bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome-wide characterization of the R2R3-MYB transcription factors in pepper (*Capsicum* spp.) unveils the role of CaMYB101 as repressor in anthocyanin biosynthesis

Running head: R2R3-MYB transcriptional repressors in *Capsicum* spp.

Ying Liu[1, 2, 3#], Yi Wu[1,4#], Zicheng Wang[5], Shiya Zhang[5], Xintong Liu[5], Yury Tikunov[1], Richard G.F. Visser[1], Rob E. Schouten[2], Leo F.M. Marcelis[2], Zhao Zhang[1, 5*] and Arnaud Bovy[1*]

[1]Plant Breeding, Wageningen University and Research, Wageningen, Netherlands.

[2]Horticulture and Product Physiology, Wageningen University and Research, Wageningen, Netherlands.

[3]Graduate School Production Ecology & Resource Conservation, Wageningen University and Research, Wageningen, Netherlands.

[4]Graduate School Experimental Plant Sciences, Wageningen University and Research, Wageningen, the Netherlands.

[5]Beijing Key Laboratory of Development and Quality Control of Ornamental Crops, Department of Ornamental Horticulture, China Agricultural University, Beijing, China

[#]These authors made an equivalent contribution

[*]**Authors for correspondence:**

Arnaud Bovy, Plant Breeding, Wageningen University and Research, Droevendaalsesteeg 1, Wageningen, 6708PB the Netherlands. e-mail: arnaud.bovy@wur.nl

Zhao Zhang, Department of Ornamental Horticulture, China Agricultural University, Yuanmingyuan Xilu 2, Beijing 100193, China. Tel. 0086-10-62732724, e-mail: zhangzhao@cau.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Fruit colour is one of the most important commercial traits of pepper (*Capsicum* spp.), a major horticultural crop worldwide. Some pepper accessions temporarily accumulate anthocyanins during fruit development and gradually lose them upon fruit ripening. Meanwhile, anthocyanin biosynthesis gradually stops. However, how this process is exactly regulated is still largely unknown. The R2R3-MYB transcription factor is one of the largest plant transcription factor families, and it is considered the most important regulator for the biosynthesis of anthocyanins and other flavonoids. Although R2R3-MYBs are widely studied in many plants, research in pepper has been limited. In this study, we performed a genome-wide analysis of R2R3-MYBs across three cultivated pepper species (*C. annuum, C. baccatum*, and *C. chinense*) involving identification, chromosome localization, gene structure analysis, phylogenetic analysis and collinearity analysis. Candidate R2R3-MYB repressors were further identified based on repression motifs. An R2R3-MYB gene, *CaMYB101*, was selected based on its high homology with anthocyanin biosynthesis repressors in tomato and petunia as well as its high expression level in fruit when purple pigmentation started to discolour. By using virus-induced gene silencing, CaMYB101 was characterized as an anthocyanin biosynthesis repressor. To our knowledge, CaMYB101 is the first transcriptional repressor associated with anthocyanin biosynthesis identified in pepper.

Keywords: pepper, MYB transcription factor, *Capsicum annuum*, *Capsicum chinense*, *Capsicum baccatum*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Transcription factors (TFs) are involved in stimulating or repressing the transcription of target genes to control physiological and metabolic processes during plant growth and development in response to endogenous or exogenous stimuli (Fuda et al., 2009). Based on their DNA-binding domains, TFs can be classified into different families. The MYB family is one of the largest TF families in plants and is involved in various biological processes including plant growth, circadian clock control, and primary and secondary metabolism regulation (Dubos et al., 2010). A highly conserved DNA-binding domain, known as the MYB domain, contains up to four incomplete repeats at the N-terminus. Each repeat is composed of approximately 50-55 amino acid residues and forms three α-helices. The second and third helices form a helix-turn-helix architecture with three spaced tryptophan (or hydrophobic) residues (Ogata et al., 1992; Du et al., 2015). The C-terminus region of MYB proteins is highly variable and responsible for the diverse regulatory functions of MYB TFs. Depending on the number of tandem repeat(s) in the MYB domain, the MYB family has four subfamilies: 1R-MYB (MYB-related), 2R-MYB (R2R3-MYB), 3R-MYB (R1R2R3-MYB) and 4R-MYB (four R1/R2-like repeats) (Dubos et al., 2010). In plants, the first MYB (R2R3-MYB) TF COLORED1 (C1) was discovered from *Zea mays* and has been identified to be involved in the activation of anthocyanin biosynthesis (Grotewold et al. 1991). Over time, the MYB family has been identified in many plant species. In *Arabidopsis*, 198 AtMYBs were identified and the tomato genome included 127 SlMYBs (Yanhui et al., 2006; Li et al., 2016). Among the MYB subfamilies, R2R3-MYB is the dominant subfamily with various roles in, for example, flavonoid biosynthesis.

Flavonoids are a large group of secondary metabolites, including flavones, isoflavones, flavonols, anthocyanins and proanthocyanidins. The various flavonoids can serve as antioxidants, aromatic compounds, defense compounds, signaling molecules and pigments. In plants, there are two types of structural genes: (i) early flavonoid biosynthesis pathway genes (EBGs), encoding chalcone synthase (CHS), chalcone isomerase (CHI), flavonol 3-hydroxylase (F3H) and flavonol 3'-hydroxylase (F3'H), which are induced as precursors for the downstream genes, known as late biosynthetic genes (LBGs). (ii) The LBGs including dihydroflavonol 4-reductase (DFR), anthocyanin synthase (ANS) and UDP-glucose: flavonoid 3-glucosyltransferase (UFGT), leucoanthocyanidin reductase (LAR) and anthocyanin reductase (ANR) which lead to proanthocyanidin and anthocyanin biosynthesis. The regulation of flavonoid biosynthetic structural genes often involves R2R3-MYB TFs. The EGBs are activated by independent R2R3-MYBs such as AtMYB11, AtMYB12 and AtMYB111, while LBGs related with anthocyanin biosynthesis are regulated by an MYB-bHLH-WD40 (MBW) complex consisting of a MYB TF, a basic helix-loop-helix (bHLH) and a WD-repeat protein (Xu et al., 2015). There is one exception, F3'H, which is regulated by both independent R2R3-MYBs and an MBW complex (Stracke et al., 2007).

To date, the majority of the flavonoid biosynthesis R2R3-MYBs are transcriptional activators. For example,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

high anthocyanin pigmented apple fruits were observed by overexpression of apple *MdMYB10* in cv. 'Royal Gala', together with increased expression of *MdCHS*, *MdCHI*, *MdF3H*, *MdDFR*, *MdANS* and *MdUFGT* (Espley et al., 2007). In grapevine, VvMYB5a leads to an up-regulation of flavonoid biosynthesis and boosted flavonols, anthocyanins and proanthocyanins accumulation (Deluc et al., 2006; Deluc et al., 2008). In addition to flavonoid activators, R2R3-MYB repressors have also been identified and displayed a crucial role in balancing flavonoid biosynthesis. For instance, the grapevine VvMYB4-like and strawberry FaMYB1 suppress the flavonol and anthocyanin accumulation and cause a clear pigmentation loss in flowers (Aharoni et al., 2001; Perez-Diaz et al., 2016). In petunia, overexpression of *PhMYB27* causes reduced anthocyanin accumulation, while silencing of this gene enhances anthocyanin accumulation throughout the whole plant including leaves, stems, and flowers (Albert et al., 2014).

Pepper is one of the most popular horticultural crops worldwide. The cultivated pepper comprises several species within the genus *Capsicum*, such as *C. annuum*, *C. baccatum* and *C. chinense*. The fruits of pepper contain a wide range of health-related secondary metabolites, like carotenoids and flavonoids. Among peppers, there are some purple accessions with a high level of anthocyanins in fruit peels. Moreover, the overall flavonoid content of these purple accessions is generally higher compared to non-purple accessions (Liu et al., 2020). These purple pigments are due to a temporary anthocyanin accumulation before fruit ripening and after ripening the anthocyanin biosynthesis stops and degradation becomes dominant (Yamada et al., 2019). Several R2R3-MYBs have been reported to activate flavonoid biosynthesis in pepper (Borovsky et al., 2004; Zhang et al., 2015). However, no MYB repressors for flavonoid biosynthesis have been identified in pepper yet. Genome-wide identification of R2R3-MYB gene family has been studied in *C. annuum* (Wang et al., 2020; Arce-Rodríguez et al., 2021),. The recently released pepper genomes make the study of R2R3-MYB gene family in *C. baccatum* and *C. chinense* possible. We aim to globally analyse R2R3-MYB gene family in three *Capsicum* species and among them identify flavonoid-related R2R3-MYB represosrs in pepper. To accomplish this, a genome-wide analysis of R2R3-MYBs was performed within the three cultivated *Capsicum* species (*C. annuum*, *C. baccatum* and *C. chinense*), including gene structure, chromosomal location, synteny and phylogeny. As the result, candidate R2R3-MYB repressors in *Capsicum* spp. were identified. We showed that a R2R3-MYB repressor, named CaMYB101, plays a role in the negative regulation of anthocyanin biosynthesis in pepper fruits. Our results lay a foundation for characterizing R2R3-MYBs among *Capsicum* spp. and provide a better understanding for flavonoid-related repression mechanism in pepper.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Identification of R2R3-MYBs in *Capsicum* spp. genomes

The published *C. baccatum* and *C. chinense* genome sequences (Table 1) were used for putative R2R3-MYBs identification through HMM profiling (PF00249). After removing redundant proteins by BLASTP through NCBI-CDD, 106 and 110 R2R3-MYB transcription factors were identified in *C. baccatum* (CbMYBs) and *C. chinense* (CcMYBs), respectively. Meanwhile, 108 R2R3-MYBs of *C. annuum* (CaMYBs) were derived from the study of Wang et al. (2020). All R2R3-MYBs were named based on their species and chromosomal orders, with details including the accession number, chromosomal location, number of introns and exons as well as the length of coding sequence and protein (Table S1-S3). The longest CaMYB is CaMYB82 with 995 amino acids and the shortest is CaMYB58 with 110 amino acids. CbMYB21 is the longest MYB protein of all *Capsicum* spp., and contains 1210 amino acids, while CbMYB10 is the shortest CbMYB protein with 152 amino acid. The length of CcMYBs proteins varied from 137 (CcMYB66) to 1156 (CcMYB42) amino acids.

Sequence conservation within the R2 and R3 domains

Multiple sequence alignment analyses of all putative R2R3-MYBs derived from three *Capsicum* spp. as well as from *Arabidopsis* were performed to demonstrate the feature of homologous domain sequences and the corresponding amino acid frequency. Similar to *Arabidopsis*, the R2 and R3 domains of all the R2R3-MYBs among *Capsicum* spp. were highly conserved (Figure 1). Meanwhile, the tryptophans (W) are highly conserved in R2 (three) and R3(two) domains.

Chromosomal localization and synteny analysis of *Capsicum* R2R3-MYBs

To study the genetic divergence and gene duplication of R2R3-MYBs in pepper, the chromosomal location and synteny relationship of R2R3-MYBs was investigated. The Ca-, Cb- and CcMYBs were physically mapped throughout all 12 chromosomes in *C. annuum*, *C. baccatum* and *C. chinense*, respectively (Figure 2). In total 11 CaR2R3-MYBs, 9 CbR2R3-MYBs and 5 CcR2R3-MYBs were located on unanchored scaffolds and assigned to Chr. 00 (Table S1-S3). The distribution of R2R3-MYBs among three *Capsicum* spp. genomes was also not congruent. *C. annuum* and *C. chinense* had a relatively high density of R2R3-MYBs appeared at the top and bottom of chromosomes compared to *C. baccatum*. The maximum and minimum number of R2R3-MYBs allocated per genome were similar. In *C. annuum*, Chr. 03 contained the most CaR2R3-MYBs, 12, whereas Chr. 08 had the lowest number of CaR2R3-MYBs, four, over all chromosomes (Figure 2A). In *C. baccatum*, Chr. 01, 03, 05, 06 and 07 contained the highest number (10) CbR2R3-MYBs, while Chr. 04 and 09 had the lowest number of four CbR2R3-MYBs (Figure 2B). In *C. chinense*, Chr. 06 had the highest number (13) CcR2R3-MYBs while Chr. 08 was contained only two CcR2R3-MYBs (Figure 2C).

The syntenic paralogs and orthologs were identified throughout *Capsicum* spp.. Further analysis was performed for the synonymous (Ks) and non-synonymous (Ka) values to determine the selection pressures. In

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

total, 119 pairs of duplicated genes were identified (Table 2, Supplemental Table S4), including 95 pairs of inter-species orthologs and 24 pairs of intra-species paralogs. To be specific, 44 gene pairs were between *C. annuum* and *C. chinense*, 30 gene pairs were between *C. annuum* and *C. baccatum* and 21 gene pairs were between *C. baccatum* and *C. chinense* (Figure 3, Supplemental Table S4). Among the 24 intra-species duplications, 21 gene pairs were inter-chromosomal, while the remaining were within the same chromosome which are CaMYB12-CaMYB19, CaMYB13-CaMYB20 and CcMYB14-CcMYB22. All these three pairs were located on chromosome 2. The Ka/Ks ratios for segmental duplication were between 0.11 and 0.53 through the intra-species gene pairs. In the inter-species syntenic orthologs, two out of 95 pairs of duplicated genes showed Ka/Ks ratios larger than 1.0, indicating a positive selection on these gene pairs, while the rest pairs harboured the Ka/Ks smaller than 1.0, indicating these gene pairs with negative selection.

Gene structure analysis of *Capsicum* R2R3-MYBs

Phylogenetic analyses were performed on 108 CaR2R3-MYBs, 106 CbR2R3-MYBs and 110 CcR2R3-MYBs, respectively (Supplementary Figure 2). In addition, a comprehensive phylogenetic analysis was done on 125 *Arabidopsis* R2R3-MYBs, 324 summed R2R3-MYBs in three *Capsicum* spp. and 36 published anthocyanin related MYBs (Table 3, Supplemental Figure S1). The published anthocyanin-related R2R3-MYBs were composed of activators and repressors of anthocyanin biosynthesis (Table 3).

The exon-intron structure pattern of the MYBs was further investigated in *C. annuum*, *C. baccatum* and *C. chinense* (Supplementary Figure 2). Generally, over 95% of R2R3-MYBs in *Capsicum spp.* had at least one intron over the entire gene sequence. The majority of R2R3-MYBs in pepper had a structure which was composed of two introns with three exons (respectively 70 of 108 CaMYBs, 69 of 106 CbMYBs and 72 of 110 CcMYBs).

Identification of R2R3-MYB repressors

In total 53 candidate R2R3-MYB repressors were identified based on their repression motifs in *Capsicum* spp. (Supplementary Figure 2, Table 4). To be specific, 47 candidate R2R3-MYB repressors only had an EAR motif (15 CaMYBs, 15 CbMYBs and 17 CcMYBs), three had a LxLxPP motif (CaMYB37, CbMYB32 and CcMYB40) and three had both EAR- and LxLxPP motifs (CaMYB16, CbMYB13 and CcMYB16).

To identify anthocyanin-related R2R3-MYB repressors in pepper, a phylogenetic analysis was performed using candidate pepper MYB repressors together with 36 published anthocyanin MYB regulators from other plant species (Table 3, Figure 4). Meanwhile, CaMYB101, CbMYB89 and CcMYB92 were characterized within one cluster (red curve in Figure 4), together with two identified R2R3-MYB repressors from the *Solanaceae* family (petunia PhMYB27 and tomato SlMYBL2) (Albert et al., 2011; Zhang et al., 2019). In addition, two strawberry R2R3-MYB repressors (FaMYB1 and FcMYB1) and one *Medicago truncatula* R2R3-MYB repressor (MtMYB2) were the closest orthologs to the cluster of CcMYB87, CcMYB108, CbMYB67 and CaMYB70 and the cluster of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CaMYB101, CbMYB89 and CcMYB92. Moreover, CaMYB16, CbMYB13 and CcMYB16 were characterized together with an identified apple R2R3-MYB repressor (MdMYB6), while CcMYB91 was characterized together with an identified *Arabidopsis* R2R3-MYB repressor (AtMYB60).

CaMYB101, CbMYB89 and CcMYB92 had high amino acid similarity with other *Solanaceae* family members such as tomato SlMYBL2 and petunia PhMYB27, especially the R2, R3 domain and EAR motif (Figure 4B). It showed that the conserved R2, R3 domain and EAR motif of CaMYB101, CbMYB89 and CcMYB92 were identical. The tomato SlMYBL2 had two EAR motifs, LxLxL and DLNxxP, compared with the other four MYBs with one EAR motif.

Expression profile of candidate CaR2R3-MYB repressors in different tissues

The expression profile of 17 R2R3-CaMYB repressors, based on the RNA-seq transcriptome analysis of *C. annuum* cv. Zunla (Qin et al., 2014), revealed expression variation in different tissues and fruit developmental stages (Figure 5). Eight (CaMYB37, CaMYB45, CaMYB67, CaMYB70 CaMYB72, CaMYB91, CaMYB92 and CaMYB101) out of 17 R2R3-CaMYB repressors showed a relatively high expression level in fruits, especially at pre-breaker stage (3-4cm fruit length) and at post breaker stage (5 days after breaker fruits), respectively (Figure 5).

Functional analysis of CaMYB101 in *C. annuum* cv. Tequila

The *C. annuum* cv. Tequila is a transiently purple-fruited genotype of which the fruit is green in the early stages of fruit development and then turns purple and eventually red when fully ripe. To explore the involvement of MYB repressors in the regulation of fruit peel color in cv. Tequila, we investigated the candidate gene *CaMYB101*. CaMYB101 is a candidate R2R3-MYB repressor characterized by an EAR motif and, meanwhile, its orthologs in *C. baccatum* (CbMYB89) and *C. chinense* (CcMYB92) were also identified as repressors with an EAR motif (Table 4). Based on the phylogenetic analysis and transcriptome profiling, *CaMYB101* was selected as a candidate gene. CaMYB101 had high homology to tomato SlMYBL2 and petunia PhMYB27 (Figure 4) that are involved in the negative regulation of anthocyanin biosynthesis. The published transcriptome profile also demonstrated a relatively high expression level of *CaMYB101* in pepper fruits (cv. Zunla) (Figure 5), indicating it may function in fruits. In addition, the duplication analysis showed that it did not undergo a duplication event, which minimized the paralog-caused interference and, therefore, functional redundancy. First of all, the expression of *CaMYB101* in purple pepper background, cv. Tequila, was verified by RT-qPCR in different tissues and fruits at different developmental stages (Figure 6). The expression of *CaMYB101* increased from the fertilized ovary to the fruit turning stage (28 DAA) and then decreased until the fruits were fully ripe.

To investigate the role of CaMYB101 in anthocyanin biosynthesis in pepper fruit, three repeated virus-induced gene silencing (VIGS) experiments were performed with pTRV2::GUS constructs as control. All

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

three VIGS experiments showed the reproducible results. The pTRV2::GUS infected plants produced only green leaves (GUS-GL), while the pTRV2::CaMYB101 infected plants produced both green leaves (MYB-GL) and purple leaves (MYB-PL) which was due to anthocyanin accumulation (Figure 7A). These results suggested that CaMYB101 probably repressed anthocyanin biosynthesis in pepper. To verify the silencing efficiency, the expression of *CaMYB101* was examined in the silenced leaves. Compared with the control plants (GUS-GL), the expression of *CaMYB101* was significantly reduced in leaves of *CaMYB101* silenced plants (in both MYB-GL and MYB-PL) (Figure 7A), which indicated that *CaMYB101* was silenced successfully. Additionally, compared with MYB-GL, the relative expression level of *CaMYB101* was significantly lower in MYB-PL, which further implied the association of *CaMYB101* with anthocyanin accumulation.

Interestingly, in the early fruit developmental stages, the ovaries of pTRV2::GUS infected plants were green, while the ovaries of CaMYB101 silenced plants showed intense purple sectors due to anthocyanin accumulation (Figure 7B). However, the CaMYB101 silenced fruits that developed from purple ovaries (42 DAA) showed a faster purple discolouration than the pTRV2::GUS fruits (42 DAA) (Figure 7C). We examined the expression of *CaMYB101* in the silenced ovaries and fruits. In the green parts of *CaMYB101* silenced ovaries (MYB-GO), the expression of *CaMYB101* was also significantly reduced, compared to the green ovaries of control (GUS-GO) (Figure 7B), suggesting that *CaMYB101* was successfully silenced. However, unexpectedly, *CaMYB101* showed a significantly higher expression in the purple part of *CaMYB101* silenced ovaries (MYB-PO) than in the green part of the same ovaries (MYB-GO) and the green ovaries of control (GUS-GO). At a later fruit developmental stage (42 DAA), consistent with leaves, a significantly lower expression of *CaMYB101* was detected in pTRV2::CaMYB101 fruits than in pTRV2::GUS fruits (Figure 7C), which indicated the successful silencing of CaMYB101 in fruits.

The expression of anthocyanin biosynthetic genes was verified in ovaries of pTRV2::CaMYB101 and pTRV2::GUS infiltrated plants. The clear boundary of anthocyanin pigmentation in *CaMYB101* silenced ovaries (Figure 7B) enabled precise sampling of the purple parts (MYB-PO) and green parts (MYB-GO). The majority of tested anthocyanin biosynthetic genes were significantly upregulated upon *CaMYB101* silencing, including regulatory genes $CaMYB_A$ (*CaMYB82*), *CabHLH*, EBGs *CaCHS2*, *CaCHI2*, *CaF3H* and *CaF3'H* as well as LBGs *CaF3'5'H*, *CaDFR*, *CaANS* and *CaUFGT.* These genes all showed significantly higher expression level in MYB-PO compared to MYB-GO and/or GUS-GO (Figure 8). This further demonstrated that *CaMYB101* is an important transcriptional repressor for anthocyanin biosynthesis, which is likely to function upstream of the anthocyanin regulation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Identification and characterization of R2R3-MYBs in three *Capsicum* species

At present, the pepper (*Capsicum*. spp) genome sequences have been released for *C. annuum*, *C. baccatum* and *C. chinense* (Kim et al., 2014; Qin et al., 2014; Kim et al., 2017). In plants, complete and accurate gene annotation enables evolutionary and functional studies of gene families. However, different gene families are not completely annotated in pepper genus. The MYB family is one of the largest transcription factor families, with the R2R3-MYBs as the main subfamily (Stracke et al., 2001). Nevertheless, there is no comprehensive comparison of R2R3-MYBs across different pepper genomes and most of their functions are unclear. As one of the most cultivated species among the *Capsicum* genus, 108 and 116 *C. annuum* R2R3-MYBs have been genome-widely analyzed based on the Zunla 1.0 genome sequence by Wang et al. (2020) and Arce-Rodríguez et al. (2021), respectively. The reason for the disparity in different number of R2R3-MYBs could be that other known R2R3-MYB plant proteins were also used for identification in Arce-Rodríguez et al. (2021) work. The 108 CaR2R3-MYBs that Wang et al. (2020) identified in *C. annuum* were used for further analysis in this study, together with 106 CbR2R3-MYBs and 110 CcR2R3-MYBs we identified in *C. baccatum* and *C. chinense*, respectively. The number of R2R3-MYBs we identified is similar throughout three *Capsicum* spp. and *Solanum tuberosum* L. (111) (Li et al., 2020), lower than *Solanum lycopersicum* (121) (Zhao et al., 2014) and *Arabidopsis thaliana* (126) (Stracke et al., 2001), and higher than in *Solanum melongena* L. (73) (Wang et al., 2016) and *Oryza sativa* (102) (Yanhui et al., 2006). This suggests that the R2R3-MYBs in different plants evolved at different degrees, indicating that the species with closer evolutionary relationships have more similar number of R2R3-MYBs.

The genetic conservation, divergence and gene duplication cases of all R2R3-MYBs were studied throughout *Capsicum* spp. including among species (interspecies) and within species (intraspecies). The R2R3-MYB proteins of three *Capsicum* spp. and *Arabidopsis* shared highly conserved sequences within the R2 and R3 MYB domains based on amino acid frequencies (Figure 1), which confirmed the conserved nature of pepper MYB domains. This is consistent with the previous studies of R2R3-MYBs in Arabidopsis (Dubos et al., 2010). The R2 domain has three conserved tryptophans and R3 domain has two, where the first tryptophan (missing in R3) was always replaced by a hydrophobic amino acid. These results are consistent with studies in other *Solanaceous* members such as tomato (Zhao et al., 2014), eggplant (Wang et al., 2016) and potato (Li et al., 2020). Tyrosine residues showed an essential role in both R2 and R3 DNA binding domains, which could be as recognition helices that bind to the specific DNA sequence motif for different gene functions (Dubos et al., 2010).

Even though the pepper R2R3-MYBs were distributed unevenly on 12 chromosomes within each genome, the overall R2R3-MYB distribution of *C. annuum* and *C. chinense* was more similar compared to *C. baccatum*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 2). Synteny analysis showed the essential role for the expansion of R2R3-MYBs in *Capsicum* spp. Overall, 119 pairs of syntenic paralogs and orthologs were identified through *Capsicum* spp. Among different species, *C. annuum* and *C. chinense* had more duplicated gene pairs (44 pairs) than *C. annuum* and *C. baccatum* (30 pairs) or *C. baccatum* and *C. chinense* (21 pairs) (Figure 3, Supplemental Table S4). Both chromosomal distribution and syntenic relationship suggested *C. annuum* and *C. chinense* were evolutionarily closer, which agreed with Kim et al. (2017) who reported that divergence first occurred between *C. baccatum* and the progenitor of *C. annuum* and *C. chinense*. Gene duplication plays an important role in the expansion of the R2R3-MYB family. In addition, the Ka/Ks ratio of over 96 % R2R3-MYBs duplication gene pairs is smaller than 1.0, which indicates that the R2R3-MYB subfamily underwent conservative evolution.

Characterization of pepper R2R3-MYB repressors

Previous research shows that the R2R3-MYB subfamily plays an important role in anthocyanin synthesis. For example in pepper, an anthocyanin activator, $CaMYB_A$, has been functionally identified (Borovsky et al., 2004; Zhang et al., 2015). However, no anthocyanin MYB repressor has been identified yet in pepper. To investigate the transcriptional features of pepper R2R3-MYBs, we phylogeneticly analysed *Arabidopsis* R2R3-MYBs and anthocyanin related R2R3-MYB activators and repressors (Table 3, Supplemental Figure S1). The R2R3-MYBs through *C. annuum*, *C. baccatum* and *C. chinense* were clustered tightly. Meanwhile, some R2R3-MYBs were clustered with anthocyanin-related R2R3-MYBs. For example, CaMYB101, CbMYB89 and CcMYB92, were within one cluster together with FaMYB1, FcMYB1, MtMYB2, PhMYB27 and SlMYBL2, suggesting that they have similar functions (Figure 4A, Supplemental Figure 1).

The EAR motif is the most predominant repression motif with different sequence patterns, i.e. LxLxL and DLNxxP, was used in repressors identification in plants (Kagale and Rozwadowski, 2011). For further analysis, 53 R2R3-MYBs through *Capsicum* spp. were identified as repressors by containing at least one EAR motif. In order to have a better understanding of the identified R2R3-MYBs with EAR motifs in this study, a phylogenetic analysis was performed on these 53 candidate repressors and anthocyanin R2R3-MYB activators and repressors (Table 5A). It indicated that CaMYB101, CbMYB89 and CcMYB92 were closely clustered with repressor PhMYB27, SlMYBL2, MtMYB2, FaMYB1 and FcMYB1; CaMYB16, CbMYB13 and CcMYB16 were clusters together with repressor MdMYB6; CcMYB91 was clustered together with repressor AtMYB60 from *Arabidopsis*, indicating their potential function as anthocyanin repressors. Based on this phylogenetic analysis, it was also clear that most candidate repressors had three sets of paralogs in three pepper genomes, therefore, we focused on CaR2R3-MYB repressors since *C. annuum* is the most widely cultivated species through *Capsicum* spp. These 17 CaR2R3-MYB repressors exhibited varied transcript profiles in different tissues according to the public *C. annuum* (Zunla v1.0) RNA-seq data (Figure 5) (Qin et al., 2014). The relative transcript abundance of *CaMYB37*, *CaMYB45*, *CaMYB67*, *CaMYB70*, *CaMYB72*, *CaMYB91*, *CaMYB92* and *CaMYB101* were dependent on the fruit development stage. In purple pepper fruits

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cv. Tequila, anthocyanin accumulation is mainly initiated at the fruit early stage (10 DAA) and would degrade during ripening (Liu et al., 2018). Over 17 CaR2R3-MYB repressors, CaMYB101 showed high relative-expression levels at the breaker stage(when fruitwas turning red). Meanwhile, the phylogenetic analysis revealed that the CaMYB101 and published R2R3-MYB repressors, petunia PhMYB27 and tomato SlMYBL2, are within the same cluster. In this case, CaMYB101 was selected for further functional study.

CaMYB101 represses anthocyanin biosynthesis via a negative-feedback loop?

The termination of anthocyanin biosynthesis in purple pepper fruits upon ripening provides us with an opportunity to investigate anthocyanin biosynthesis repression mechanisms. The relative gene expression level of *CaMYB101* in different tissues of cv. Tequila confirmed its activity in fruits (Figure 6). In addition, the expression of *CaMYB101* was strongly associated with anthocyanin accumulation in fruits, which agreed with the expression of *TrMYB133*, an R2R3-MYB repressor for anthocyanin biosynthesis in red forage legumes (Albert, 2015). In addition, the *TrMYB133* as well as *PhMYB27*, the petunia ortholog of *CaMYB101* (Figure 4), both provide feedback repression for anthocyanin biosynthesis via activation by the MBW complexes containing anthocyanin R2R3-MYB activators, and this feedback repression might be conserved across eudicots (Albert et al., 2014; Albert, 2015). Notably, *CaMYB101* had a significantly higher expression level in leaf and stem of cv. Tequila than in fruits, which was different from the RNA-seq profile (Figure 5 & 8). This could be due to the fact that Qin et al., (2014) used a non-purple cultivar, i.e. *C. annuum* cv. Zunla for RNA-seq.

Anthocyanin R2R3-MYB repressors suppress anthocyanin biosynthesis; by silencing the repressors, anthocyanin accumulationis expected. Through inoculation of pTRV2::CaMYB101 in cv. Tequila at the cotyledon stage, we found anthocyanin accumulation in both ovaries and leaves, resulting in ovaries and leaves being partially purple and partially green (Figure 7A and 8B). However, VIGS of *CaMYB101* did not result in a more purple fruit phenotype in the later stages of fruit ripening, instead it accelerated fruit purple discoloration (Figure 7C). At the same time, *CaMYB101* showed a lower expression in leaves and fruits of pTRV2::CaMYB101 plants compared to pTRV2::GUS plants, confirming the successful silencing of *CaMYB101* in pTRV2::CaMYB101 plants (Figure 7A and 8B). R2R3-MYB repressors suppress anthocyanin biosynthesis via two approaches. On the one hand, R2R3-MYB repressors inhibit the assembling of functional MBW complexes by competing with MYB activators to bind to bHLH transcription factor (LaFountain and Yuan, 2021). On the other hand, they suppress the transcription of anthocyanin structural genes through repressive motifs. CaMYB101 has a bHLH binding domain and an EAR motif (Figure 4), suggesting it may have both repressive abilities.

In purple areas of pTRV2::CaMYB101 ovaries, the anthocyanin regulator genes, including R2R3-MYB activators $CaMYB_A$, and most structural genes were activated and showed higher transcript levels than in green parts of the pTRV2::CaMYB101 and pTRV2::GUS ovaries. (Figure 8A-C). Unexpectedly, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression level of *CaMYB101* in purple areas of pTRV2::CaMYB101 ovaries was also significantly higher compared to green parts of pTRV2::CaMYB101 ovaries, as well as in green pTRV2::GUS ovaries (Figure 7B). The high expression level of *CaMYB101* in the purple area of pTRV2::CaMYB101 ovaries is in line with a study that also reported a high expression level of its tomato orthologous gene *SlMYBL2* in purple fruits of transgenic tomato lines overexpressing *BrTT8* (Zhang et al., 2019). *SlMYBL2* was transcriptionally activated to counterbalance the active transcription of the MBW complex to prevent excessive anthocyanin synthesis. Zhang et al. (2019) proposed that an excess of anthocyanin synthesis could act as a signal to activate the expression of *SlMYBL2*. Furthermore, *AtMYB4*, *PhMYB27* and *Tr-MYB133* are the direct target of MBW complexes containing anthocyanin R2R3-MYB activators and they in turn provide negative feedback regulation to MBW complexes (Jin et al., 2000; Albert et al., 2011; Albert, 2015). We hypothesize that anthocyanin accumulation in the purple ovary is attributed to the initial silencing of *CaMYB101* by VIGS. We further hypothesise that either high anthocyanin levels or high transcription levels of the MBW complex acts as a signal to activate *CaMYB101* expression to avoid overproduction of anthocyanins in a negative-feedback loop. It also indicates that the VIGS effect that should be still present in the silenced tissue is not enough to cope with that feedback induction effect.

Even though VIGS of *CaMYB101* caused anthocyanin accumulation in both pepper leaves and ovaries, feedback regulation seemed not to occur in leaves (Figure 7A), suggesting that different molecular mechanisms of anthocyanin pigmentation may exist in vegetative tissues and in ovaries. However, why silencing of *CaMYB101* accelerated anthocyanin discolouration in fruits is unclear (Figure 7C) and this needs further research.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusion

The present study provide a fundamental layout for pepper R2R3-MYB gene family by conducting a comprehensive global genome analysis throughout three *Capsicum.*spp., with specific focus on R2R3-MYB repressors. Functionaly analysis of *CaMYB101* showed that *CaMYB101* repressed anthocynain accumulation in pepper leaves and ovaies. Our study sheds new light on the negative regulation of anthocyanin biosynthesis in pepper. Moreover, our results provide potential genetic engineering approach to alter anthocyanin accumulation in this important horticultural crop.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Plant material

Bell pepper (*Capsicum annuum*) cv. Tequila (Enza Zaden) was grown under standard greenhouse conditions in Wageningen, the Netherlands. Nine plants were grown for determining gene expression profile of *CaMYB101* in different plant tissues. Plant leaf, stem, seed, ovary, flower and fruits at day after anthesis (DAA) of 5, 7, 21, 28, 42 and 56 were sampled for tissue specific expression. Tissues and fruits from each three plants were pooled together as one biological replicate.

Identification of the R2R3-MYB subfamily genes in *Capsicum* spp.

Multiple *de novo* pepper genome sequences of *Capsicum baccatum* and *Capsicum chinense* were downloaded from the Pepper Genome Platform (http://peppergenome.snu.ac.kr/). A Hidden Markov Models (HMM) profile of MYB DNA-binding domain (PF00249) was downloaded from Pfam database (http://pfam.xfam.org/). It was used as a query to search the *C. baccatum* and *C. chinense* genomes to identify all MYB containing sequences with an E-values < $1e^{-3}$. All candidate protein sequences were examined using the NCBI conserved domain database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and ExPASy (https://web.expasy.org/protparam/) to verify the presence of R2 and R3 domains. The R2R3-MYB genes of *C. annuum* used in this study were identified by Wang et al. (2020). The pepper R2R3-MYBs were named based on their species, namely CaMYB for *C. annuum*, CbMYB for *C. baccatum* and CcMYB for *C. chinense*, and numbered based on chromosomal orders (chr1 – chr12, chr0). The amino acid sequences of R2 and R3 MYB repeats of CaMYBs, CbMYBs and CcMYBs were aligned with ClustalW (MEGA-7) and manually adjusted referring to *Arabidopsis* (Dubos et al., 2010).The sequence logos of R2 and R3 MYB repeats were constructed by WebLogo (Crooks et al., 2004). The chromosomal distribution of pepper R2R3-MYBs was mapped using Mapchart 2.2 software (Voorrips, 2002).

Gene structure motif and phylogenetic analysis

The MEME v5.1.0 online tool was used for the conserved domain investigation. The map of exon-intron structure of *Capsicum* spp. R2R3-MYBs was constructed by TBtools software using coding sequences with the corresponding protein sequences (Chen et al., 2020). The complete protein sequences of *Capsicum* spp. R2R3-MYBs were aligned by ClusterW method then the phylogenetic analysis of R2R3-MYBs was constructed by NJ method with bootstrap test 1000 replicates using MEGA6.0.

Collinearity/Synteny analyses

The Capsicum spp. genome sequences were downloaded from Pepper Genome Platform (http://peppergenome.snu.ac.kr/). Furthermore, a Multiple Collinearity Scan toolkit (reference) was used to

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

investigate the syntenic relationship with E-value < 1 e $^{-10}$ between *Capsicum* spp. R2R3-MYBs sequences. TBtools software was used for the calculation of synonymous (Ks) and non-synonymous (Ka) value.

Identification of repression motif-containing R2R3-MYB proteins

Repression motifs, included EAR motif (LxLxL, DLNxxP), TLLLFR motif (TLLLFR), R/KLFGV motif (R/KLFGV) and LxLxPP motif (LxLxPP) were screened across all R2R3-MYBs in three pepper genomes (Kagale and Rozwadowski, 2011).

Expression Profiles of CaMYB Genes Based on RNA-Seq

The expression levels of R2R3-MYBs in *Capsicum annuum* cv. Zunla have been investigated in the different growth development. The RNA-seq atlas was downloaded from China National GeneBank (https://db.cngb.org/search/project/CNPhis0000547/) (Qin et al., 2014). Six different organs including root, stem, leaf, bud, flower and fruits were used for analysis. Furthermore, the relative gene expression level of R2R3=-MYBs were calculated by Reads Per Kilo bases per Million (RPKM). The hierarchical clustering was carried out with Pearson's correlation distance and the gene clusters were clustered by gene expression.

VIGS and agrobacteria inoculation

The specific region of CaMYB101 (Capana00g002497) used for silencing was selected by SGN VIGS Tool (Fernandez-Pozo et al., 2015). A 223 bp fragment of CaMYB101 was amplified within the specific region using primers vigsCaMYB101-For (5'- CACCAGGATCTTGGTCTAAACAAGAAGA -3') and vigsCaMYB101-Rev (5'- CCTATTGCCAAGAAGAGCAT -3') from cv. Tequila cDNA to design VIGS construct. The fragment was cloned into pTRV2 vector (Liu et al., 2002), then the pTRV2 vectors carrying specific CaMYB101 fragment were transformed into *Agrobacterium tumefaciens* strain GV3101. *A. tumefaciens* culture containing pTRV1, pTRV2::CaMYB101, pTRV2::PDS or pTRV2::GUS was prepared for VIGS experiment as described by Romero et al. (2011). The pTRV1 culture was mixed with pTRV2 culture in a 1:1 ratio. Then TRV infection was done through mixed culture infiltration on the abaxial of cotyledon of 100 three-week-old cv. Tequila seedlings using syringes without a needle. Among them, 45 plants were inoculated with pTRV2::CaMYB101, 45 plants were inoculated with pTRV2::GUS and 10 plants were inoculated with pTRV2::PDS.

Real-Time Quantitative PCR

The qPCR reaction was performed by using SYBR Green Master Mix then detected by BioRad Real-Time PCR System. The delta-delta-Ct was used for expression analysis. The primers used in this paper are listed in Supplemental Table S5.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author contribution

YL defined the research question, proposed the methodology and the experimental design. YL and YW carried out the experiments and analysed the experimental data together. ZZ, ZW, SZ and XL performed the bioinformatic analysis. YL and YW wrote the first draft, ZZ revised it. YT, RV, LM, RS and AB provided comments for the final version. All authors have approved the manuscript. This paper has not been accepted or published elsewhere.

Funding

The authors are grateful for the financial support provided by China Scholarship Council (CSC) under grant number 201607720005 and 201707720017.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference

Aguilar-Barragán A, Ochoa-Alejo N (2014) Virus-induced silencing of MYB and WD40 transcription factor genes affects the accumulation of anthocyanins in chilli pepper fruit. Biologia Plantarum **58:** 567-574

Aharoni A, De Vos CHR, Wein M, Sun Z, Greco R, Kroon A, Mol JNM, O'Connell AP (2001) The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant Journal **28:** 319-332

Albert NW (2015) Subspecialization of R2R3-MYB repressors for anthocyanin and proanthocyanidin regulation in forage legumes. Frontiers in Plant Science **6**

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE (2014) A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. The Plant Cell **26:** 962-980

Albert NW, Lewis DH, Zhang H, Schwinn KE, Jameson PE, Davies KM (2011) Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. Plant J **65:** 771-784

Arce-Rodríguez ML, Martínez O, Ochoa-Alejo N (2021) Genome-wide identification and analysis of the myb transcription factor gene family in chili pepper (Capsicum spp.). International Journal of Molecular Sciences **22:** 1-24

Bond DM, Albert NW, Lee RH, Gillard GB, Brown CM, Hellens RP, Macknight RC (2016) Infiltration-RNAseq: TRANSCRIPTOME profiling of Agrobacterium-mediated infiltration of transcription factors to discover gene function and expression networks in plants. Plant Methods **12:** 1

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12:** 2383-2393

Borovsky Y, Oren-Shamir M, Ovadia R, De Jong W, Paran I (2004) The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of Petunia. Theoretical and Applied Genetics **109:** 23-29

Cavallini E, Zenoni S, Finezzo L, Guzzo F, Zamboni A, Avesani L, Tornielli GB (2014) Functional diversification of grapevine MYB5a and MYB5b in the control of flavonoid biosynthesis in a petunia anthocyanin regulatory mutant. Plant and Cell Physiology **55:** 517-534

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Molecular Plant **13:** 1194-1202

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. Genome Research **14**: 1188-1190

Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde JP, Mérillon JM, Hamdi S (2006) Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiology **140:** 499-511

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, Robinson SP, Barrieu F (2008) The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol **147:** 2041-2053

Du H, Liang Z, Zhao S, Nan MG, Tran LSP, Lu K, Huang YB, Li JN (2015) The evolutionary history of R2R3-myb proteins across 50 eukaryotes: New insights into subfamily classification and expansion. Scientific Reports **5**

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. Trends in Plant Science **15:** 573-581

Elomaa P, Uimari A, Mehto M, Albert VA, Laitinen RAE, Teeri TH (2003) Activation of Anthocyanin Biosynthesis in Gerbera hybrida (Asteraceae) Suggests Conserved Protein-Protein and Protein-Promoter Interactions between the Anciently Diverged Monocots and Eudicots. Plant Physiology **133:** 1831-1842

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007) Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant Journal **49:** 414-427

Feng S, Wang Y, Yang S, Xu Y, Chen X (2010) Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. Planta **232:** 245-255

Fuda NJ, Ardehali MB, Lis JT (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature **461:** 186-192

Gao JJ, Shen XF, Zhang Z, Peng RH, Xiong AS, Xu J, Zhu B, Zheng JL, Yao QH (2011) The myb transcription factor MdMYB6 suppresses anthocyanin biosynthesis in transgenic Arabidopsis. Plant Cell, Tissue and Organ Culture **106:** 235-242

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant Journal **53:** 814-827

Heppel SC, Jaffé FW, Takos AM, Schellmann S, Rausch T, Walker AR, Bogs J (2013) Identification of key amino acids for the evolution of promoter target specificity of anthocyanin and proanthocyanidin regulating MYB factors. Plant Molecular Biology **82:** 457-471

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C (2000) Trancriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO Journal **19:** 6150-6161

Jun JH, Liu C, Xiao X, Dixon RA (2015) The transcriptional repressor MYB2 regulates both spatial and temporal patterns of proanthocyandin and anthocyanin pigmentation in Medicago truncatula. The Plant Cell **27:** 2860-2879

Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, De Jong WS (2009) The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. Theoretical and Applied Genetics **120:** 45-57

Kagale S, Rozwadowski K (2011) EAR motif-mediated transcriptional repression in plants: An underlying mechanism for epigenetic regulation of gene expression. Epigenetics **6:** 141-146

Kiferle C, Fantini E, Bassolino L, Povero G, Spelt C, Buti S, Giuliano G, Quattrocchio F, Koes R, Perata P, Gonzali S (2015)

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tomato R2R3-MYB Proteins SlANT1 and SlAN2: Same Protein Activity, Different Roles. Plos One **10**

Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, Kim SB, Han K, Lee J, Park M, Lee HA, Lee HY, Lee Y, Oh S, Lee JH, Choi E, Choi E, Lee SE, Jeon J, Kim H, Choi G, Song H, Lee JK, Lee SC, Kwon JK, Lee HY, Koo N, Hong Y, Kim RW, Kang WH, Huh JH, Kang BC, Yang TJ, Lee YH, Bennetzen JL, Choi D (2017) New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biology **18**

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 JKC, Sørensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nature Genetics **46:** 270-278

Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C, Smeekens S, Tonelli C, Paz-Ares J, Weisshaar B (1998) Towards functional characterisation of the members of the R2R3-MYB gene family from Arabidopsis thaliana. Plant Journal **16:** 263-276

LaFountain AM, Yuan YW (2021) Repressors of anthocyanin biosynthesis. New Phytologist **231:** 933-949

Li Y, Lin-Wang K, Liu Z, Allan AC, Qin S, Zhang J, Liu Y (2020) Genome-wide analysis and expression profiles of the StR2R3-MYB transcription factor superfamily in potato (Solanum tuberosum L.). International Journal of Biological Macromolecules **148:** 817-832

Li Z, Peng R, Tian Y, Han H, Xu J, Yao Q (2016) Genome-wide identification and analysis of the MYB transcription factor superfamily in solanum lycopersicum. Plant and Cell Physiology **57:** 1657-1677

Lin-Wang K, McGhie TK, Wang M, Liu Y, Warren B, Storey R, Espley RV, Allan AC (2014) Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca). Frontiers in Plant Science **5**

Liu Y, Lv J, Liu Z, Wang J, Yang B, Chen W, Ou L, Dai X, Zhang Z, Zou X (2020) Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (Capsicum annuum L.). Food Chemistry **306:** 125629

Liu Y, Tikunov Y, Schouten RE, Marcelis LFM, Visser RGF, Bovy A (2018) Anthocyanin Biosynthesis and Degradation Mechanisms in Solanaceous Vegetables: A Review. Frontiers in Chemistry **6**

Liu Y, Wang L, Zhang J, Yu B, Wang J, Wang D (2017) The MYB transcription factor StMYBA1 from potato requires light to activate anthocyanin biosynthesis in transgenic tobacco. Journal of Plant Biology **60:** 93-101

Liu YH, Lin-Wang K, Espley RV, Wang L, Yang HY, Yu B, Dare A, Varkonyi-Gasic E, Wang J, Zhang JL, Wang D, Allan AC (2016) Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. Journal of Experimental Botany **67:** 2159-2176

Ma D, Constabel CP (2019) MYB Repressors as Regulators of Phenylpropanoid Metabolism in Plants. Trends in Plant Science **24:** 275-289

Mano H, Ogasawara F, Sato K, Higo H, Minobe Y (2007) Isolation of a regulatory gene of anthocyanin biosynthesis in tuberous roots of purple-fleshed sweet potato. Plant Physiology **143:** 1252-1268

Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, Wagner RD (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell **15:** 1689-1703

Mur LA (1995) Characterization of members of the MYB gene family of transcription factors from Petunia Hybrida. [S.l. : s.n.]

Nakatsuka T, Haruta KS, Pitaksutheepong C, Abe Y, Kakizaki Y, Yamamoto K, Shimada N, Yamamura S, Nishihara M (2008) Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. Plant and Cell Physiology **49:** 1818-1829

Ogata K, Hojo H, Aimoto S, Nakai T, Nakamura H, Sarai A, Ishii S, Nishimura Y (1992) Solution structure of a DNA-binding unit of Myb: A helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proceedings of the National Academy of Sciences of the United States of America **89:** 6428-6432

Park JS, Kim JB, Cho KJ, Cheon CI, Sung MK, Choung MG, Roh KH (2008) Arabidopsis R2R3-MYB transcription factor AtMYB60 functions as a transcriptional repressor of anthocyanin biosynthesis in lettuce (Lactuca sativa). Plant Cell Rep **27:** 985-994

Pattanaik S, Kong Q, Zaitlin D, Werkman JR, Xie CH, Patra B, Yuan L (2010) Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. Planta **231:** 1061-1076

Peel GJ, Pang Y, Modolo LV, Dixon RA (2009) The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in Medicago. Plant Journal **59:** 136-149

Perez-Diaz JR, Perez-Diaz J, Madrid-Espinoza J, Gonzalez-Villanueva E, Moreno Y, Ruiz-Lara S (2016) New member of the R2R3-MYB transcription factors family in grapevine suppresses the anthocyanin accumulation in the flowers of transgenic tobacco. Plant Molecular Biology **90:** 63-76

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, Changying LH, 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 (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proceedings of the National Academy of Sciences of the United States of America **111:** 5135-5140

Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R (1999) Molecular analysis of the anthocyanin2

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene of petunia and its role in the evolution of flower color. Plant Cell **11:** 1433-1444

Salvatierra A, Pimentel P, Moya-León MA, Herrera R (2013) Increased accumulation of anthocyanins in Fragaria chiloensis fruits by transient suppression of FcMYB1 gene. Phytochemistry **90:** 25-36

Sapir M, Oren-Shamir M, Ovadia R, Reuveni M, Evenor D, Tadmor Y, Nahon S, Shlomo H, Chen L, Meir A, Levin I (2008) Molecular aspects of Anthocyanin fruit tomato in relation to high pigment-1. J Hered **99:** 292-303

Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E, Oyama R, Bailey P, Davies K, Martin C (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the Genus antirrhinum. Plant Cell **18:** 831-851

Shen X, Zhao K, Liu L, Zhang K, Yuan H, Liao X, Wang Q, Guo X, Li F, Li T (2014) A role for PacMYBA in ABA-regulated anthocyanin biosynthesis in red-colored sweet cherry cv. Hong Deng (Prunus avium L.). Plant and Cell Physiology **55:** 862-880

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling. Plant Journal **50:** 660-677

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Current Opinion in Plant Biology **4:** 447-456

Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR (2006) Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiology **142:** 1216-1232

Vimolmangkang S, Han Y, Wei G, Korban SS (2013) An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. BMC Plant Biology **13**

Voorrips RE (2002) Mapchart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity **93:** 77-78

Wang J, Liu Y, Tang B, Dai X, Xie L, Liu F, Zou X (2020) Genome-Wide Identification and Capsaicinoid Biosynthesis-Related Expression Analysis of the R2R3-MYB Gene Family in Capsicum annuum L. Frontiers in Genetics **11**

Wang S, Chen Z, Ji T, Di Q, Li L, Wang X, Wei M, Shi Q, Li Y, Gong B, Yang F (2016) Genome-wide identification and characterization of the R2R3MYB transcription factor superfamily in eggplant (Solanum melongena L.). Agri Gene **2:** 38-52

Xu WJ, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends in Plant Science **20:** 176-185

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of Arabidopsis: Expression analysis and phylogenetic comparison with the rice MYB family. Plant Molecular Biology **60:** 107-124

Zhang Y, Li Y, Li W, Hu Z, Yu X, Tu Y, Zhang M, Huang J, Chen G (2019) Metabolic and molecular analysis of nonuniform anthocyanin pigmentation in tomato fruit under high light. Horticulture Research **6:** 56

Zhang Z, Li DW, Jin JH, Yin YX, Zhang HX, Chai WG, Gong ZH (2015) VIGS approach reveals the modulation of anthocyanin biosynthetic genes by CaMYB in chili pepper leaves. Front Plant Sci **6:** 500

Zhao P, Li Q, Li J, Wang L, Ren Z (2014) Genome-wide identification and characterization of R2R3MYB family in Solanum lycopersicum. Molecular Genetics and Genomics **289:** 1183-1207

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Assembly descriptions and sources for three genomes used to characterize R2R3-MYB diversity in *Capsicum* spp.

Table 2. Duplication analysis of the R2R3-MYBs within *Capsicum annuum*, *Capsicum baccatum* and *Capsicum chinense*, respectively.

Table 3. Plant R2R3-MYB family genes involved in anthocyanin biosynthesis.

Table 4. Capsicum spp. R2R3-MYB repressors with conserved repressor motifs.

Table 1. Assembly descriptions and sources for three genomes used to characterize R2R3-MYB diversity in *Capsicum* spp.

Genome features	*Capsicum annuum* [1]	*Capsicum baccatum* [2]	*Capsicum chinense* [2]
Assembled genome size (Gb)	2.9	3.2	3.0
Number of Scaffolds	6478	2083	1557
Scaffold N50 (Mb)	1.4	2.0 Mb	3.3 Mb
Coverage	99X	98X	80X

[1] Pepper Zunla 1 Ref_v1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCA_000710875.1);

[2] Kim et al. (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Duplication analysis of the R2R3-MYBs within *Capsicum annuum*, *Capsicum baccatum* and *Capsicum chinense*, respectively. The minimum block size for collinear analysis is set at 5 in MCSanX for intraspecies analysis.

Sequence1	Sequence 2	Ka[1]	Ks[2]	Ka/Ks	Effective Len[3]	Average S-sites[4]	Average N-sites[5]
CaMYB12	CaMYB19	0.20	0.84	0.24	858	186.67	671.33
CaMYB13	CaMYB20	0.13	0.63	0.20	585	120.67	464.33
CaMYB19	CaMYB24	0.34	2.54	0.13	864	190.33	673.67
CaMYB16	CaMYB34	0.27	1.44	0.18	783	190.42	592.58
CaMYB15	CaMYB35	0.46	1.56	0.29	864	186.42	677.58
CaMYB12	CaMYB54	0.33	1.47	0.23	957	209.58	747.42
CaMYB16	CaMYB93	0.34	1.72	0.20	636	145.67	490.33
CaMYB26	CaMYB52	0.22	0.43	0.50	867	176.58	690.42
CaMYB23	CaMYB91	0.37	3.55	0.11	765	154.25	610.75
CaMYB34	CaMYB93	0.26	1.71	0.15	726	162.67	563.33
CaMYB51	CaMYB85	0.31	0.82	0.38	858	173.83	684.17
CaMYB64	CaMYB79	0.15	0.80	0.19	942	199.08	742.92
CaMYB76	CaMYB83	0.14	0.75	0.19	381	79.25	301.75
CbMYB106	CbMYB13	0.25	1.41	0.18	867	206.17	660.83
CbMYB12	CbMYB29	0.44	1.91	0.23	858	186.08	671.92
CcMYB107	CcMYB14	0.33	1.53	0.22	957	208.58	748.42
CcMYB12	CcMYB50	0.33	0.83	0.39	807	162.25	644.75
CcMYB14	CcMYB22	0.20	0.89	0.23	858	186.67	671.33
CcMYB16	CcMYB39	0.25	1.42	0.17	855	203.17	651.83
CcMYB19	CcMYB38	0.40	1.70	0.24	855	185.08	669.92
CcMYB31	CcMYB56	0.23	0.44	0.53	882	179.67	702.33
CcMYB40	CcMYB103	0.51	2.17	0.24	615	140.25	474.75
CcMYB39	CcMYB103	0.26	1.70	0.15	726	163.83	562.17
CcMYB41	CcMYB101	0.26	0.91	0.29	849	170.25	678.75

[1] Ka, non-synonymous substitution rate;
[2] Ks, synonymous substitution rate;
[3] Effective Len: effctive length of compared sequences;
[4] S-site, number of synonymous site;
[5] N-site, number of non-synonymous sites.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3. Plant R2R3-MYB family genes involved in anthocyanin biosynthesis.

Gene name	Gene ID	Species	Function (activator+/repressor-)	References
AmROSEA1	DQ275529	*Anitirrhinum majus*	+	Schwinn et al. (2006)
AmROSEA2	DQ275530	*Anitirrhinum majus*	+	Schwinn et al. (2006)
AmVenosa	DQ275531	*Anitirrhinum majus*	+	Schwinn et al. (2006)
AtPAP1/AtMYB75	AT1G56650	*Arabidopsis thaliana*	+	Borevitz et al. (2000)
AtPAP2/AtMYB90	AT1G66390	*Arabidopsis thaliana*	+	Borevitz et al. (2000)
AtPAP3/AtMYB113	AT1G66370	*Arabidopsis thaliana*	+	Gonzalez et al. (2008)
AtPAP4/AtMYB114	AT1G66380	*Arabidopsis thaliana*	+	Gonzalez et al. (2008); Heppel et al. (2013)
AtMYB60	AT1G08810	*Arabidopsis thaliana*	-	Kranz et al. (1998); Park et al. (2008)
CaMYBA	AJ608992	*Capsicum annuum*	+	Borovsky et al. (2004); Aguilar-Barragán and Ochoa-Alejo (2014)
FaMYB1	MN689833	*Fragaria x ananassa*	-	Aharoni et al. (2001)
FaMYB10	EU155162	*Fragaria vesca*	+	Lin-Wang et al. (2014)
FcMYB1	GQ867222	*Fragaria chiloensis*	-	Salvatierra et al. (2013)
GhMYB10	AJ554700	*Gerbera hybrida*	+	Elomaa et al. (2003)
GtMYB3	AB289445	*Gentiana triflora*	+	Nakatsuka et al. (2008)
IbMYB1	AB258984	*Ipomoea batatas*	+	Mano et al. (2007)
MdMYB1	DQ886414	*Malus domestica Borkh.*	+	Takos et al. (2006)
MdMYB3	JN544704	*Malus × domestica*	+	Vimolmangkang et al. (2013)
MdMYB6	DQ074461	*Malus × domestica*	-	Gao et al. (2011)
MdMYB10	EU518249	*Malus × domestica*	+	Espley et al. (2007)
MtLAP1	FJ199998	*Medicago truncatula*	+	Peel et al. (2009); Bond et al. (2016)
MtMYB2	AES99346	*Medicago truncatula*	-	Jun et al. (2015)
NtAN2	FJ472647	*Nicotiana tabacum*	+	Pattanaik et al. (2010)
PacMYBA	JN166079	*Prunus avium L*	+	Shen et al. (2014)
PhAN2	AAF66727	*Petunia hybrida*	+	Quattrocchio et al. (1999)
PhDPL	HQ116169	*Petunia hybrida*	+	Albert et al. (2011)
PhMYB27	KF985023	*Petunia hybrida*	-	Mur (1995); Albert et al. (2011)
PhPHZ	HQ116170	*Petunia hybrida*	+	Albert et al. (2011)

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PyMYB10	KF387520	*Pyrus pyrifolia*	+	Feng et al. (2010)
SlANT1^{AC}/LeANT1	AY348870	*Solanum lycopersicum cv. Ailsa Craig*	+	Mathews et al. (2003)
SlANT1^{Aft}	ABO26065	*Solanum lycopersicum.L accession LA1996*	+	Sapir et al. (2008)
SlAN2	FJ705320	*Solanum lycopersicum.L*	+	Kiferle et al. (2015)
SlMYBL2	XM_004239350	*The Pro35S:BrTT8 transgenic Solanum lycopersicum Mill. cv Ailsa Craig, AC*	-	Zhang et al. (2019)
StAN1	DQ917781	*Solanum tuberosum L.*	+	Jung et al. (2009)
StAN2	AY841131	*Solanum tuberosum L.*	+	Jung et al. (2009)
StMYBA1	KP317177	*Solanum tuberosum L.*	+	Liu et al. (2016); Liu et al. (2017)
VvMYB5b	AY899404	*Vitis vinifera*	+	Deluc et al. (2008); Cavallini et al. (2014)

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. Capsicum spp. R2R3-MYB repressors with conserved repressor motifs.

Repression motif	EAR (LxLxL, DLNxxP) motif	LxLxPP motif
Capsicum annuum	*CaMYB5, CaMYB16, CaMYB34, CaMYB45, CaMYB48, CaMYB50, CaMYB55, CaMYB67, CaMYB68, CaMYB70, CaMYB72, CaMYB82, CaMYB91, CaMYB92, CaMYB93, CaMYB101*	*CaMYB16, CaMYB37*
Capsicum baccatum	*CbMYB4, CbMYB13, CbMYB29, CbMYB46, CbMYB50, CbMYB66, CbMYB67, CbMYB69, CbMYB76, CbMYB85, CbMYB89, CbMYB92, CbMYB93, CbMYB94, CbMYB102, CbMYB106*	*CbMYB13, CbMYB32*
Capsicum chinense	*CcMYB5, CcMYB16, CcMYB39, CcMYB47, CcMYB51, CcMYB52, CcMYB61, CcMYB72, CcMYB79, CcMYB87, CcMYB91, CcMYB92, CcMYB93, CcMYB103, CcMYB104, CcMYB105, CcMYB108, CcMYB110*	*CcMYB16, CcMYB40*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

Figure 1. The conserved motifs of the R2 and R3 domains in Capsicum spp. R2R3-MYB proteins.

Figure 2. Chromosome localization of Capsicum spp. R2R3-MYB family members.

Figure 3 Synteny of the *R2R3-MYB* transcription factors in the *Capsicum* spp. genome.

Figure 4. (A)Phylogenetic tree of R2R3-MYB with repression motif and published anthocyanin-related MYBs.

Figure 5. The gene expression patterns in root, stem, leaf, bud, flower and fruits of *Capsicum annuum*.

Figure 6. Expression profile of *CaMYB101* genes in different tissues and in fruits during different stages obtained using qRT-PCR.

Figure 7. Functional analysis of candidate repressor *CaMYB101* of anthocyanin biosynthesis by VIGS.

Figure 8. The relative expression level of anthocyanin biosynthetic genes in the ovaries

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

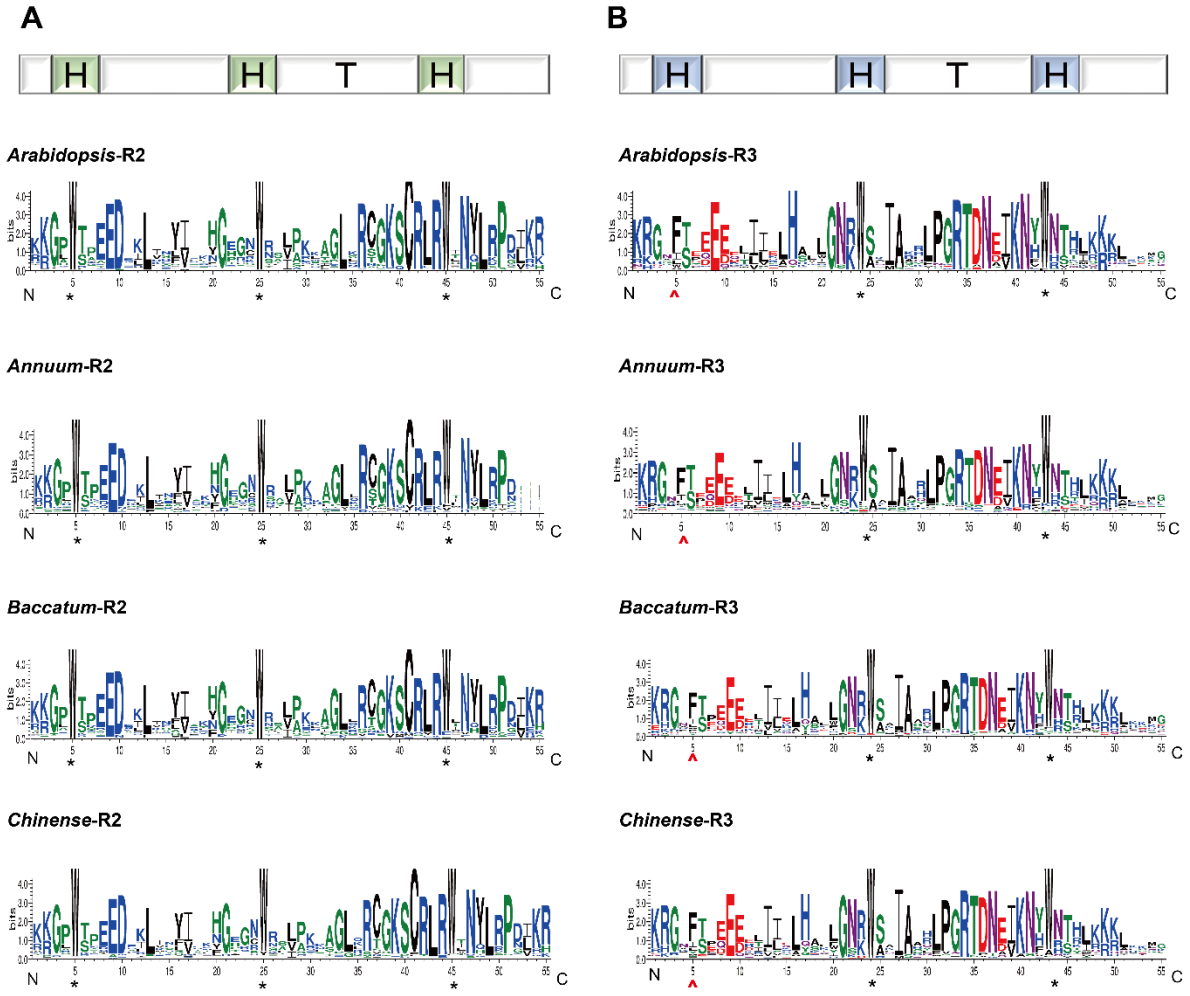

Figure 1. The conserved motifs of the R2 and R3 domains in *Capsicum* spp. R2R3-MYB proteins. (A) The R2 domain of *C. annuum, C. baccatum and C. chinense*; (B) the R3 domain of *C. annuum*, *C. baccatum* and *C. chinense.* These sequence logos were determined from the multiple alignment analysis of 108, 106 and 110 R2R3-MYB proteins in *C. annuum*, *C. baccatum* and *C. chinense* respectively. Each MYB repeat contains three α-helices (H). The second and third helices form a helix-turn-helix architecture (HTH). The bit score shows the information content for each position in the sequence. The conserved tryptophan residues (W) are marked with black asterisks and the replacement of tryptophan in the R3 repeat are marked by red circumflex accent.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

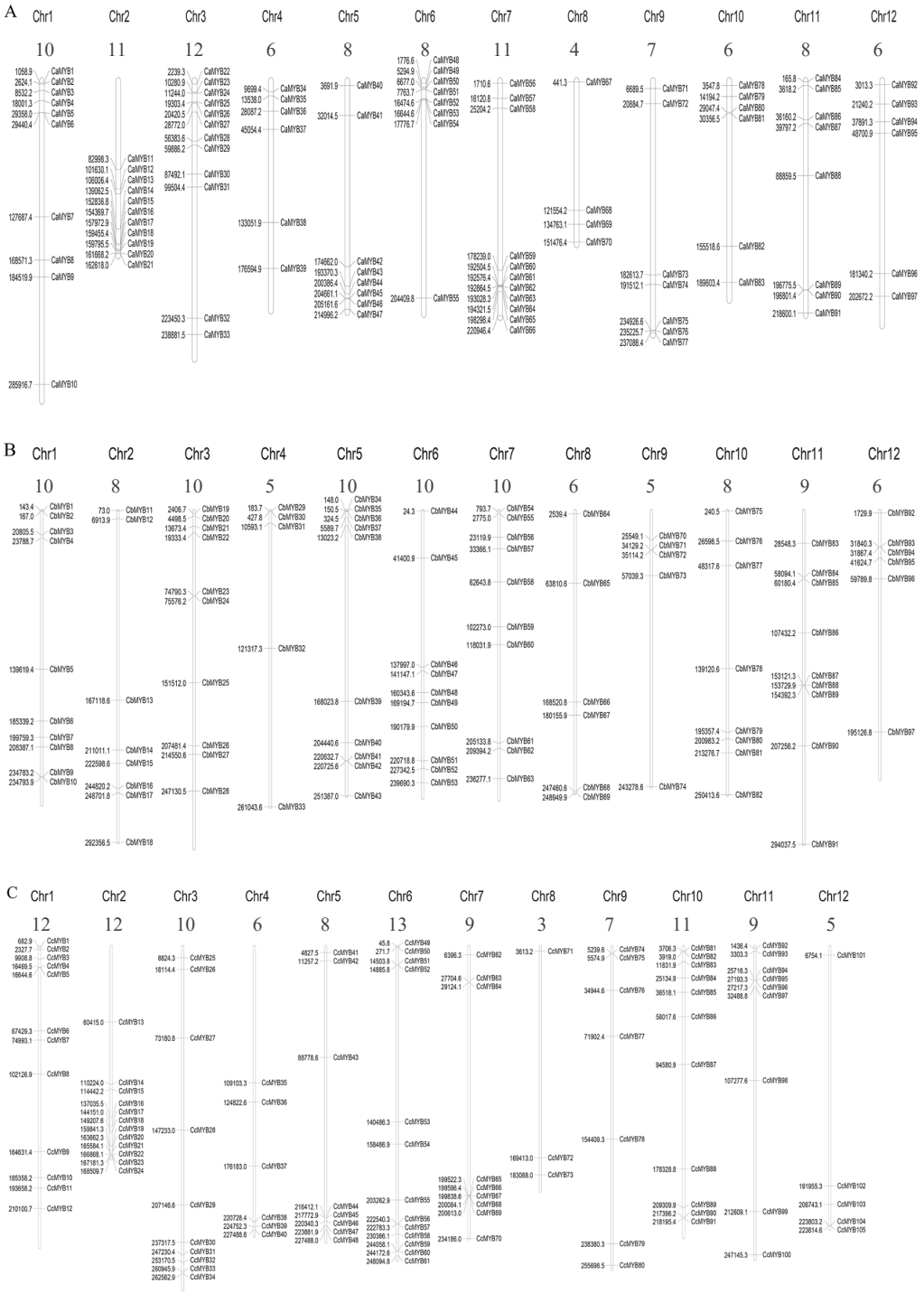

Figure 2. Chromosome localization of *Capsicum* spp. R2R3-MYB family members. The physical distribution of each R2R3-MYB gene is listed on the chromosomes of (A) *Capsicum annuum*, (B) *Capsicum baccatum* and (C) *Capsicum chinense*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

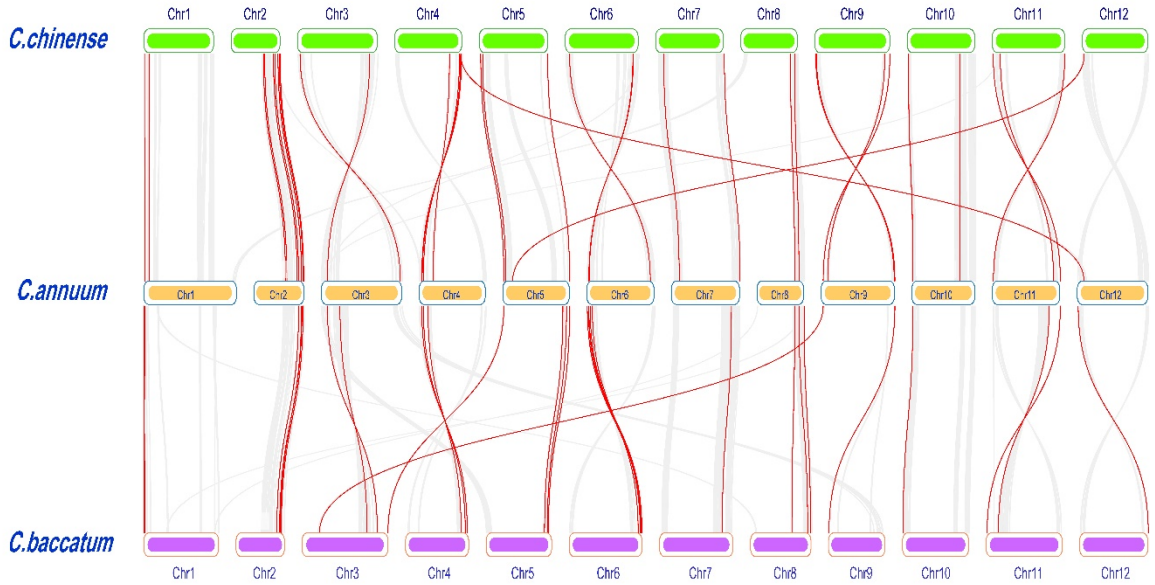

Figure 3 Synteny of the *R2R3-MYB* transcription factors across the *Capsicum* spp. genome. The *Capsicum* R2R3-MYB were mapped to the corresponding chromosomes of *C. chinense*, *C. annuum* and *C. baccatum*. Those with a syntenic relationship are joined by red lines. The gray lines indicate all syntenic blocks in the *Capsicum* spp. genome. Specific gene pairs are listed in Supplemental Table S4. The minimum block size for collinear analysis is set at 30 in MCSanX for interspecies analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

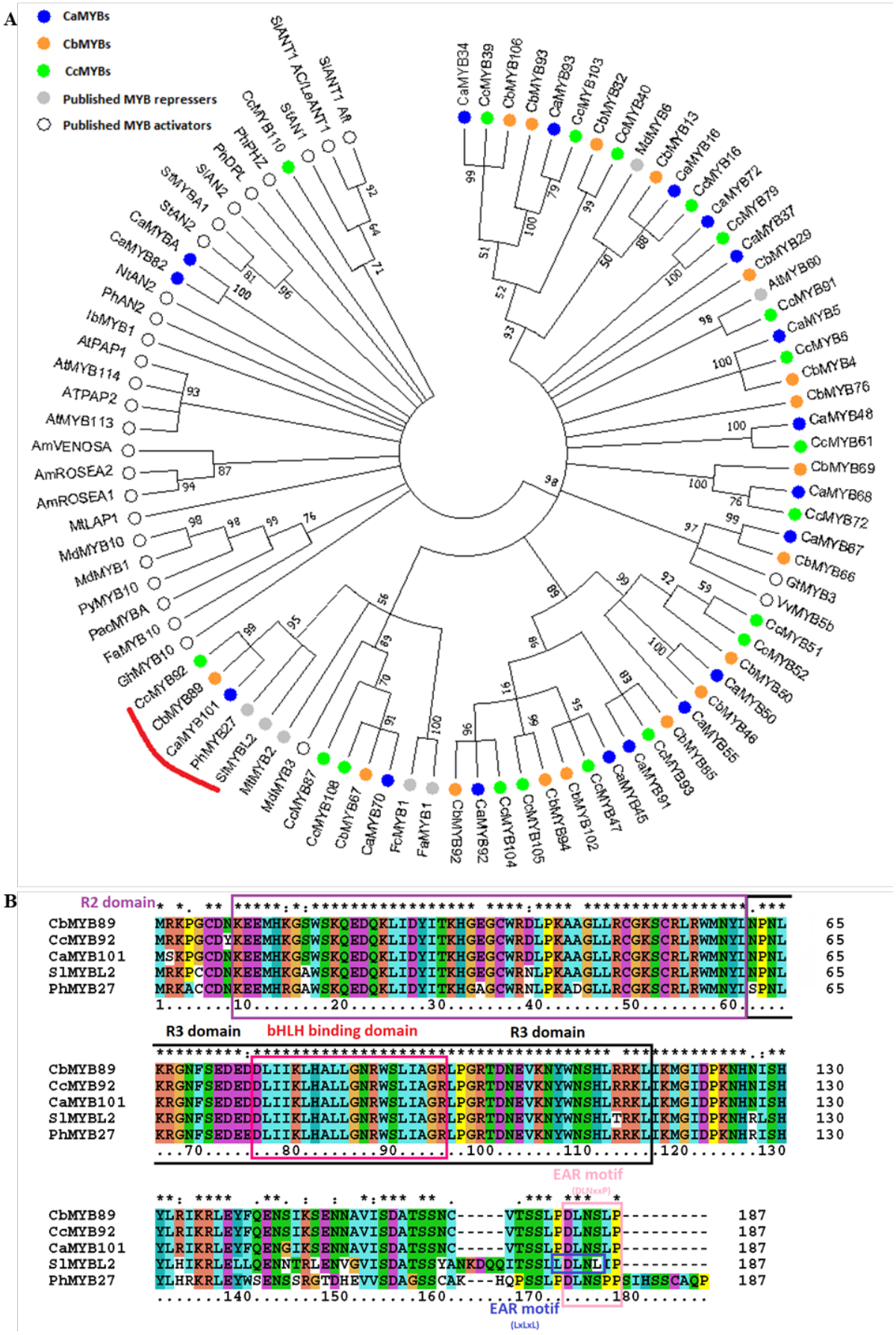

Figure 4. (A)Phylogenetic tree of R2R3-MYB with repression motif and published anthocyanin-related MYBs. CaMYB82

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and $CaMYB_A$ is the same gene. The red curve shows the cluster of CaMYB101, CbMYB89 and CcMYB89 with petunia and tomato MYB repressors. The blue, orange and green dots indicate CaMYBs, CbMYBs and CcMYBs, respectively (Table 4). The white and grey dots indicate published MYB activators and MYB repressors (Table 3). (B) Multiple alignment of predicted protein sequences of CaMYB101, CbMYB89, CcMYB92, SlMYBL2 and PhMYB27. The R2, R3 domains, EAR motifs (LxLxL and DLNxxP) and bHLH binding domain is labeled above the alignment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

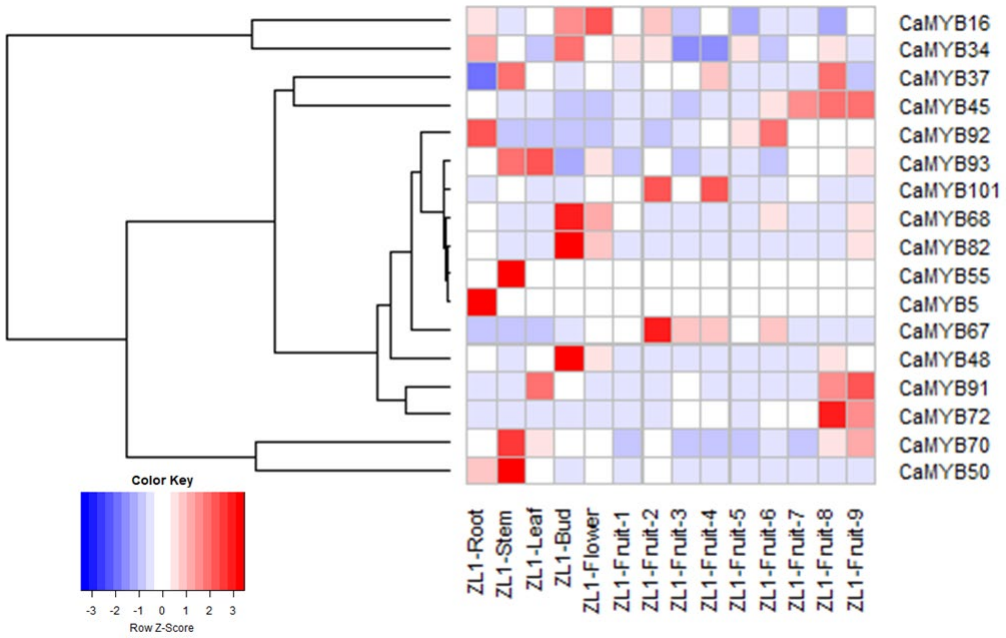

Figure 5. The gene expression patterns in root, stem, leaf, bud, flower and fruits of *Capsicum annuum* cv. Zunla. The color indicates relative gene expression in selected tissues based on rpkm (reads per kb per million mapped reads). Fruit stages 1-3 represent pre-breaker stages (1-3cm, 3-4cm, 4-5cm fruit length; mature green), fruit stages 4-6 represent the breaker stage (when the fruit was turning red) and fruit stages 7-9 represent post-breaker stages (3, 5, and 7 days after breaker).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

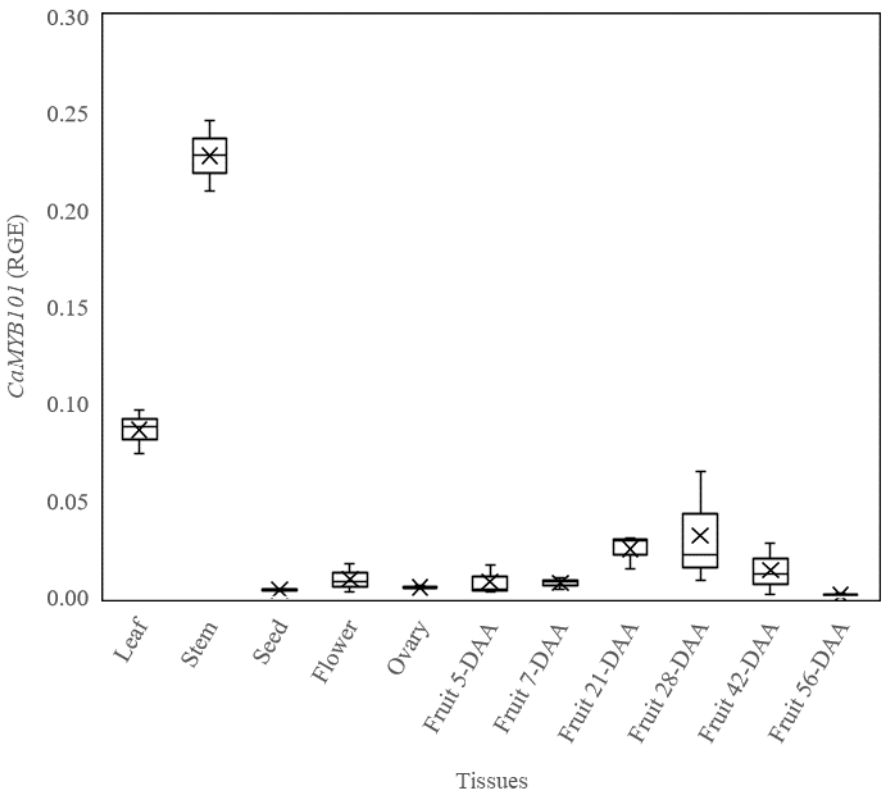

Figure 6. Expression profile of *CaMYB101* genes in different tissues and in fruits during different stages obtained via qRT-PCR. A *Ubiqutin* (*Ub*) gene was used as an internal control. Box plot is generated based on three biological replicates of each tissue (N = 3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

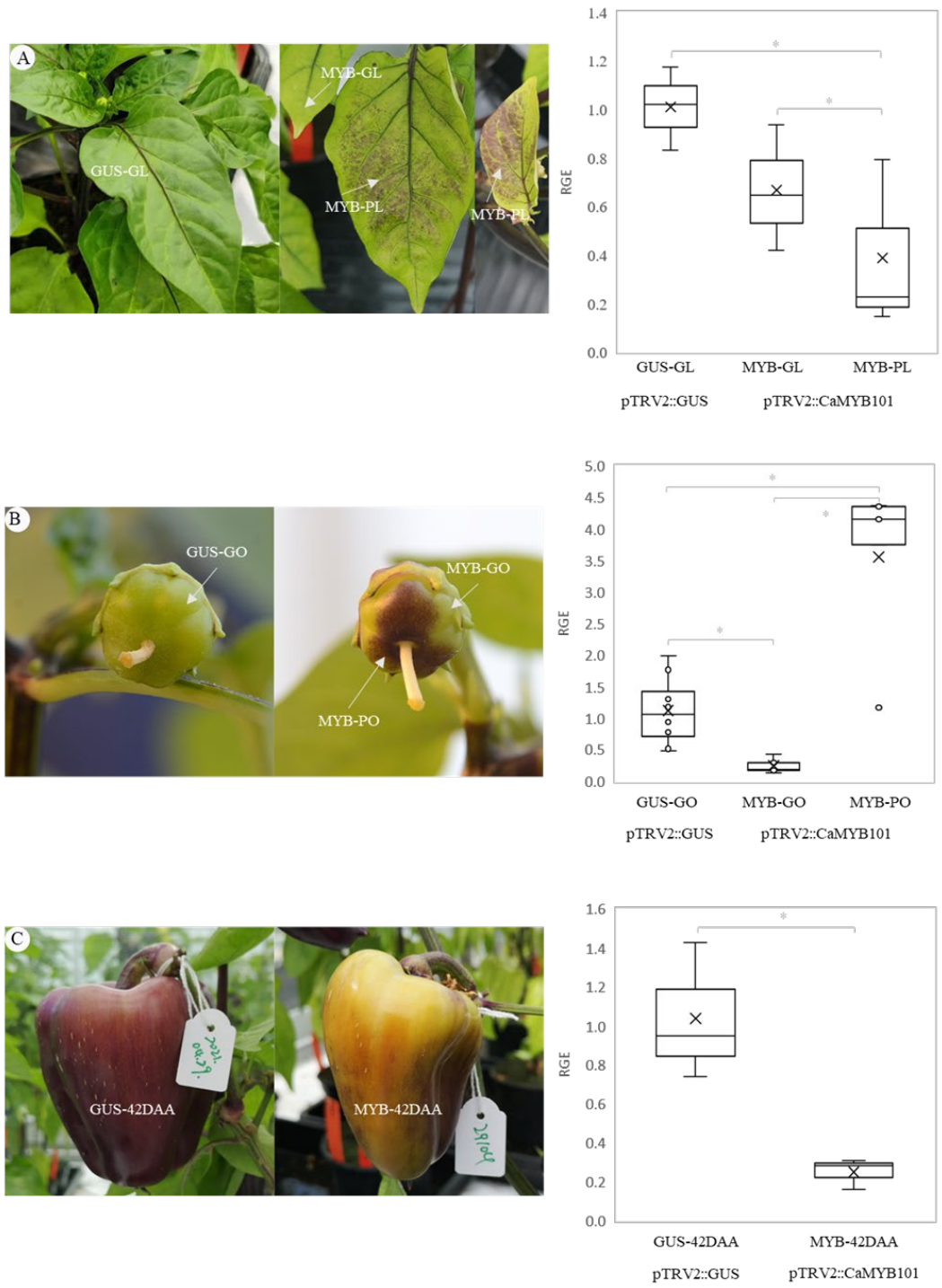

Figure 7. Functional analysis of candidate repressor *CaMYB101* of anthocyanin biosynthesis by VIGS. The VIGS experiments were repeated three times with same phenotypes. Data from the third repetition are presented. Phenotyping of pTRV2::CaMYB101 and pTRV2::GUS plants and relative gene expression level of *CaMYB101* in (A) leaves (N = 3), (B) ovaries (N = 5), and (C) fruits (42 DAA; N = 3). GUS-GL, GUS-GO and GUS-42 DAA represent the pTRV2::GUS green leaves, green ovaries and fruits at 42 DAA, respectively. MYB-GL, MYB-GO represent the green leaves and green parts of ovaries of pTRV2::CaMYB101 silenced plants. MYB-PL and MYB-PO represent the purple leaves and purple parts of ovaries of pTRV2::CaMYB101 silenced plants. MYB-42 DAA represents the pTRV2::CaMYB101 42 DAA fruits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.24.457473; this version posted August 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

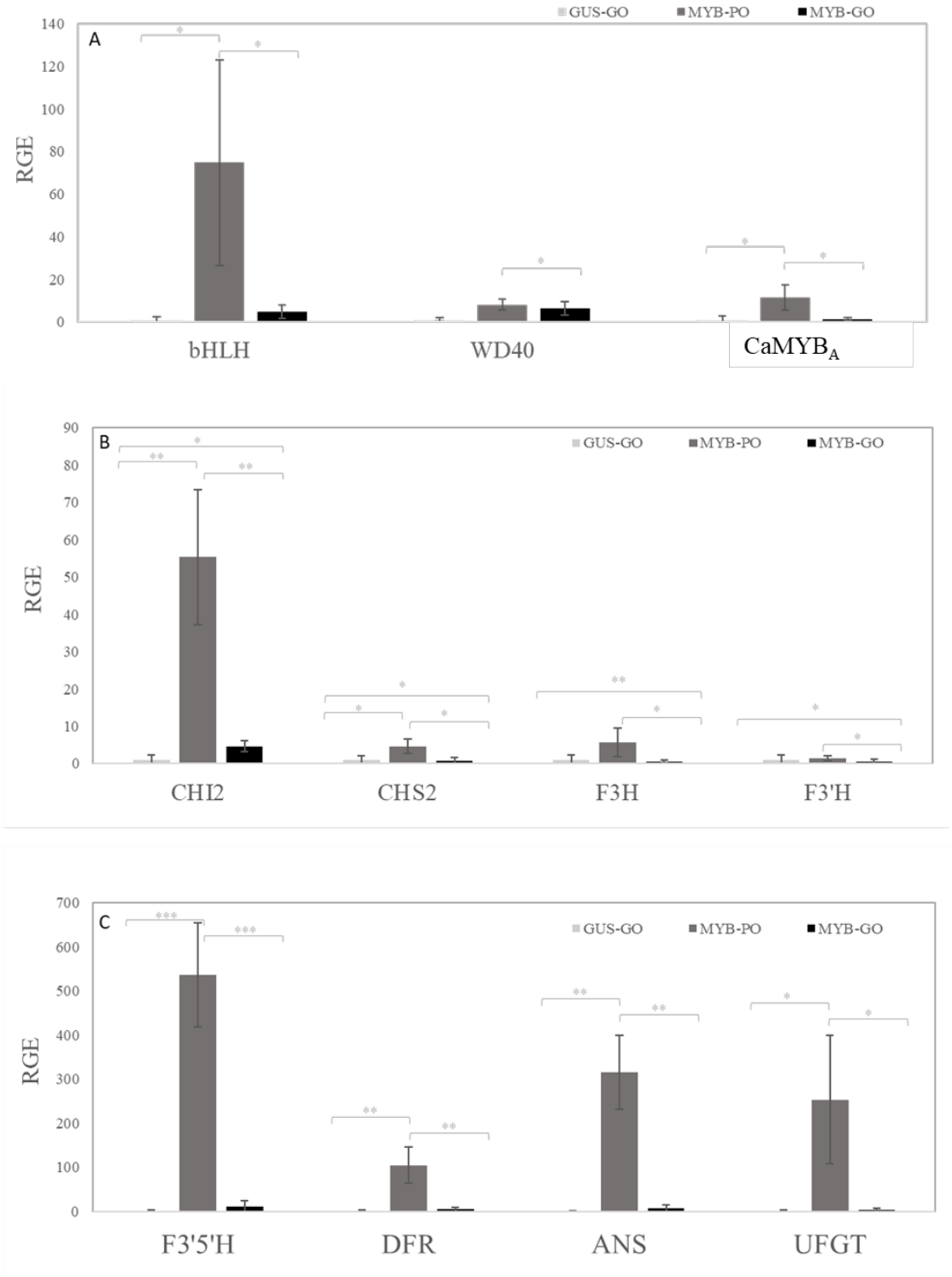

Figure 8. The relative expression level of anthocyanin biosynthetic genes in ovaries (A) *CabHLH*, *CaWD40* and *CaMYB$_A$* activator (CaMYB82 in this study and NCBI accession number AJ608992), (B) anthocyanin early biosynthetic genes *chalcone synthase 2* (CaCHS2), *chalcone isomerase 2* (*CaCHI2*), *flavanone 3-hydroxylase* (CaF3H) and *flavonoid 3'-hydroxylase* (CaF3'H) and (C) anthocyanin late biosynthetic genes *flavonoid 3', 5'-hydroxylase* (*CaF3'5'H*), *dihydroflavonol 4-reductase* (*CaDFR*), *anthocyanin synthase* (*CaANS*) and a*nthocyanidin 3-O-glucosyltransferase* (CaUFGT). GUS-GO, MYB-GO and MYB-PO represent the pTRV2::GUS ovaries, purple spots of pTRV2::CaMYB101 ovaries and green parts of pTRV2::CaMYB101 ovaries, separately.