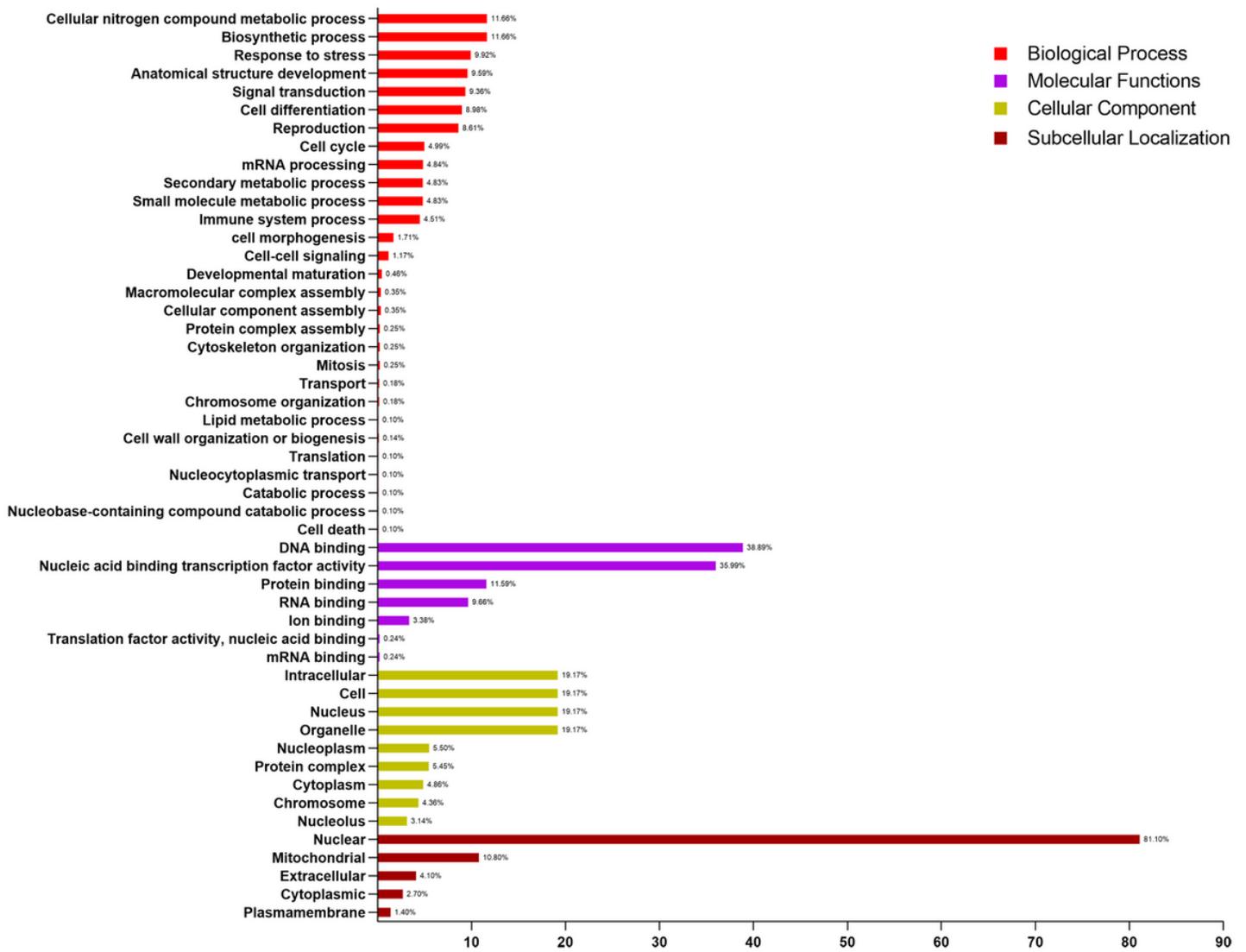


Figure 8

Predicted biological process, molecular functions, cellular components and subcellular localization of PavMYB gene family.

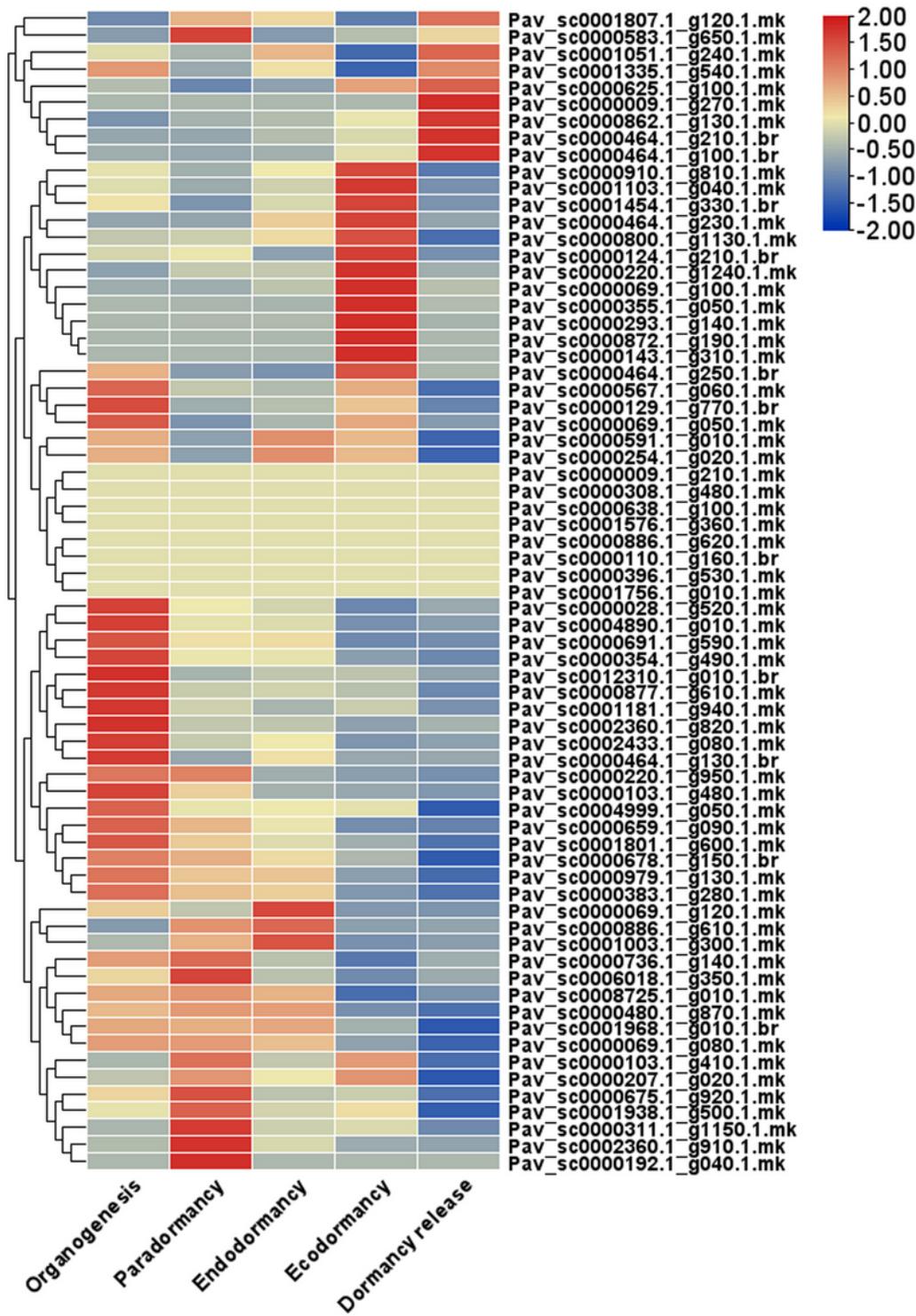


Figure 9

Heat map of PavMYB genes expression in different dormancy related phases data (organogenesis, paradormancy, endodormancy, ecodormancy and dormancy release). Blue, violet and red color indicated no expression, low expression and high expression respectively.

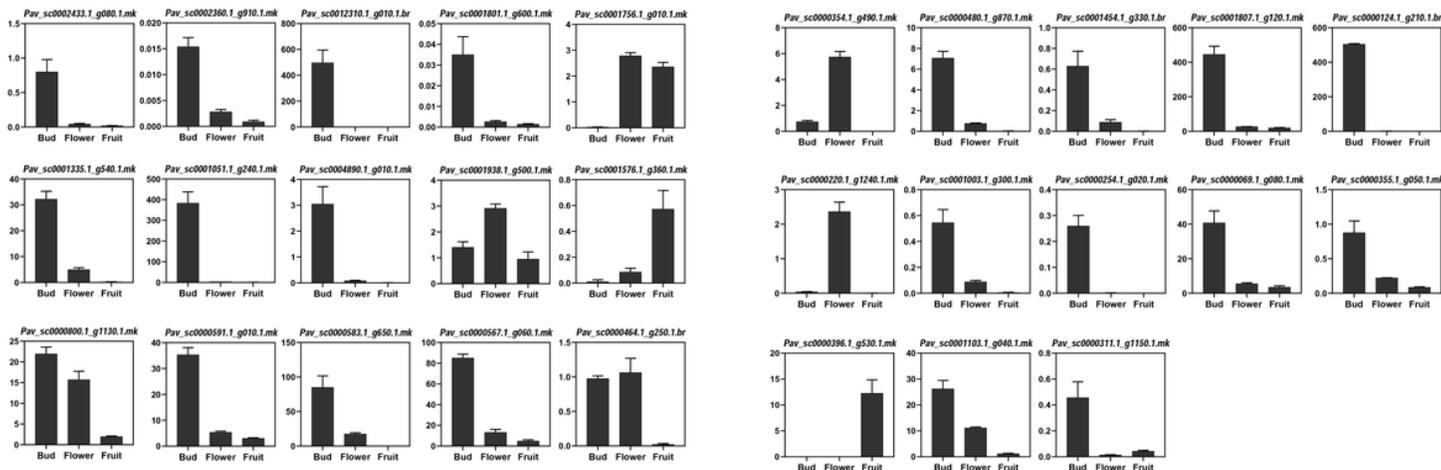


Figure 10

Relative expression patterns of PavMYB genes (C1-C28) through qRT-PCR on different tissues (bud, flower and fruit) in sweet cherry. Mean \pm SE of three biological replicates (each having 3 technical replicates).

Supplementary Files

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

- [Fig.S1.tif](#)
- [Fig.S2.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)