Planta (2022) 256: 67
https://doi.org/10.1007/s00425-022-03979-z

ORIGINAL ARTICLE

The R2R3-MYB gene family in *Cicer arietinum*: genome-wide identification and expression analysis leads to functional characterization of proanthocyanidin biosynthesis regulators in the seed coat

Ruchika Rajput[1] · Shivi Tyagi[1] · Jogindra Naik[1] · Boas Pucker[2,3] · Ralf Stracke[2] · Ashutosh Pandey[1]

Received: 29 June 2022 / Accepted: 19 August 2022 / Published online: 29 August 2022

Abstract

Main conclusion **We identified 119 typical CaMYB encoding genes and reveal the major components of the proanthocyanidin regulatory network. CaPARs emerged as promising targets for genetic engineering toward improved agronomic traits in *C. arietinum*.**

Abstract Chickpea (*Cicer arietinum*) is among the eight oldest crops and has two main types, i.e., desi and kabuli, whose most obvious difference is the color of their seeds. We show that this color difference is due to differences in proanthocyanidin content of seed coats. Using a targeted approach, we performed in silico analysis, metabolite profiling, molecular, genetic, and biochemical studies to decipher the transcriptional regulatory network involved in proanthocyanidin biosynthesis in the seed coat of *C. arietinum*. Based on the annotated *C. arietinum* reference genome sequence, we identified 119 typical CaMYB encoding genes, grouped in 32 distinct clades. Two CaR2R3-MYB transcription factors, named CaPAR1 and CaPAR2, clustering with known proanthocyanidin regulators (PARs) were identified and further analyzed. The expression of *CaPAR* genes correlated well with the expression of the key structural proanthocyanidin biosynthesis genes *CaANR* and *CaLAR* and with proanthocyanidin levels. Protein–protein interaction studies suggest the in vivo interaction of CaPAR1 and CaPAR2 with the bHLH-type transcription factor CaTT8. Co-transfection analyses using *Arabidopsis thaliana* protoplasts showed that the CaPAR proteins form a MBW complex with CaTT8 and CaTTG1, able to activate the promoters of *CaANR* and *CaLAR in planta*. Finally, transgenic expression of *CaPARs* in the proanthocyanidin-deficient *A. thaliana* mutant *tt2-1* leads to complementation of the *transparent testa* phenotype. Taken together, our results reveal main components of the proanthocyanidin regulatory network in *C. arietinum* and suggest that CaPARs are relevant targets of genetic engineering toward improved agronomic traits.

Keywords *Arabidopsis thaliana* · *Cicer arietinum* · *Cis*-element · Co-transfection · MBW complex · Proanthocyanidin · R2R3-MYB · Transcription factor

Communicated by Dorothea Bartels.

✉ Ashutosh Pandey
ashutosh@nipgr.ac.in

Ruchika Rajput
ruchikarajput@nipgr.ac.in

Shivi Tyagi
shivityagi@nipgr.ac.in

Jogindra Naik
jogindra95@nipgr.ac.in

Boas Pucker
bpucker@cebitec.uni-bielefeld.de

Ralf Stracke
stracke@cebitec.uni-bielefeld.de

[1] National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi 110067, India

[2] Chair of Genetics and Genomics of Plants, Bielefeld University, 33615 Bielefeld, Germany

[3] Institute of Plant Biology and Braunschweig Integrated Centre of Systems Biology (BRICS), TU Brunswick, Brunswick, Germany

Abbreviations

ANR	Anthocyanidin reductase
BBS	BHLH-binding sequences
BiFC	Bimolecular fluorescence complementation
DMACA	4-Dimethylaminocinnamaldehyde
LAR	Leucoanthocyanidin reductase
MBS	MYB-binding sequences
PA	Proanthocyanidin
PAR	Proanthocyanidin regulator
TF	Transcription factors

Introduction

v-Myeloblastosis viral oncogene (v-MYB) homolog proteins, called MYB protein family, are one of the largest and well-studied family of transcription factors (TFs). MYB proteins are present in all eukaryotes. Like other TF proteins, they are composed of modules: the N-terminal region carries the conserved DNA-binding MYB domain, whereas the C-terminal part is very diverse and important for the regulatory TF activity. The MYB domain contains one to three imperfect repeats, named R1, R2, and R3. Each repeat is around 50–53 amino acids and forms a helix-turn-helix structure (summarized in Ogata et al. 1996). Based on the number of adjacent repeats in the MYB domain, MYB proteins are classified into three categories (1) R1-MYB with one repeat, (2) R2R3-MYB with two repeats, and (3) R1R2R3-MYB having three repeats (Jin and Martin 1999). The majority of plant MYB TFs belongs to the R2R3-MYB class thought having evolved from an R1R2R3-MYB gene ancestor by the loss of the sequences encoding the R1 repeat and subsequent expansion of the gene family (Kranz et al. 2000). R2R3-MYB proteins are characterized by the presence of the N-terminal domain specific for DNA-binding motif with conserved signature bHLH-interacting motif ([DE]Lx(2)[RK]x(3)Lx(6)Lx(3)R) (Zimmermann et al. 2004) and the C-terminal domain consists of distinct subgroups (Ss) known to regulate different biological functions, where the members of S5, constituting a S5-motif, have a role in proanthocyanidin (PA) formation (Stracke et al. 2001; Schwinn et al. 2016). Based on conserved amino acid sequence motifs present in most C-terminal domains, R2R3-MYBs are classified into so called 28 subgroups, which have been introduced from *A. thaliana* work (Stracke et al. 2001; Dubos et al. 2010) and have been extended by works on Lotus (Shelton et al. 2012). These R2R3-MYB TFs are considered to be plant-specific, because they were only found in plants and are associated with the regulation of plant-specific processes including specialized metabolism, biotic and abiotic stress responses, defense, phytohormone signaling, root genesis, plant growth and development (Dubos et al. 2010).

Some R2R3-MYB proteins are known as key regulatory proteins involved in the biosynthesis of flavonoids, a group of specialized plant metabolites, which are synthesized through the phenylpropanoid pathway. Flavonoids serve multiple physiological functions in plants, such as attraction of pollinators, facilitation of seed dispersal, and protection against adverse environmental conditions (Winkel-Shirley 2001). Besides, flavonoids act as beneficial micronutrients in human and animal diet and are associated with health-promoting properties, such as anti-cancerous, immunosuppressive, anti-inflammatory, protection against cardiovascular disease (Bondonno et al. 2019; Carmona-Gutierrez et al. 2019). Flavonoids are divided according to their chemical aglycon structure into various classes, including phlobaphenes, flavanones, flavones, flavonols, isoflavones, anthocyanin and PA (Winkel-Shirley 2001).

Regulation of the flavonoid biosynthesis pathway is well understood in *A. thaliana.* However, research on flavonoids in crops is gaining momentum during the last years (Pang et al. 2013; Rajput et al. 2022). Structural genes encoding enzymes of the flavonoid biosynthesis pathway are primarily regulated at the transcriptional level through regulatory proteins belonging to diverse TF families like MYB, basic helix–loop–helix (bHLH), WD40 and WRKY (Quattrocchio et al. 2006).

R2R3-MYB proteins play a key role in the regulation of the different branches of flavonoid biosynthesis. One subgroup of R2R3-MYBs acts independently of known cofactors, regulating one of the evolutionary old branches of the flavonoid biosynthesis pathway: the flavonol biosynthesis branch (Stracke et al. 2007; Naik et al. 2021). Other R2R3-MYB subgroup members form complexes with bHLH and WD40 TFs (called MBW complexes), regulating later flavonoid biosynthesis branches leading to the formation of anthocyanins and proanthocyanidins (PAs, also named condensed tannins) (Zimmermann et al. 2004).

The regulation of PA synthesis has been well characterized by the analysis of *transparent testa (tt)* mutants that fail to accumulate PAs in the seed coat, making the seed coat transparent and revealing the yellow color of the underlying cotyledons (Nesi et al. 2001). In *A. thaliana*, a ternary complex of AtMYB123/TT2, AtbHLH42/TT8 and AtTTG1 was found to be the major regulator of PA metabolism in seeds (Nesi et al. 2001; Baudry et al. 2004). Hereby, the R2R3-MYB protein AtTT2 confers target gene specificity to the MBW complex, being the key determinant for PA accumulation (Baudry et al. 2004). Similar TT2-like specific R2R3-MYB regulators have been characterized in grapevine [VvMYBPA1, VvMYBPA2, VvMYB5a, and VvMYBPAR (Deluc et al. 2006, 2008; Bogs et al. 2007; Terrier et al. 2009; Koyama et al. 2014)], apple [MdMYB9, MdMYB11, and MdMYB12 (Gesell et al. 2014; An et al. 2015; Wang et al. 2017)] and peach [PpMYBPA1 and

PpMYB7 (Ravaglia et al. 2013; Zhou et al. 2015)]. Most of them exhibit tissue- or organ-specific expression and responsiveness to different stresses, but all are capable of activating the *anthocyanidin reductase (ANR)* gene promoter as part of a specific MBW complex (Appelhagen et al. 2014).

C. arietinum, a grain legume, is an economically significant crop upholding the second position globally among pulses in terms of production and importance for food security in Eastern Africa and South Asia (Khandal et al. 2020). There it is the most preferred crop since it contains high protein, calcium, iron, phosphorus, and other minerals. Because of these properties, *C. arietinum* has become a significant diet for vegetarians all over the world. The chickpea genome has been sequenced and more than 28,000 genes have been identified (Varshney et al. 2013). The color of the seed coat is an important agronomic characteristic in chickpea and varies between cultivars, germplasms, landraces, and accessions (Bajaj et al. 2015). *C. arietinum* is categorized into two distinct types: desi and kabuli, differing in size, color and surface of seeds, flower color and morphology. Seeds of desi type are darker in color, whereas kabuli seeds have a lighter seed coat color. Penmetsa et al. (2016) identified a bHLH-type TF whose allelic variants are associated with seed coat and flower pigmentation in desi and kabuli types of *C. arietinum.* However, no functional characterization of transcriptional PA biosynthesis regulators has apparently been reported for *C. arietinum*.

In the present study, we have identified *R2R3-MYBs* in *C. arietinum* and analyzed their plant part-specific gene expression. The expression of two *CaMYB* genes which putatively encode PA regulators (CaPARs), showed consistent correlation with the PA content in seeds of the color contrasting types of *C. arietinum*. The role of *CaPAR* genes in PA biosynthesis was characterized by ectopic expression in PA-deficient *tt2* mutants of *Arabidopsis*. Further, by Yeast two-hybrid and BiFC assays, we showed how previously characterized CaTT8 (Penmetsa et al. 2016) interacted with identified CaPAR proteins. Co-transfection experiments were also performed to study the regulatory role of CaPARs and CaTT8 on the promoters of late biosynthetic genes of chickpea. Therefore, the present data shed further light on the molecular mechanisms underlying the transcriptional regulation of the PA biosynthesis in *C. arietinum*.

Materials and methods

Plant material

Mature seeds of seed coat color contrasting *C. arietinum* types, desi (ICC 4958, BGD256) and kabuli (ICCV2, ICC 6253) were collected from the fields of National Institute of Plant Genome Research (NIPGR), New Delhi, India. The *A. thaliana* L*er* wild type and *transparent testa 2–1* (*tt2-1,* NASC stock NW83, L*er* background, X-ray) mutant plants as well as the corresponding transgenic *CaPAR* overexpressing *A. thaliana* lines were grown with a photoperiod of 16 h light and 8 h dark at 22ºC in a plant growth chamber (AR-41L3; Percival, Perry, IA, USA). *Nicotiana benthamiana* plants were grown in a plant growth chamber with 8 h light/16 h dark photoperiod at 22 °C. The *A. thaliana* suspension cell culture At7 (Trezzini et al. 1993) was derived from hypocotyl of the reference accession Columbia (Col) and handled as described in Stracke et al. (2016).

Identification and chromosomal distribution of *CaMYB* genes

The availability of an advanced draft genome sequence assembly (version 2.0) of desi (ICC 4958)-type chickpea (Parween et al. 2015) enabled the genome-wide identification of *R2R3-MYB* genes in *C. arietinum.* A set of previously known R2R3-MYB protein sequences from various plant species were considered as the query in BLASTp searches (Altschul et al. 1990) with a 10^{-10} *e*-value cutoff against the annotated protein sequences of *C. arietinum* retrieved from NCBI (GCF_000331145.1; Annotation Release 101). All identified CaMYB proteins were manually inspected for the presence of an integral R2R3-MYB domain. The sequences with redundancy or incomplete open reading frames were excluded from the data set. A customized Python script was deployed to summarize the genomic position and structural gene details including the exon and intron number of the *CaMYB* genes (https://github.com/bpucker/CaMYBs). The *CaR2R3-MYB* nomenclature was assigned to the order of annotated genes on the *C. arietinum* pseudochromosomes. The figural mapping of *CaMYB* genes on the various pseudochromosomes was done using MapInspect (https://mapinspect.software.informer.com/). To predict gene duplication events, genes with ≥ 80% similarity and distanced by > 5 Mb were considered as the result of segmental duplications, however, genes localized within the 5 Mb region were considered as the result of tandem duplications (He et al. 2016).

Multiple sequence alignment and phylogenetic study

The sequences of the R2R3-MYB domains were extracted from the full-length CaMYB protein sequences. Alignment of each R2 and R3 domain was constructed through MAFFT v7 (Katoh and Standley 2013) and represented by weblogo (Crooks et al. 2004). A phylogenetic analysis was carried out by aligning the R2R3-MYB domain sequences from *C. arietinum, A. thaliana* and selected plant landmark MYB sequences (Stracke et al. 2014) using ClustalW at default parameters (Thompson et al. 1994).

The randomized axelerated maximum likelihood (RAxML) method was used for the construction of phylogeny at 100 bootstrap value and a cladogram figure was visualized with iTOLv6.4 (https://itol.embl.de/). The multiple sequence alignment of CaPARs with the functionally characterized R2R3-MYB-type PA regulator proteins including AtTT2 (NM122946) from *A. thaliana*, MtPAR (HQ337434) from *M. truncatula*, LjTT2a (BAG12893.1) from *Lotus japonicus*, DkMYB2 (AB503699), and DkMYB4 (AB503701) from *Diospyros kaki*, AaMYB3 (MH349476) from *Anthurium andraeanum*, TaMYB14 (AFJ53054.1) from *Trifolium arvense*, MtMYB14 (Medtr4g125520.1) from *Medicago truncatula*, VvMYBPA1 (CAJ90831.1) and VvMYBPA2 (ACK56131.1) from *Vitis vinifera* was carried out using CLUSTALW with default parameters.

In silico gene expression analysis

C. arietinum RNA-Seq data sets for 10-day-old seedlings, 10-day-old seedling roots, 10-day-old seedling shoots, shoot apical meristem, developmental stages of seeds, germinating seedling, developmental stages of flower, reproductive plant, root, vegetative plant and young leaves (BioProjects: PRJNA182724, PRJNA79731, PRJNA316844, PRJNA316845, SAMN00794551)(Garg et al. 2011; Niraj et al. 2016; Rajkumar et al. 2020) were retrieved from the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra). STAR v2.5.1b (Dobin et al. 2013) was applied to align the reads to the *C. arietinum* genome sequence in the 2-pass mode. Reads were considered mapped if the alignment similarity exceeds 95% and cover 90% of the read length as previously described (Haak et al. 2018). FeatureCounts v1.5.0-p3 (Liao et al. 2014) was deployed with default settings for the quantification of gene expression. Resulting count tables were processed and combined by previously developed Python scripts (Haak et al. 2018). Heatmap construction was performed with HCE (Hierarchical Clustering Explorer 3.5) and the hierarchical clustering of genes was executed by Euclidean distance method (Seo et al. 2006).

Gene expression analysis (RT-qPCR)

RNA was isolated from seed samples using GSure® Plant RNA Isolation Kit (GCC Biotech, Kolkata, West Bengal, India) and subsequently treated with RNase-free DNase I (Thermo Fisher Scientific, Waltham, MA, USA). Total RNA was subjected to reverse transcription to generate first-strand cDNA using oligo (dT) primers. RT-pPCR analysis of a set of selected genes was carried out using 2 × PCR Master mix (Applied Biosystems, Waltham, MA, USA). The PCR mix contained 1 µL of diluted cDNA (corresponding to 10 ng total RNA), 5 µL of 2 × SYBR Green PCR Master Mix (Applied Biosystems), and 10 nM of each gene-specific primer in a final volume of 10 µL. Expression of different genes involved in proanthocyanindin biosynthesis was studied through 7500 Fast Real time PCR System (Applied Biosystems). All RT-qPCRs were performed under the following conditions: 20 s at 95 °C, 3 s at 95 °C, and 40 cycles of 30 s at 60 °C in 96-well optical reaction plates (Applied Biosystems). The integrity of amplicons was verified by melting curve analysis (60 to 95 °C) after 40 cycles. Transcript abundance were normalized to the constitutively expressed chickpea elongation factor 1-α (EF-1α) (GenBank: AJ004960.1) and Ca βtubulin (CaβTub) (LOC101495306) coding genes and relative gene expression levels were calculated using the cycle threshold (Ct) $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). Mean values were derived from two independent biological replicates, each having three technical replicates. Primers used in this work are given in Table S1.

Cloning of transcriptional regulator cDNAs

First strand cDNA of *C. arietinum* desi accession (BGD 256) seed coat was used as template for PCR amplification of full-length *CaPAR1*, *CaPAR2*, *CaPAR*-like, *CaMYB16*, *CaMYB30*, *CaTT8* (XP_027189966.1) and *CaTTG1* (Ca _XP004502764.1) coding sequence (CDS, without the stop codon), using a set of Gateway™ attB-site-containing oligonucleotides (Table S1), designed on the basis of sequence information from the *C. arietinum* genome database. The resulting amplicons were recombined into the Gateway™ cloning vector pDONR^{TM}zeo (Invitrogen) and transformed in *E. coli* TOP10 cells. Integrity of resulting Entry clones was proven by Sanger sequencing by the sequencing core facility of NIPGR (New Delhi, India).

Sub-cellular localization of CaPAR

For the analysis of subcellular localization of CaPAR1,CaPAR2 and CaPAR-like CaMV35S-driven C-terminal YFP fusion constructs were generated by Gateway™ LR-recombination into the binary vector pGWB441 (Nakagawa et al. 2007). Resulting plasmids were transformed into *Agrobacterium tumefaciens* GV3101::pMP90 (Koncz and Schell 1986). Agrobacteria harboring CaPAR-YFP constructs and Agrobacteria harboring a plasmid with a NLS-RFP nuclear marker fusion-protein construct (Kumar et al. 2018) were re-suspended in freshly made infiltration medium (10 mM $MgCl_2$, 10 mM MES/KOH, pH 5.7, 150 µM acetosyringone) and co-infiltrated with a syringe onto the abaxial surface of *N. benthamiana* leaf and kept at 22 °C for 48 h. YFP and RFP fluorescence was observed using a Leica TCS SP5 (Leica Microsystems, Wetzlar, Germany) laser-scanning confocal microscope at 514–527 nm and 558–583 nm, respectively.

Quantification of soluble and insoluble PA

Seed coat and embryo/endosperm were dissected from mature seeds of the color contrasting *C. arietinum* types desi and kabuli. Extraction and quantification of soluble and insoluble PAs was performed according to Pang et al. (2007). For soluble PA quantification, catechin was used to prepare a calibration curve (200, 400, 600, 800 and 1000 µg/mL) and content were determined as catechin equivalent. Samples, blanks, and standards were read within 15 min on a FLUOstar® Omega Microplate Reader (BMG, Ortenberg, Germany) equipped with a 640-nm emission filter. Dried seed coat and embryo/endosperm were used for quantification of insoluble PAs. Plant material was treated with 1 mL butanol–HCl reagent (5:95, v/v) containing 0.7 g $FeCl_3$/L, homogenized and sonicated at room temperature for 1 h with 30 kHz frequency. Sonicated extract was filtered with PVDF filters (Merck Millipore) and transferred to cuvettes for determination of absorption at 550 nm. Afterward, samples were boiled for 1 h and after cooling to room temperature, the A_{550} was determined again. Absorbance differences were converted into PA equivalents using a standard curve (50, 100, 150, 200, and 250 µg/mL) of procyanidin B1 (Merck, Darmstadt, Germany).

Ultra-high-performance liquid chromatography (UHPLC) analysis

Separation for qualitative and quantitative analysis of monomeric and oligomeric PAs was performed on the 1290 Infinity II series HPLC system (Agilent Technologies, Santa Clara, CA, USA) equipped with 1290 Infinity II series pump, auto-sampler, column compartment, and thermostat using a Zorbax Eclipse Plus C_{18} column (2.1 X 100 mm, 1.8 µm) maintained at 30 °C. The mobile phase consisted of an aqueous solution of 0.1% LC–MS-grade formic acid (solution A) and 0.1% formic acid in LC–MS-grade acetonitrile (solution B). The gradient for solution B was programmed as follows: 0 to 2 min, 5%; 2 to 10 min, 5%–15%; 10 to 32 min, 15%–30%; 32 to 40 min, 30%–80%; 40 to 42 min, 80%–90%; and 42 to 47 min, 90%– 0%. Other chromatographic parameters included a constant flow of 270µL/min (injection volume, 3 µL), and run time of 47 min including equilibration. Before analysis, all samples were filtered through 0.22 µm PVDF syringe filter (Merck). The filtrate was collected in a clear-glass HPLC vial (Agilent Technologies).

LCMS analysis

Analysis of the PAs was performed using an UPLC system (Exion LC Sciex, Framingham, MA, USA) coupled to a triple quadrupole system (QTRAP6500 + ; ABSciex) using an electrospray ionization. The voltage was set at 5500 V for positive ionization. The values of gas 1 and gas 2 (70 psi), curtain gas (40 psi), collision-assisted dissociation (medium), and temperature of the source (650 °C) were used. The mass spectrometer was used in multiple reaction monitoring mode (MRM) for qualitative and quantitative analysis. Analytical standards were purchased from Merck. Identification and quantitative analysis were carried out using Analyst software (version 1.5.2).

Yeast two-hybrid assay

For yeast two-hybrid (Y2H) assays, the CDSs of CaPARs and CaTT8 were recombined from Entry clones into the destination prey vector pGADT7g (Clontech Laboratories Inc., San Jose, CA, USA) and the bait vector pGBKT7g (Clontech Laboratories Inc.), respectively. The resulting plasmids were co-transformed into the *Saccharomyces cerevisiae* strain Y2HGold (Clontech Laboratories Inc.) and cultured at 30 °C on medium lacking Trp and Leu. For interaction analysis, the bait and prey-containing yeast colonies were patched in medium lacking Trp, Leu, His and adenine (QDO) with 40 mg/mL X-gal, 20 ng/mL aureobasidin (AbA) and 5 mM 3-amino-1,2,4-triazole (3-AT). p53/large T antigen interaction (Pipas and Levine 2001) was used as positive control and empty pGADT7g /pGBKT7g vectors as negative control.

Bimolecular fluorescence complementation (BiFC) assay

For bimolecular fluorescence complementation (BiFC) assays, CaPAR CDSs were recombined from Entry clones into the vector 35S-pSITE-nYFP-C1 and CaTT8 into the 35S- vector pSITE-cYFP (Martin et al. 2009). The resulting plasmids were transformed individually into *Agrobacterium tumefaciens* strain GV3101::pMP90 (Koncz and Schell 1986). Agrobacteria, harboring the nYFP- and cYFP fusion proteins to be tested, were re-suspended in freshly prepared infiltration medium (10 mM $MgCl_2$, 10 mM MES/KOH, pH 5.7 and 150 µM acetosyringone) and co-infiltrated into *N. benthamiana* leaves. Empty vectors (nYFP-cYFP) were used as negative control. After incubation for 48 h at 22 °C, YFP fluorescence was detected using a Leica TCS SP5 (Leica Microsystems, Wetzlar, Germany) laser-scanning confocal microscope at 514 nm excitation and 527 nm emission wavelength.

Complementation analysis in *A. thaliana*

Entry clones of *CaPAR* were recombined into the T-DNA binary destination vector pMDC-32 (Curtis and Grossniklaus 2003) by LR-reaction, putting *CaPAR* CDSs under the control of a 2 × 35S CaMV promoter. Resulting plasmids were

transformed into *A. tumefaciens* strain GV3101::pM90RK (Koncz and Schell 1986) by electroporation and T-DNA was transformed via *Agrobacterium*-mediated transfer into the PA-deficient *A. thaliana tt2-1* regulatory mutant by floral dip method (Clough and Bent 1998). Selection of T1 transgenic plants was conducted on MS solid medium containing 15 mg/L hygromycin. Resistant seedlings were transferred into soil and grown in plant growth chamber to seed production. T2 seeds were analyzed for PA accumulation.

4-dimethylaminocinnamaldehyde (DMACA) staining of seeds and PA content determination

DMACA staining was performed to confirm the presence of PAs in *A. thaliana* seeds according to a previously described protocol (Wang et al. 2017). PA content was measured using DMACA reagent and catechin as standard using a spectrofluorometer (Varian).

Structural analysis of the *CaANR and CaLAR* promoter

A 1328 bp fragment upstream of the transcriptional start of the *CaANR* gene *(proCaANR-1328,* Ca _XP004502764.1*)* and *CaLAR* gene *(proCaLAR-1289,* Ca_XP004496202.1) was analyzed for the presence of putative MYB and bHLH candidate-binding sites. Several regulatory-binding sites were identified using the search in plant cis-acting regulatory DNA elements in PLACE database (http://www.dna.affrc.go.jp/PLACE/signalscan.html) (Higo et al. 1999) as well as the motif previously identified (Wang et al. 1997; Hartmann et al. 2005).

Co-transfection analysis in *A. thaliana* At7 protoplasts

A promoter fragment of 1328 bp *CaANR (anthocyanidin reductase)* and 1289 bp *CaLAR* genes were amplified (Table S1) and cloned in the pDISCO vector (Stracke et al. 2010) giving a *proCaANR-1328-uidA*(GUS) and *proCaLAR-1289-uidA*(GUS) constructs used as reporters in the co-transfection assay. To generate 35S-driven TF effector constructs, the *CaPARs, CaMYB16* and *CaMYB30, CaTT8* and *CaTTG1* CDS were recombined from Entry clones into the vector pBTdest (Baudry et al. 2004).

The cultivation of At7 cells, co-transfection of effector and reporter into protoplasts and determination of activation capacity was performed as described in Stracke et al. (2016). In short, a total of 25 µg premixed plasmid DNA containing 10 µg of reporter construct, 1 µg of each effector construct, 5 µg of LUC plasmid (transfection control and standardization) and a non-reacting filling plasmid were transfected into At7 protoplasts. Transfected protoplasts were incubated for 20 h at 26 °C in the dark followed by harvesting the protoplast for LUC, GUS and Bradford assay. Specific GUS activity is given in pmol 4-methylumbelliferone (4-MU) mg^{-1} of protein min^{-1}. Standardized GUS activity was calculated by multiplication of the specific GUS activity value with a correction factor derived from the ratio of the specific LUC activity in the given sample to the mean specific LUC activity (describing the transformation efficiency) of a set of six experiments.

Co-expression analysis

Co-expression analysis of *C. arietinum* genes was performed using publically available RNA-Seq data of different tissues and developmental stages. A dedicated Python script (https://github.com/bpucker/CaMYBs) was applied for the calculation of the correlation in gene expression between different gene pairs using the Spearman correlation function of the scipy package (https://www.scipy.org/). Genes having expression value < 10 FPKM (Fragments Per Kilobase Million) were filtered out. The downstream analysis was carried out with genes showing $r > 0.7$ and P-value < 0.05. Expression of all genes in the selected data set was successively compared with the expression of *CaPAR1, CaPAR2, CaTT8* and *CaTTG1*. Results were ranked by decreasing correlation coefficient (File S1). Functional annotation was assigned to co-expressed genes using the *A. thaliana* (Araport11) annotation based on reciprocal best BLAST hits or best BLAST hits, respectively, as previously described (Pucker 2016; Haak et al. 2018). The genes with co-expression values ≥ 0.85 and other co-expressed genes with known functions were considered for the construction of an interaction network using Gephi 0.9.1 software (Bastian et al. 2009).

Results

Genome-wide identification of *CaMYB* genes, genomic organization and duplications

Initially, we identified 150 putative R2R3-MYB protein-coding genes based on amino acid sequence similarity to previously characterized MYB proteins. From these, we excluded 31 *MYB* genes having incomplete open reading frame. Finally, we identified a total set of 119 typical *CaMYB* genes of which 108 *CaMYB* genes (91%) could be assigned to chromosomes (Fig. S1), while eleven *CaMYB* genes could not, due to their position on unassigned contigs. Therefore, 108 *CaMYB* genes were designated as *CaMYB1* to *CaMYB108* ordered by their position on the pseudochromosomes (Table 1). The designation of the remaining genes *CaMYB109* to *CaMYB119* was done in alphanumerical order of the gene identifier from the

Table 1 List of annotated *R2R3-MYB* genes in the *C. arietinum* genome with functional classification

Gene name	Gene ID	Chr.	Orientation	Protein length [aminoacid]	Coding exon nr.	Functional assignment	Landmark MYB in clade
CaMYB37	XP_012569365	3	+	271	3	Abiotic stress response, flower morphogenesis, stilbene	NtMYB1, NtMYB2, AtMYB13 (AtMYBLFGN), AtMYB14, AtMYB15 (ATY19), OsMYB4, PhMYB2
CaMYB42	XP_012569580	3	+	207	3		
CaMYB45	XP_004495379	4	−	276	3		
CaMYB54	XP_012570437	4	+	272	3		
CaMYB13	XP_004489482	2	−	339	3	Anther and tapetum development	AtMYB80, AtMYB35 (TDF1)
CaMYB63	XP_004500173	5	+	325	3		
CaMYB109	XP_004513209	Un	+	301	3		
CaMYB111	XP_004514305	Un	+	550	4	Anther and leaf development, Stress response	PhMYB3, AtMYB33, AtMYB65, AtMYB81, AtMYB97, AtMYB101 (ABS7), AtMYB104, AtMYB120, AtMYB91 (AS1), HvGAMYB, GmGA-MYB2
CaMYB8	XP_004487418	1	−	478	3		
CaMYB25	XP_004491297	2	+	307	3		
CaMYB87	XP_004507522	6	−	710	3		
CaMYB93	XP_012573637	7	−	476	3		
CaMYB33	XP_004493026	3	−	605	3		
CaMYB62	XP_004500080	5	−	542	3		
CaMYB117	XP_004516525	Un	+	504	5		
CaMYB32	XP_004492354	3	−	355	3		
CaMYB16	XP_004489415	2	+	250	6	Anthocyanin, betalains	MdMYB10, LeANT1, BvMYB1Y, VvMYBA1, AmROSEA1, AmVENOSA, PhAn2, AtMYB113, AtMYB90 (AtPAP2), AtMYB75 (AtPAP1), AtMYB114
CaMYB30	XP_004491866	3	+	236	3		
CaMYB6	XP_004487066	1	+	298	3	Axillary meristem, Root growth	AtMYB37, AtMYB38, AtMYB87, AtMYB68, AtMYB84, AtMYB36, LeBLIND
CaMYB38	XP_004494064	3	−	251	3		
CaMYB57	XP_004498932	4	−	296	3		
CaMYB58	XP_004498936	4	+	345	3		
CaMYB72	XP_004503043	5	−	263	3		
CaMYB82	XP_004506057	6	−	320	3		
CaMYB88	XP_004507824	7	−	355	3		
CaMYB94	XP_004509048	7	+	280	3		
CaMYB105	XP_004511803	8	+	324	3		
CaMYB31	XP_012568996	3	+	569	8	Cell cycle control	
CaMYB56	XP_012570676	4	+	302	7		
CaMYB55	XP_004498481	4	+	1003	22		
CaMYB67	XP_004501185	5	−	923	24		
CaMYB104	XP_004511594	8	+	471	14		
CaMYB5	XP_004487036	1	+	222	3	Cell cycle regulation	AtMYB48, AtMYB59

Table 1 (continued)

Gene name	Gene ID	Chr.	Orientation	Protein length [aminoacid]	Coding exon nr.	Functional assignment	Landmark MYB in clade
CaMYB17	XP_004489664	2	+	374	3	Cell wall, Lignin, Axillary meristem	AtMYB52 (BW52), AtMYB54, AtMYB56, AtMYB69, AtMYB89, AtMYB105, AtMYB110, AtMYB117
CaMYB19	XP_004490084	2	−	415	3		
CaMYB35	XP_004493239	3	−	333	10		
CaMYB73	XP_004502570	5	−	291	2		
CaMYB74	XP_004503065	5	+	266	2		
CaMYB79	XP_004505415	6	+	432	3		
CaMYB112	XP_012567212	Un	−	231	4		
CaMYB1	XP_004485696	1	−	357	3	Defense, Stress response	AmMYB306, AtMYB49, AtMYB41, AtMYB60, AtMYB30, AtMYB31 (ATY13), AtMYB94 (AtMYBCP70), AtMYB47, AtMYB95, AtMYB96 (MYBCOV1), AtMYB74, AtMYB78, AtMYB73, AtMYB70, AtMYB77, AtMYB44 (AtMYBR1), AtMYB1, AtMYB25, AtMYB102 (ATM4), AtMYB108, AtMYB109, AtMYB2, AtMYB112, AtMYB116, AtMYB62 (BW62B; BW62C), CpMYB5, CpMYB10
CaMYB2	XP_004486524	1	+	322	3		
CaMYB7	XP_004487343	1	+	327	3		
CaMYB9	XP_004487648	1	−	320	3		
CaMYB14	XP_004489392	2	+	323	3		
CaMYB24	XP_004491024	2	+	258	1		
CaMYB28	XP_004491789	3	+	320	3		
CaMYB36	XP_004493248	3	−	327	3		
CaMYB48	XP_004495882	4	+	293	3		
CaMYB49	XP_004495885	4	+	275	2		
CaMYB52	XP_004497077	4	−	314	3		
CaMYB59	XP_004499445	5	+	324	3		
CaMYB68	XP_004501595	5	+	237	1		
CaMYB78	XP_004504959	6	−	330	3		
CaMYB20	XP_004490111	2	+	262	3		
CaMYB90	XP_004508108	7	+	305	3		
CaMYB116	XP_012567497	Un	−	346	2		
CaMYB95	XP_012573654	7	−	293	2		
CaMYB106	XP_004511885	8	−	323	1		
CaMYB70	XP_012571729	5	+	411	3	Embryogenesis, Seed maturation	AtMYB98, AtMYB22, AtMYB118, AtMYB115, AtMYB100, AtMYB119, AtMYB64 (URP10)
CaMYB91	XP_004508897	7	−	469	4		
CaMYB21	XP_004490644	2	−	364	4		
CaMYB26	XP_004491750	3	+	428	3		
CaMYB60	XP_004499811	5	−	407	3		
CaMYB64	XP_004500263	5	−	456	4		
CaMYB65	XP_004500264	5	+	420	3		

Table 1 (continued)

Gene name	Gene ID	Chr.	Orientation	Protein length [aminoacid]	Coding exon nr.	Functional assignment	Landmark MYB in clade
CaMYB39	XP_004494411	3	−	363	3	Flavonols, phlobaphenes	VvMYBF1, AtMYB11 (PFG2), AtMYB12, AtMYB111 (PFG3), BvMYB12, ZmMYBP-Wr
CaMYB34	XP_004492711	3	−	303	3	Flower meristem identity	AtMYB17 (LMI2)
CaMYB50	XP_012570165	4	−	304	7		
CaMYB100	XP_004510807	7	+	293	3		
CaMYB18	XP_004490015	2	−	471	12	Guard cell division, root gravitropism	AtMYB88, AtMYB4R1, AtMYB124
CaMYB101	XP_004511228	7	+	454	14		
CaMYB40	XP_012569502	3	+	336	3	Mucilage, Lignin, Stomatal closure	AtMYB50, AtMYB61, AtMYB86, AtMYB55
CaMYB41	XP_004494503	3	−	360	3		MYB50, MYB61
CaMYB46	XP_004495813	4	+	375	3		MYB50, MYB61
CaMYB47	XP_004495851	4	−	428	3		MYB50, MYB61
CaMYB69	XP_004501883	5	−	353	3		MYB50, MYB61
CaMYB115	XP_004515875	Un	−	351	3		MYB50, MYB61
CaMYB53	XP_004497798	4	+	251	3	Phenyl-propanoid, Lignin	AmMYB305, AtMYB121, AtMYB79, AtMYB71, AtMYB58, AtMYB72, AtMYB10, AtMYB63
CaMYB44	XP_004495255	4	−	258	3		
CaMYB11	XP_004488797	1	−	368	3		
CaMYB85	XP_004507244	6	−	241	3		
CaMYB86	XP_004507232	6	−	249	3		
CaMYB110	XP_004513558	Un	+	309	3		
CaMYB51	XP_004496837	4	−	271	3	Photomorpho-genesis	AtMYB19, AtMYB18 (LAF1), AtMYB45
CaMYB113	XP_004515699	Un	−	278	3		
CaMYB43	XP_012569639	3	−	256	2	General flavonoids, trichome	AtMYB5
CaMYB71	XP_004502203	5	+	320	2		
CaMYB99	XP_004510651	7	+	291	2		
CaMYB89	XP_004507986	7	+	310	3	Proantho-cyanidin, anthocyanin	AtMYB123 (AtTT2), TaMYBA6, OsC1, ZmMYBC1
CaMYB98	XP_004510460	7	+	276	3		
CaMYB92	XP_004508935	7	−	199	3		
CaMYB23	XP_004490862	2	+	218	3	Flavonoid repressor	FaMYB1
CaMYB81	XP_004505789	6	−	286	2	Repressors, Phenyl-propanoids, Sinapate, Lignin	AmMYB308, AmMYB315, AtMYB4, LeMYB27, AtMYB7, AtMYB3, AtMYB32, AtMYB6, AtMYB8 (HOS10)
CaMYB77	XP_004503863	6	−	293	3		
CaMYB107	XP_004512511	8	+	329	3	Suberin	AtMYB9, AtMYB107, AtMYB39
CaMYB66	XP_004500626	5	+	358	3	Root development	AtMYB92, AtMYB93, AtMYB53
CaMYB102	XP_012574351	8	−	314	2		
CaMYB80	XP_004505589	6	−	338	4		

Table 1 (continued)

Gene name	Gene ID	Chr.	Orientation	Protein length [aminoacid]	Coding exon nr.	Functional assignment	Landmark MYB in clade
CaMYB27	XP_004491748	3	+	292	3	Stamen/anther development, phenylpropanoid, secondary cell wall	AtMYB99, PtMYB1, AtMYB40
CaMYB22	XP_004490820	2	−	319	4	Secondary cell wall, lignin	PtMYB4, AtMYB43, AtMYB20, AtMYB42, AtMYB85, AtMYB46, AtMYB83, EgMYB2
CaMYB4	XP_004486969	1	+	311	2		
CaMYB61	XP_004499875	5	+	342	2		
CaMYB75	XP_004503126	6	+	334	3		
CaMYB84	XP_004507092	6	−	331	2		
CaMYB114	XP_004515780	Un	−	275	2		
CaMYB12	XP_004489258	2	+	210	3	Stamen development, Flavonols	PsMYB26, AtMYB57, AmMYB340, AtMYB24, AtMYB21
CaMYB108	XP_004512906	8	+	280	3		
CaMYB118	XP_004516766	Un	−	233	4		
CaMYB15	XP_004489411	2	+	421	3	Trichome branching, Petal morphogenesis	PhMYB1, AtMYB16 (AtMIXTA), AmMYBML3, GhMYB25, AmMIXTXA, AtMYB106 (NOK), TaMYB38, AmML2, AmML1
CaMYB29	XP_004491833	3	+	361	8		
CaMYB119	XP_004517012	Un	+	406	3		
CaMYB3	XP_004486846	1	+	343	3	Anther, trichome development	AtMYB103, AtMYB26 (MS35), AtMYB67 (ATY53)
CaMYB10	XP_004488555	1	+	319	3		
CaMYB103	XP_012574477	8	+	523	8		
CaMYB76	XP_012572125	6	+	532	7		
CaMYB97	XP_004510395	7	−	333	3		
CaMYB96	XP_004510118	7	−	309	3	Leaf, shoot, germination, morphogenesis	AtMYB125
CaMYB83	XP_004506466	6	−	302	3		

chickpea genome sequence annotation. *CaMYB* genes were found on all eight chromosomes (Fig. S1). The number of coding exons ranged from one to 24, where three exons were most frequent and occurred in 77 *CaMYB* (65%) genes. Three *CaMYB* genes have a single coding exon, 14 *CaMYB* genes have two, four have four, two have six, five have seven and three have eight coding exons. Moreover, single *CaMYB* genes have ten *(CaMYB35)*, 14 *(CaMYB104)*, 22 (*CaMYB55)*, and 24 (*CaMYB67)* coding exons (Table 1). Additionally, two gene duplications were found on chromosome 5, including one segmental duplication (*CaMYB60*, *CaMYB70*) and one tandem duplication (*CaMYB64*, *CaMYB65*) (Fig. S1).

R2R3-MYB domain and phylogenetic analysis

To investigate the R2R3-MYB domain sequence features, and the frequencies of the most prevalent amino acids at each position within the two repeats of the *C. arietinum* R2R3-MYB domain, sequence logos were produced (Fig. S2). The R2R3-MYB domain covered about 106 amino acid residues and consist of highly conserved and regularly spaced tryptophan (W) residues at position 7, 28, and 48 of the R2 repeat, and 61, 80, and 99 of the R3 repeat. However, the W_{61} in the R3 repeat showed major substitutions with phenylalanine (F) and few substitutions with isoleucine (I), leucine (L), or tyrosine (Y) residues. Also, a few substitutions of W_{99} with F in the R3 repeat were identified. Apart from conserved W residues, various other residues including glutamate (E_{11}), aspartate (D_{12}), cysteine (C_{44}), arginine (R_{47}), E_{65}, and R_{90} were also entirely conserved in the CaMYB proteins.

To gain insight into the putative functions of CaMYBs, their phylogenetic relationship was analyzed in the context of functionally known R2R3-MYBs from *A. thaliana* and several landmark R2R3-MYBs. The topology of the resulting rooted cladogram showed clustering of the analyzed MYBs into 32 functionally distinct clades (Fig. 1), where clade 2 (C2) seems to be somehow chickpea-specific, i.e., only containing CaMYBs in this analysis. The maximum number of CaMYB proteins was found in C17, while only one CaMYB protein belongs to C26, C28, and C29. Moreover, C12, C21, and C31 were specific to *Arabidopsis* in this cladogram, i.e., only containing AtMYB proteins. Further, these clades were assigned to the 28 sub-groups of which S1-25 corresponds to subgroups reported in *A. thaliana* and S26-28 in *Lotus japonicas* (Fig. 1). The putative functions of different CaMYBs were inferred on the basis of homology to AtMYBs and the other landmark MYBs (Table 1).

Expression pattern of *CaMYB* genes in different organs and developmental stages

An expression profiling of *CaMYB* genes was performed using publically available RNA-Seq data. This revealed that *C. arietinum MYB* genes have a variety of expression patterns in different organs. All *CaMYBs* are found to be expressed in at least one or more organs/tissues, indicating the absence of pseudogenes in the dataset (Table S2). We undertook hierarchical clustering of the profiles to identify similar transcript abundance patterns and generated an expression heatmap to visualize the different expression profiles of the *CaMYBs* (Fig. 2). The expression of *CaMYB* genes was broadly clustered into five groups, i.e., I to V (Fig. 2). Group I genes were predominantly expressed in roots, while group II genes were more or less exclusively expressed in the reproductive organs of the plant. Group III genes showed elevated expression in 10-day-old seedlings. Group IV genes showed specificity toward expression in developing seeds, while group V genes are expressed in various organs.

Identification of candidate proanthocyanidin regulators (PAR)

In the context of functional analysis, we selected the biosynthesis of chickpea seed PA pigments to identify and characterize the involved R2R3-MYB proteins. Homology-based functional assignment (Table 1) suggested two CaMYBs, clustering together with landmark MYBs that have been implicated as proanthocyanin-specific activators (Fig. 1, C29 and C30). This includes especially the *A. thaliana* factor TRANSPARENT TESTA2 (TT2), controlling the accumulation of PA pigments in the seed coat. We designated these CaMYBs as CaPAR1 (CaMYB89) and CaPAR2 (CaMYB98). One CaPAR-like CaMYB (CaMYB92) was also identified in clade C30. The expression analysis revealed that transcripts of this PAR-like gene are abundant in root tissue. Thus, we subsequently selected CaPAR1 and CaPAR2 for further research.

CaPAR proteins were compared with well-characterized PA regulators from *A. thaliana* (AtTT2), *Medicago truncatula* (MtPAR), *Vitis vinifera* (VvMYBPA1, VvMYBPA2), *Diospyros kaki* (DkMYB2, DkMYB4), and *Anthurium andraeanum* (AaMYB3). A multiple amino acid alignment revealed a high conservation of the N-terminal protein region comprising the R2R3-MYB domain (Fig. 3a). The W_{61} residue found in the R3 repeat of other R2R3-MYB proteins was substituted with I_{61} in all PAR proteins (Fig. 3a and Fig. S2). The highly conserved bHLH-binding motif ([D/E][L/M]IIRLH[R/K] LLGNRWSLIA[G/K]R) was present within the R3 repeat of all two CaPARs. However, the S5-motif ([V/L][W/I]

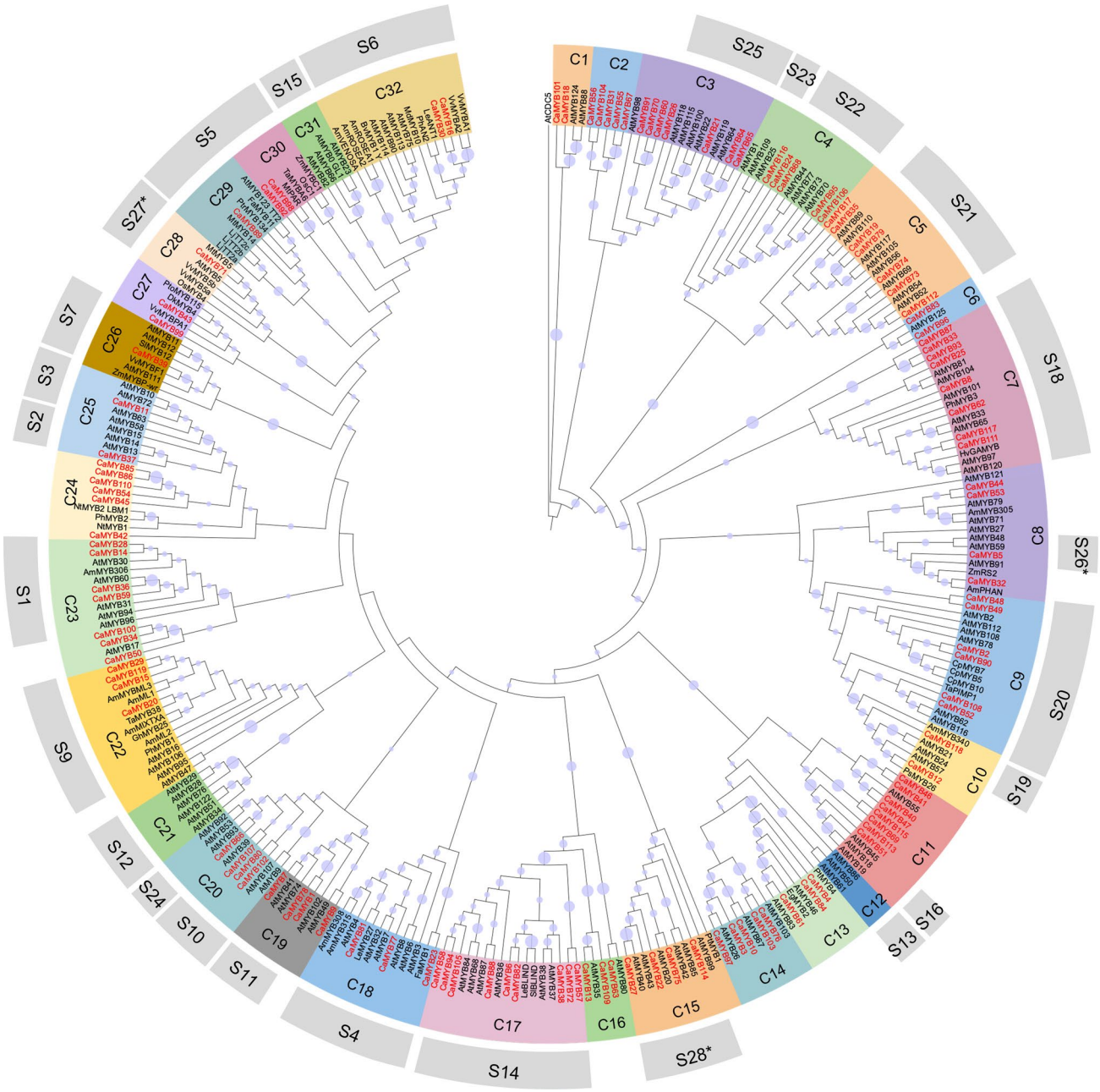

Fig. 1 Cladogram of R2R3-MYB-proteins from *C. arietinum*, *A. thaliana* and selected landmark R2R3-MYBs. RAxML method with 100 bootstraps was used for the construction of a cladogram of 119 CaMYBs, 126 AtMYBs from *A. thaliana***,** certain landmark R2R3-MYBs from different plant species and an AtCDC5 outgroup from *A. thaliana.* The figure was visualized with iTOLv6.4 (https://itol.embl.de/). Bootstrap values are shown with light blue circles at the nodes, indicating higher values in larger circles. Functionally distinct clades are labeled from C1 to C32 along with the corresponding, motif-defined sub-groups from *Arabidopsis* (S1-S25) and Lotus (S26*-S28*), shown with gray background. The cladogram does not show branch length (genetic distance) data to provide increased readability

xxKAxRCT), being described as typically for PA regulators, was not perfectly conserved in all of them. CaPAR1 showed (V[V/I]RTKAx[R/K]C[T/S]) being similar to that of AtTT2, while the S5-motif of CaPAR2 showed resemblance to that of MtPAR.. The S5-motif was absent from CaPAR-like (Fig. 3a).

CaPAR genes are co-expressed with structural PA genes in seeds

The *CaMYB* expression data (Fig. 2) indicate that *CaPAR1* and *CaPAR2* are expressed in a seed-specific manner. *CaPAR-like* gene expression was also found in developing

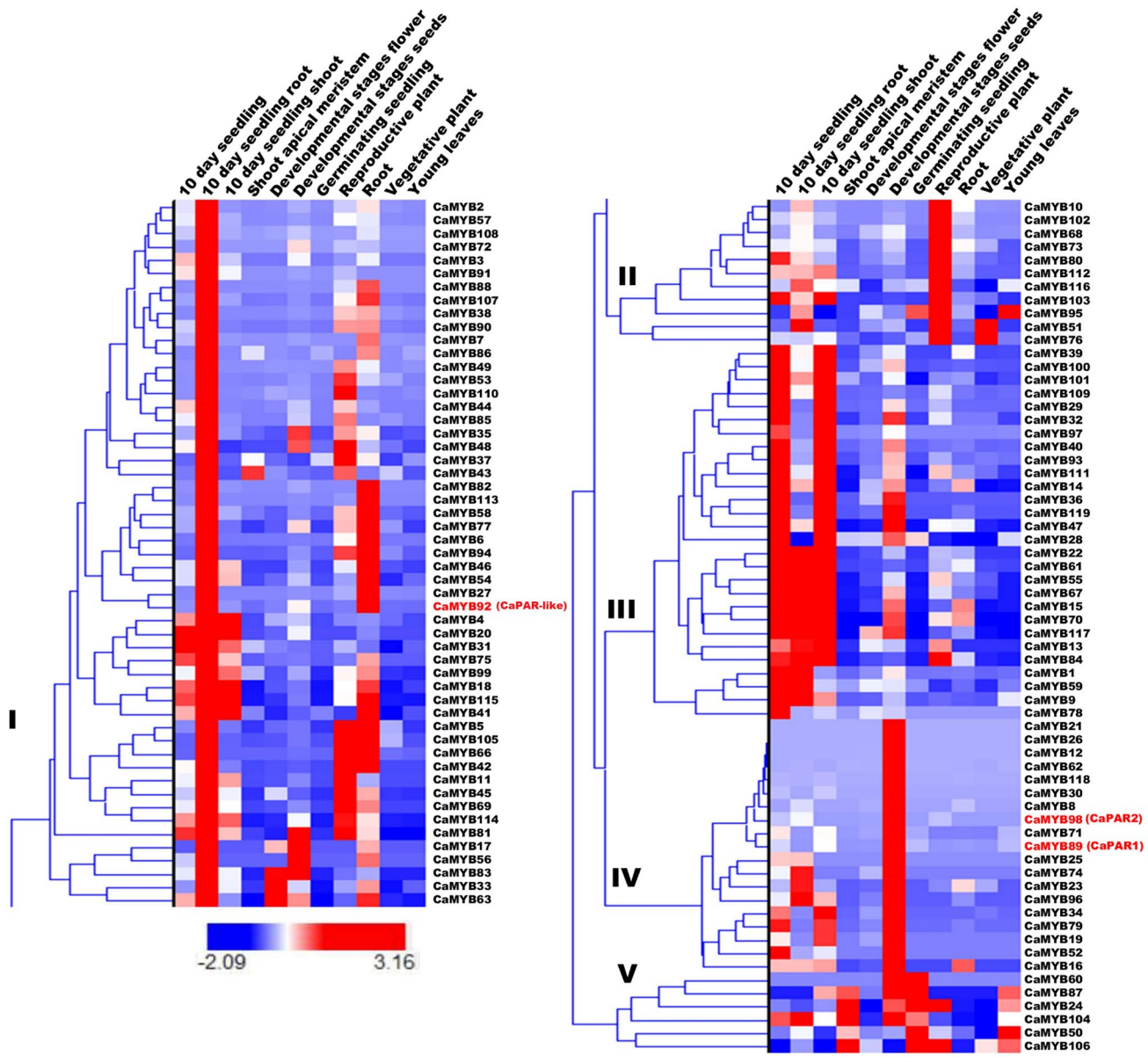

Fig. 2 Expression profiling of *CaMYB* genes. Heatmap shows the differential expression of *CaMYB* genes calculated from publically available RNA-Seq data sets of various plant parts and developmental stages of flowers and seeds. Hierarchical clustering forms five distinct *CaMYB* expression groups (termed I-V). Red to blue in the color-scale bar represents higher (red) to lower (blue) expression. *CaMYB* genes which encode PA regulators (PAR) are given in red: *CaPAR1* (*CaMYB89*) and *CaPAR2* (*CaMYB98*) within in expression group IV and *CaPAR*-like (*CaMYB92*) in expression group I

seeds, but was highest in root tissue. In addition, the expression of two anthocyanin-related R2R3-MYBs, i.e., *CaMYB16* and *CaMYB30* was also found higher in developmental stages of seeds similar to the expression of *CaPAR1* and *CaPAR2* in group IV (Fig. 2). We used an RNA-Seq dataset of developing seed stages to analyze the expression of *CaPAR* genes and two selected PA biosynthesis enzyme-coding genes *CaANR* (anthocyanidin reductase) and *CaLAR* (leucoanthocyanidin reductase). As shown in Fig. 3b, the two *CaPAR* genes show highest transcript abundance in late stages of seed development. Moreover, *CaANS* and *CaLAR*, as putative *CaPAR* target genes, show strong correlated expression (Fig. 3b).

CaPAR proteins are nuclear localized

Subcellular localization of CaPAR1, CaPAR2 and CaPAR-like was analyzed using CaPAR-YFP fusion proteins in tobacco epidermal cells. A nuclear marker NLS-RFP protein was used as localization control. Fluorescence of

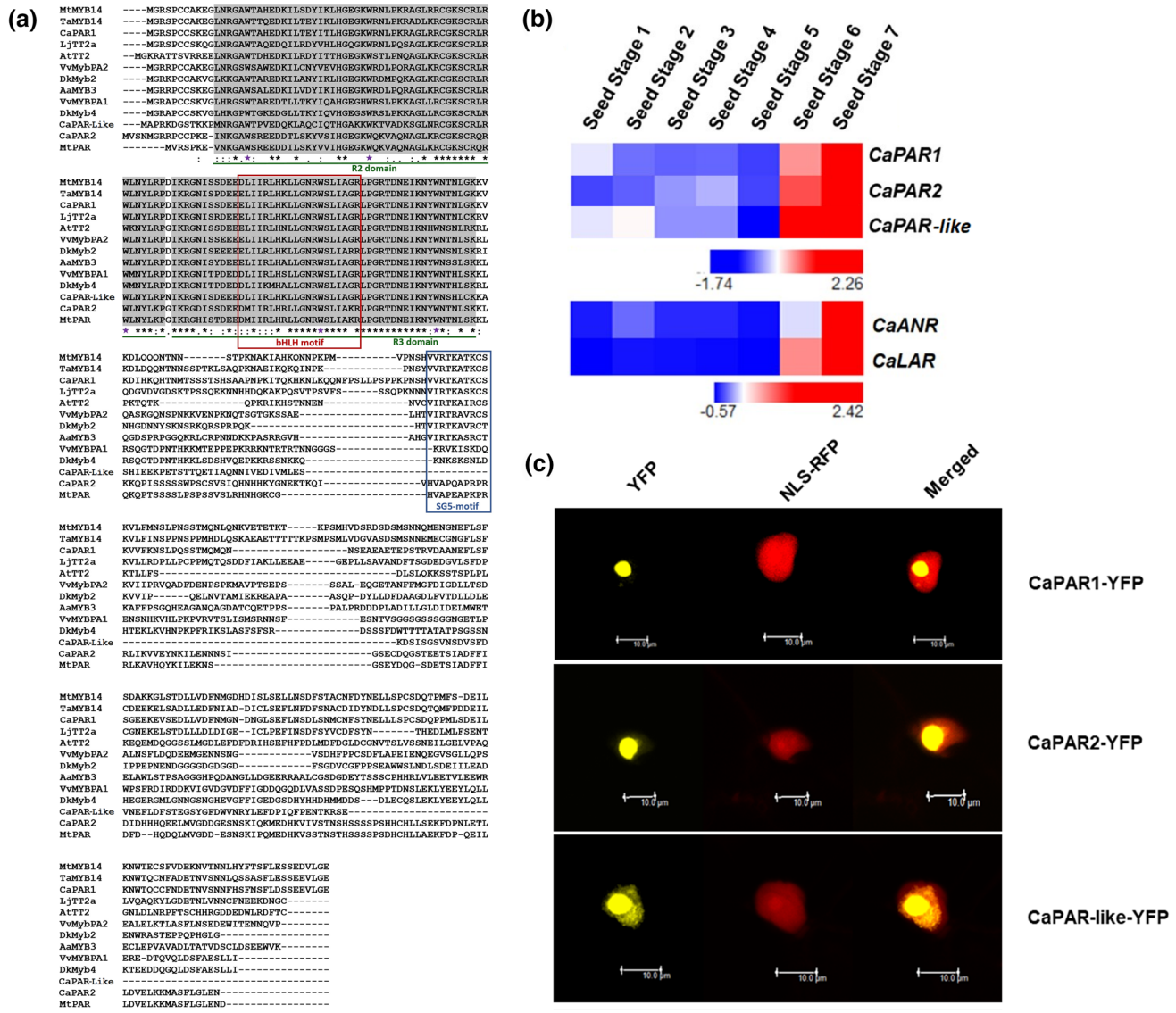

Fig. 3 Multiple sequence alignment of candidate PA regulators from *C. arietinum* (CaPAR1, CaPAR2 and CaPAR-like), digital expression and subcellular localization. **a** The R2 and R3 repeats of the MYB domain, the bHLH interaction motif and the S5 defining motif are marked according to previously characterized PA-specific R2R3-MYB regulators. Regularly spaced tryptophan residues of the MYB domain are indicated by purple asterisks. **b** Publically available RNA-Seq data sets were explored to analyze expression of *CaPAR* and *CaPAR*-like genes and candidate PA-branch biosynthesis genes (*CaANR* and *CaLAR*) in different stages of *C. arietinum* seed development. **c** Subcellular localization of CaPAR-YFP fusion proteins in Agrobacteria-infiltrated tobacco leaves analyzed by confocal microscopy. NLS-RFP was used as nuclear marker

CaPAR1-YFP, CaPAR2-YFP and CaPAR-like fusion proteins were observed exclusively in the nucleus as indicated in overlay images. We conclude from these results that CaPARs are localized in the nucleus (Fig. 3c).

PAs accumulate abundantly in the seed coat of chickpea desi type

Seeds are the dominant site of PA accumulation in *C. arietinum*. For analysis of PAs, we selected seed color contrasting chickpea types, the dark brown seeded desi type (cultivars ICC4959 and BGD256) and lighter colored kabuli types (cultivars ICC6253 and ICCV2) (Fig. 4a). To check whether the color variations are caused by different amounts of PAs, we determined the levels of soluble as well as insoluble PAs. The analysis revealed that soluble PA content was mainly restricted to the seed coat of desi types (Fig. 4b). The seed coat of the desi cultivars ICC4958 and BGD256 contain 509 ± 72 and 645 ± 82 µg/gm DW of catechin equivalent, respectively. In contrast, the seed coat of the kabuli cultivars

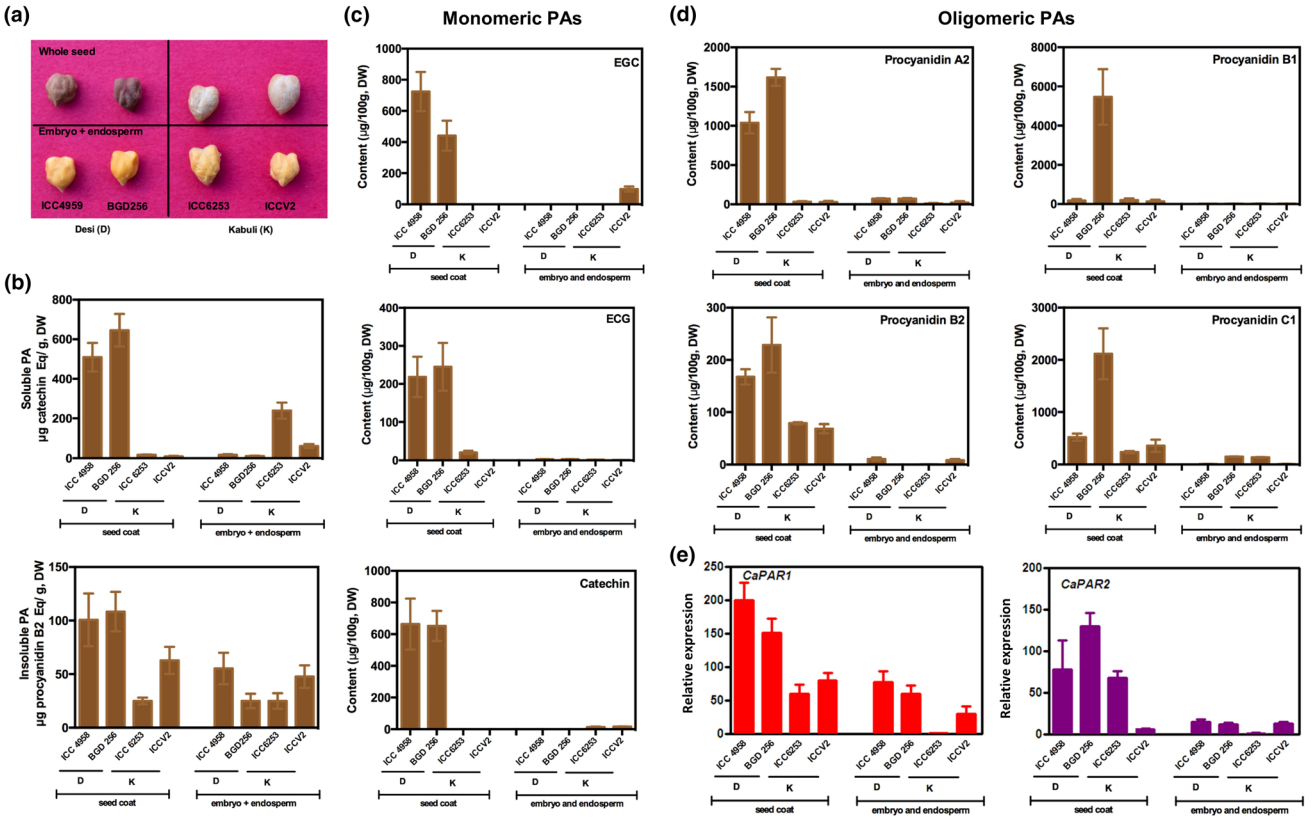

Fig. 4 Soluble and insoluble PAs in seed tissues of desi and kabuli types. **a** Mature seeds of two brown seeded desi cultivars (ICC4958 and BGD256) and two yellow seeded kabuli cultivars (ICC6253 and ICCV2) are shown; whole seeds in the upper lane and embryo and endosperm in the lower lane. **b** Spectrophotometric quantification of soluble PA content in separated seed coat and embryo + endosperm of desi and kabuli types, determined by reaction with DMACA reagent. Catechin was used as standard. Insoluble PA content in seed tissue, determined by hydrolysis with butanol-HCL reagent. Procyanidin B2 was used as standard. Values are means ± SD of two independent biological replicates with three technical replicates of each. Estimation of **c** monomeric and **d** oligomeric forms of PAs in seed coat and embryo and endosperm of color contrasting cultivars of *C. arietinum*. Values are means ± SD of two biological replicates having three technical replicates of each. EGC, epigallocatechin; ECG, epicatechin gallate. **e** Expression of *CaPAR* genes in mature seed coat and embryo + endosperm. Presented RT-qPCR data cover two independent biological replicates with three technical replicates of each and the error bars give ± SD values. CaEF-1$^{\alpha}$ and CaβTub expression was used as reference control

ICC6253 and ICCV2 contain 16.7 ± 2.4 and 9.0 ± 2.5 µg/gm DW of catechin equivalent, respectively. However, the embryo and endosperm of both, the desi and kabuli types, are deficient in soluble PAs except for the ICC6253 cultivar, in which we detected 239 ± 40 µg/gm DW of catechin equivalent. The insoluble PA content was also higher in desi seed coat than in kabuli seed coat. The seed coat of ICC4958 and BGD256 contains 101 ± 24 and 108 ± 18 µg/gm DW of procyanidin B2 equivalent, respectively. The seed coats of ICC6253 and ICCV2 contain 25 ± 3.1 and 63 ± 12.6 µg/gm DW of procyanidin B2 equivalent, respectively. While the difference in insoluble PA content in embryo and endosperm of the desi and kabuli types was insignificant (Fig. 4b and Fig. S3).

Our results of the spectrophotometric quantification made us keen to analyze both the monomeric and oligomeric forms of PAs in the similar tissues of *C. arietinum*. Therefore, we performed targeted metabolite profiling by UHPLC (Fig. S4) and LCMS (Fig. S5) for quantitative estimation of individual metabolites. We have quantified three monomeric and four oligomeric forms of PAs. The monomeric forms, i.e., epigallocatechin, epicatechin gallate and catechin were predominantly accumulated in the seed coat of ICC4958 and BGD256, respectively. In contrast to this, we could not detect the analyzed metabolites in the seed coat of kabuli types except for epicatechin gallate in ICC6253 type. Interestingly, in the embryo and endosperm of both desi and kabuli types, either we could not identify the analyzed metabolites, or the quantified metabolites were present at very low abundance except for ICCV2 where we detected epigallocatechin (Fig. 4c). In addition to this, we have also identified four different forms of procyanidins, i.e., procyanidin A2, procyanidin B1, procyanidin B2, procyanidin C1. Similar to the monomeric forms of PAs, we also detected

oligomeric forms of PAs in higher amount in the seed coat of desi type (Fig. 4d). The oligomeric forms, i.e., procyanidin A2, procyanidin B1, procyanidin B2 and procyanidin C1 were predominantly accumulated in the seed coat of ICC4958 and BGD256, respectively. We have also identified the oligomeric forms in the seed coat of kabuli types, however, the concentration of the analyzed metabolites was substantially lower than in the seed coat of desi type (Fig. 4d). Procyanidin A2, procyanidin B1, procyanidin B2, and procyanidin C1 were predominantly accumulated in the seed coat of ICC6253 and ICCV2, respectively (Fig. 4d). The analyzed metabolites were also less abundant in the embryo and endosperm of both desi and kabuli type (Fig. 4d). Taken together, it is clear that seed coat of desi types accumulated higher concentration of both the monomeric and oligomeric forms of PAs than the seed coat of kabuli types.

Expression of *CaPAR* genes is correlated with PA content in seeds

We used the dissected seed coat, embryo, and endosperm samples also for *CaPAR* gene expression analysis. RT-qPCR analysis points to higher *CaPAR* expression in the seed coat of desi type compared to kabuli type. *CaPAR* expression does not differ strongly in the embryo and endosperm of desi and kabuli types (Fig. 4b).

If the previously shown results are taken together, the correlation of gene expression and metabolite accumulation, together with amino acid similarity, indicate that the R2R3-MYB proteins encoded by *CaPAR1* and *CaPAR2* are positive regulators of PA biosynthesis in seeds of *C. arietinum.*

Ectopic expression of *CaPAR* partially restores the PA-deficient phenotype of *A. thaliana tt2-1* mutants

We used the PA-deficient *A. thaliana tt2-1* mutant for functional *in planta* complementation experiments. *CaPAR1* and *CaPAR2* were heterologously expressed in the *tt2-1* mutant under the control of the CaMV-35S promoter. PA accumulation in seeds of stable transgenic complementation lines was analyzed by DMACA staining (Fig. S6). The seeds of *Ler* (wild type control) turned blue-black, indicating high PA accumulation, while *tt2-1* mutant seeds appeared yellow. Seeds of different transgenic lines bearing the *35S*::*CaPAR1* and *35S::CaPAR2* constructs, showed dark pink staining as indication for PA accumulation (Fig. 5a).

Spectrophotometric quantification analysis indicated that *35S::CaPAR1 Arabidopsis* seeds produced approximately eight times more PAs than the *tt2-1* mutant seeds. Similarly, *35S::CaPAR2* produced approximately six times more PAs than the *tt2-1*-mutant seeds (Fig. 5b). The complementation analysis suggests that the analyzed

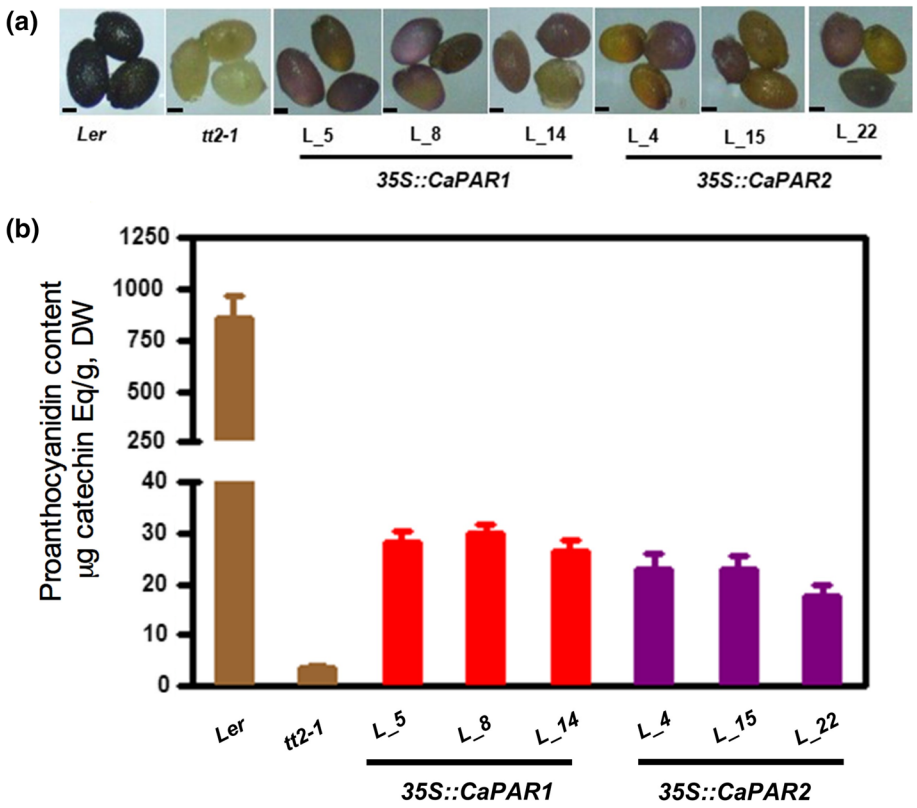

Fig. 5 *CaPAR* partially complement the *A. thaliana tt2-1* mutant phenotype. **a** Visualization of representative, DMACA stained seeds of the PA-deficient *A. thaliana tt2-1* mutant, the corresponding L*er* wild type and transgenic *35S::CaPAR1, 2* lines in *tt2-1* background. Scale bar = 5 µm. **b** Relative PA content of seeds. Values are means ± SD of two independent biological replicates each having three technical replicates

CaR2R3-MYBs have the potential to (partially) complement the PA-deficient phenotype of the *A. thaliana tt2-1* mutant, clearly support *in planta* PA regulator function.

CaPAR1 and CaPAR2 interact with CaTT8

PA biosynthesis required the formation of a regulatory MBW complex. A chickpea bHLH protein (CaTT8) with high similarity to AtTT8, being associated with petal pigmentation and seed coat color, was taken as candidate to analyze the interaction with CaPAR1- and CaPAR2 R2R3-MYBs by Y2H experiments. The Y2H results (Fig. 6a) indicate that CaPAR1 and CaPAR2 interact with CaTT8. The interaction of CaPAR1 with CaTT8 is stronger than the interaction of CaPAR2 with CaTT8. Furthermore, we performed BiFC assay in *N. benthamiana* leaves to confirm the MYB-bHLH interaction *in planta* (Fig. 6b). Here, reconstructed YFP fluorescence was seen with the CaPAR1-YFP^N with CaTT8-YFP^C pair, while it was not detected with CaPAR2-YFP^N and CaTT8-YFP^C.

Structural analysis of the *CaANR*, *CaLAR* promoters

Three and four putative MYB-binding sequences (MBS) have been identified in *proCaANR-1328* and *proCaLAR-1289*, respectively. In *proCaANR-1328*, two MBS are present between positions − 1205 to − 1200 and − 1232 to − 1227 which fits the MYBCORE motif (CNGTTR), one MRS is located between position − 996 to − 989 with the consensus motif AAMAATCT, named as MBS1, MBS2 and MBS3, respectively. In addition to this, we have also identified bHLH-binding sequences (BBS) including the MYCCONSENSUSAT motif (CANNTG), also known as R-response element (RRE), at positions − 280 to − 275, − 300 to − 295, − 618 to − 613, − 1245 to − 1240, and − 1267 to − 1262. These elements were named BBS2, BBS3, BBS4, BBS6, and BBS7, respectively. Two bHLH-binding sequences (BBS) were found at positions − 103 to − 98 and − 1230 to − 1225, both fitting the consensus motif CACGTG of palindromic G boxes. These elements were named BBS1 and BBS5, respectively. In *proCaLAR-1289*, four MBS are present between positions − 764 to − 758, − 800 to − 794, − 875

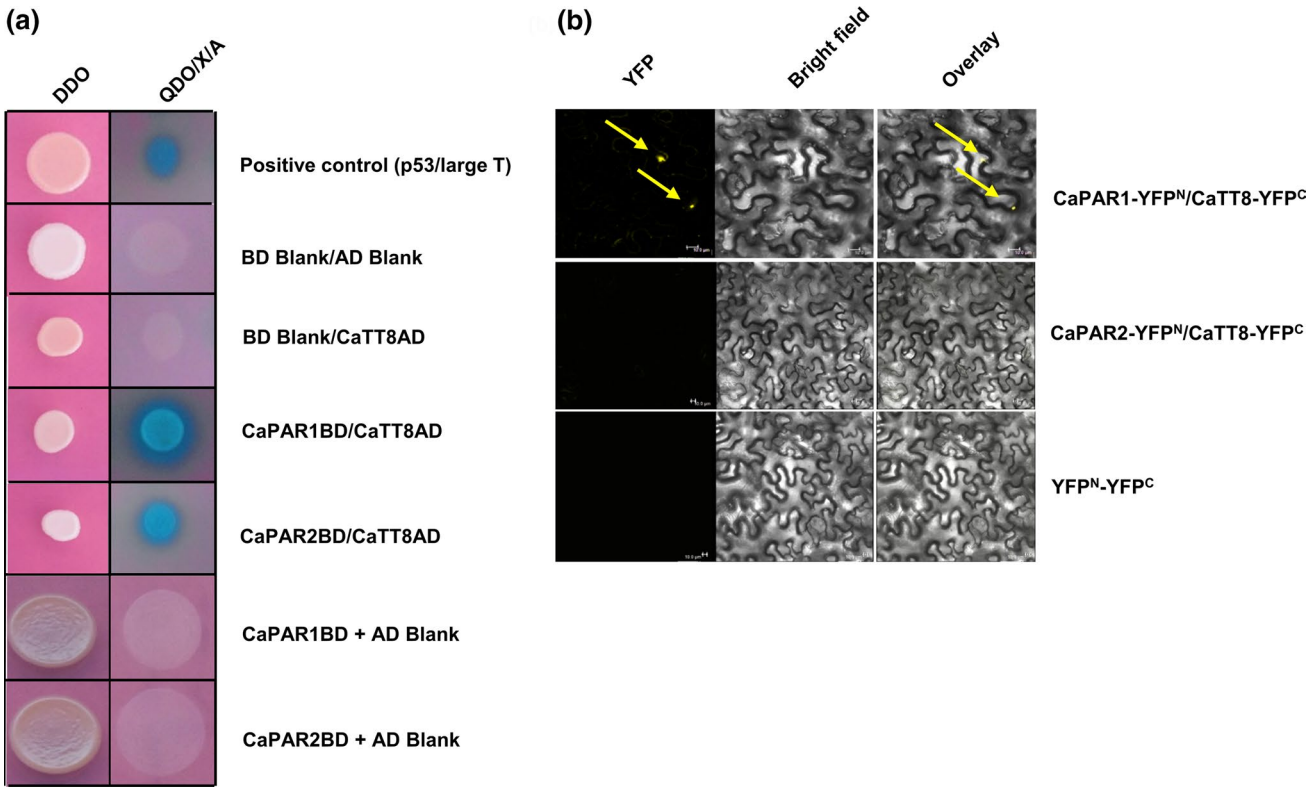

Fig. 6 Interaction of CaPAR proteins with CaTT8. **a** Yeast two-hybrid assays showed in vivo interaction between CaPAR1 and CaPAR2 with CaTT8. AD, Gal4 activation domain; BD, Gal4 DNA-binding domain; DDO, double synthetic dropout (SD)-Leu-Trp medium; QDO/X/A, quadruple SD-Leu-Trp-Ade-His + X-α-Gal + AbA medium. **b** Bimolecular fluorescence complementation (BiFC) assay indicates *in planta* interaction of CaPAR1 with CaTT8. Reconstructed YFP signals are highlighted by yellow arrows. Scale bar = 10 µm

to − 869, and − 904 to − 898. These elements were complements of the MYBCORE motif and were designated as MBS1, MBS2, MBS3 and MBS4, respectively. Identified bHLH-binding sequences (BBS) at positions − 280 to − 275, − 300 to − 295, − 618 to − 613, − 1245 to − 1240, and − 1267 to − 1262 were fitting the consensus motif CANNTG. These elements were named as BBS1 to BBS5 (Fig. 7a).

CaPAR-containing MBW complex activates the *CaANR* and *CaLAR* promoters

To analyze the transactivation potential of CaPAR1, CaPAR-2, CaMYB16, and CaMYB30 in a complex with CaTT8 on the *CaANR* and *CaLAR* promoters, we used a co-transfection-based transient *A. thaliana* protoplast expression system. Therefore, we placed the β-glucuronidase (GUS) reporter gene under the control of promoters used (Fig. 7b). Quantified, normalized GUS activity was taken as a measure of promoter activation. Effectors CaPAR1, CaPAR2,

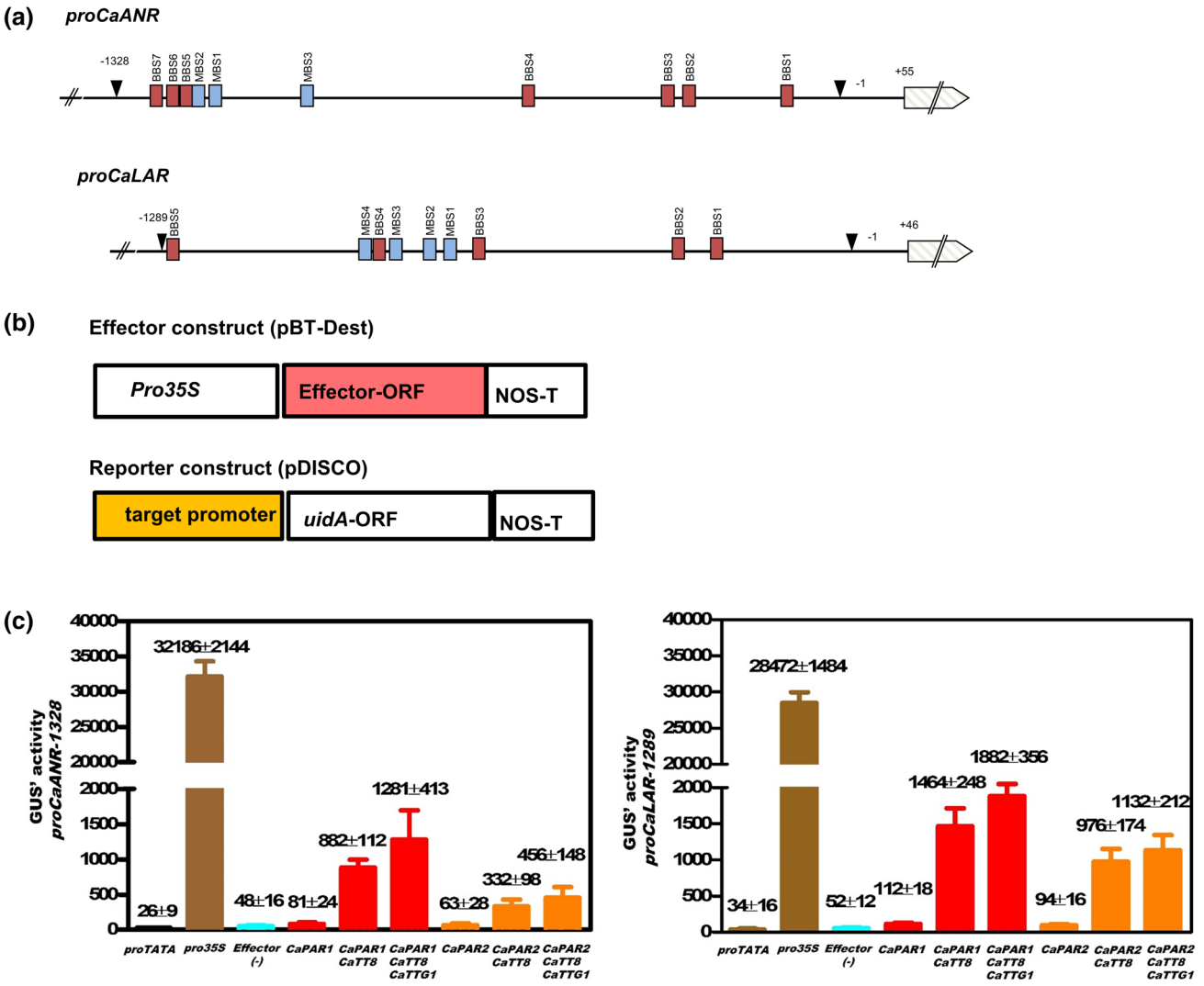

Fig. 7 CaPAR proteins function in an MBW complex and are able to activate the PAs-related *CaANR* and *CaLAR* gene promoters in *A. thaliana* protoplasts. **a** Schematic presentation of the *C. arietinum ANR* and *LAR* promoters. Putative *cis*-acting elements are indicated: MBS (light blue) are putative MYB-binding sites and BBS (red) are putative bHLH-binding sites. The black "triangles" mark (− 1) transcription start site (TSS), position + 55 and + 46 shows the translation start sites of *proCaANR* and *proCaLAR*, respectively. **b** Schematic diagram of the effector and reporter constructs used in the co-transfection experiment. **c** Results from co-transfection experiments in *A. thaliana* protoplasts. A GUS-fused 1328 bp *CaANR* and 1289 bp *CaLAR* promoter fragment (reporter) was assayed for its responsiveness to various 35S promoter-driven effectors CaPAR1, CaPAR2, CaTT8 and CaTTG1, either alone or in combinations. The figure shows ean normalized GUS' activity resulting from the influence of tested effector proteins on reporters *proCaANR-1328* and *proCaLAR1289*. Data from a set of six replicates are presented

CaMYB16, CaMYB30 and CaTT8 were driven by 35S promoters (Fig. 7b). Since a functional MBW complex also needed a WD40-repeat protein, we included CaTTG1 (Fig. S7) in this experiment. The co-transfection results suggested that none of CaPARs is able to trans-activate *proCaANR* and *CaLAR* alone. However, in combination with CaTT8 and even at a higher level with additional CaTTG1, the CaPAR R2R3-MYBs showed the potential to transactivate *proCaANR-1328* and *proCaLAR-1289* promoters at different levels (Fig. 7c). In *proCaANR-1328*, the highest transactivation capacity was found for CaPAR1 in combination with CaTT8. Further addition of CaTTG1 enhanced the *CaANR* promoter activity by a factor of 1.45. The transactivation of CaPAR2, in combination with CaTT8 gave strong GUS´ activity which further increased by a factor of 1.51 with CaTTG1. In *proCaLAR-1289*, the highest transactivation capacity was found for CaPAR1 in combination with CaTT8. Further addition of CaTTG1 enhanced the *CaLAR* promoter activity by a factor of 1.28. The transactivation of CaPAR2 in combination with CaTT8 and CaTTG1 increased the GUS´ activity by a factor of 1.15. To check the involvement of additional IV R2R3-MYBs in the regulation of the PA biosynthesis, we considered additional IV MYBs, i.e., CaMYB16 and CaMYB30 for protoplast assays. The co-transfection results suggested that neither CaMYB16 nor CaMYB30 has the potential to transactivate *proCaANR* and *proCaLAR* (Fig. S8). In summary, these results indicate that two CaPARs form *in planta* functional MBW complexes with CaTT8 and CaTTG1. These complexes are able to activate the *CaANR* and the *CaLAR* promoters.

Co-expression analysis of chickpea PA regulator MBW complex encoding genes

A total of 1197 genes were found to be co-expressed with genes encoding components of the *C. arietinum* PA regulating MBW complex. In detail, 36 genes are co-expressed with *CaPAR1*, 192 with *CaPAR2*, 294 with *CaTT8,* and 628 with *CaTTG1*. Genes annotated to encode a flavanone 3-hydroxylase (F3H), a chalcone-flavanone isomerase (CHI), and a MATE efflux protein were co-expressed with all five MBW complex genes. Some genes showed co-expression with selective MBW complex genes. For instance, genes annotated to encode RING/U-box and zinc finger family proteins are co-expressed with *CaPAR2*, *CaTT8* and *CaTTG1*; ABA-responsive element-binding protein (AREB3), auxin-responsive factor AUX/IAA-like protein, transferases like O-methyltransferase (O-MTase), S-adenosyl-L-methionine-dependent methyltransferases (SAM MTase), UDP-Glycosyltransferase (UGT) and protein kinase (PK) encoding genes are co-expressed with *CaTT8* and *CaTTG1*; a gene annotated to encode a jasmonate ZIM -domain protein8 (JAZ8) homolog is co-expressed with *CaPAR1* and *CaPAR2*, etc. (Fig. 8; Table S3). Apart from this, many of the *C. arietinum* genes annotated to encode different enzymes, kinases, and transporters were found to be co-expressed with one of the PA regulator MBW complex genes (Table S3).

Discussion

In this study, we aimed to identify and characterize proteins transcriptionally regulating the PA biosynthesis in *C. arietinum.* Harnessing knowledge from other plant species, we started to identify R2R3-MYB-type PA regulators, known as the players that determine the specificity for the different branches of flavonoid biosynthesis. In an initially broad approach, we have captured and pictured the entire *R2R3-MYB* gene family from *C. arietinum* (Table 1, Fig. 1). We identified a total of 119 *CaMYB* genes encoding MYB proteins with more than one repeats, including 113 CaR2R3-MYB encoding genes. This number is somewhere in the middle of the range known from other plant species, ranging from 45 in *Ginkgo biloba* (Liu et al. 2017), 88 in *Oryza sativa* (Katiyar et al. 2012), over 126 in *A. thaliana* (Stracke et al. 2001), 192 in *Populus trichocarpa* (Wilkins et al. 2009), and 285 in *Musa acuminata* (Pucker et al. 2020). Variability in the number of *R2R3-MYB* genes might be attributed to the ploidy level of species and the number of gene duplication events in the different genomes evolution. For example, the high number of *R2R3-MYB* genes in *M. acuminata* is thought to be caused by three whole-genome duplications (γ 100 Myr ago and α, β 65 Myr ago) that occurred during *Musa* genome evolution (Pucker et al. 2020). The number of 119 *R2R3-MYB* genes in *C. arietinum* can be explained by the genome duplication event that occurred at the base of the Faboideae, where the galegoid (*M. truncatula*, *L. japonicus* and *C. arietinum*) and millettioid (soybean, pigeonpea) clades separated ~54 Myr ago (Lavin et al. 2005). Within the *C. arietinum* genome, two *R2R3-MYB* gene duplications were found including one segmental duplication and one tandem duplication (Fig. S1). This is a small number compared with *R2R3-MYB* gene duplications reported for other species, with three tandem duplications in the *B. vulgaris* genome (Stracke et al. 2014) and 27 segmental duplications and eight tandem duplications in the *G. raimondii* genome (He et al. 2016). These consistently observed gene duplication events are the driving force that leads to an ongoing expansion of the *R2R3-MYB* gene families (Chang et al. 2020).

Homology-based functional classification of the *C. arietinum* R2R3-MYB proteins (Fig. 1, Table 1) led to assignment of a potential function to each CaMYB protein. Interestingly, the cladogram does point to a *C. arietinum*-specific clade (C2) of MYB proteins, which could indicate that there are functionalities regulated by R2R3-MYB

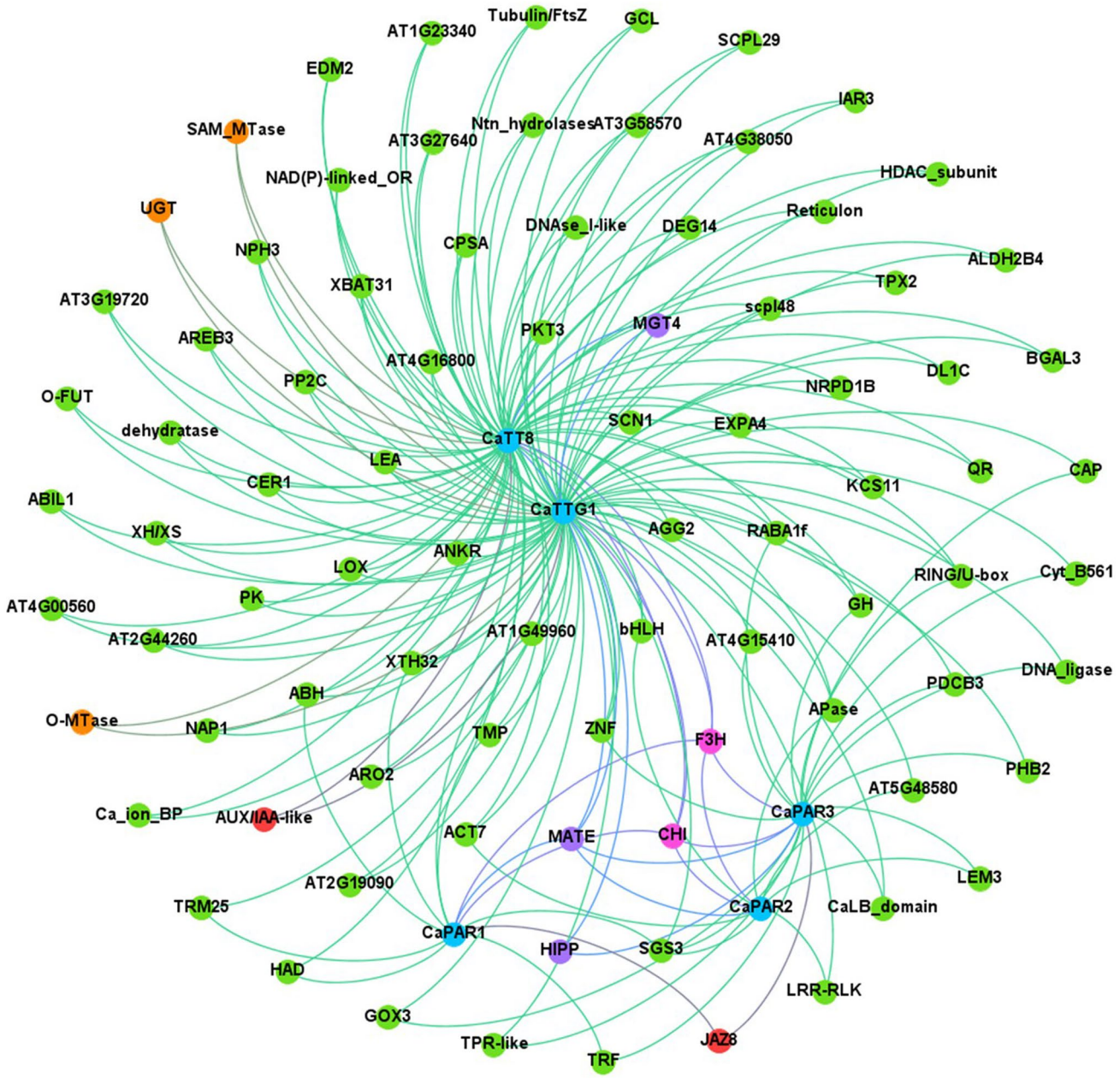

Fig. 8 Co-expression network of the *C. arietinum* MBW complex genes**.** Co-expression analysis of *CaPAR1, CaPAR2, CaTT8* and *CaTTG1* (MBW complex genes) with other genes in *C. arietinum* genome. The genes co-expressed with at least two of the MBW complex genes are shown with different colors, where genes encoding flavonoid biosynthesis-related proteins (pink), proteins related to hormone activity (red), methyl- and glycosyl-transferases (orange), transporters (purple) and other proteins (green) are connected with respective color lines to their co-expressed MBW complex genes (blue). Genes are labeled with already known abbreviations for the annotated chickpea genes or with the gene ID of homologues *A. thaliana* genes where the chickpea annotation is patchy. Details are given in Table S3

proteins found exclusively in chickpea and maybe closely related species. But this assumption would need to be tested in a larger phylogenetic context and followed up with additional experiments, which is not the intention of this work. *R2R3-MYB* genes with diverse role in various biological processes including specialized metabolism, growth and development, defense response, and response to stress conditions are grouped together based on their expression patterns. Since anthocyanins and PAs both share several common biosynthesis genes and their regulation often overlaps (Wang et al. 2018), the expression of anthocyanin-specific *R2R3-MYB* genes, i.e., *CaMYB16* and *CaMYB30* was corroborated with the expression of C*aPAR1* and *CaPAR2*.

The CaMYBs, which are placed in the clade of PA regulators (CaPARs), have been analyzed in depth in this work. We present strong evidence, that the CaPARs are involved in the regulation of PA biosynthesis in *C. arietinum.* The expression pattern of two *CaPAR* genes indicated, that their expression is confined to the seeds (Fig. 3b). Further analyses, suggested that their expression is restricted to the seed coat (Fig. 4e), which is the site of PA accumulation (Fig. 4b). The expression of both the *CaPARs* is highest in the developing seeds, which indicates that some upstream regulators might be working transcriptionally or post-transcriptionally. Examples for such regulators are, e.g., the WRKY-type protein MdWRKY41, which inhibits the expression of *MdMYB12* resulting in the repression of PA synthesis in apple (Mao et al. 2021) or PtrBBX23, which activates the expression of *PtrMYB115* leading to the activation of PA accumulation in poplar (Li et al. 2021). In grapevine a microRNA encoded peptide, miPEP164c, enhances the transcription of its own transcript to down-regulate PA synthesis in berries. miR164c targets *VvMYBPA1* via degradation through post-transcriptional gene silencing, leading to decreased expression of *LAR* and *ANR*, resulting in decreased PA accumulation (Vale et al. 2021). In poplar, a histone H3K9 demethylase JMJ25 epigenetically regulates PA synthesis. JMJ25 demethylates H3K9me2 in the chromatin body of *MYB182* to enhance its expression leading to down regulation of PA synthesis. It also affects DNA methylation at *MYB182* locus which helps in the efficient expression (Fan et al. 2018). Taking this knowledge, we speculate that some upstream regulation also exists in *C. arietinum*, modulating *CaPAR1* and*CaPAR2* expression. Maybe the genes being co-expressed with *CaPAR* genes, like *AREB3* and *ZNF* act as a transcriptional inducers in chickpea.

However, the expressions of *CaPAR2* in ICC4958 and ICC6253 do not appear to differ. The previous study by Penmetsa et al. (2016) showed the involvement of natural allele of bHLH responsible for flower color (anthocyanins) and seed coat color (PAs). Thus, it can be explained by the fact that the anthocyanin and PA biosynthesis in chickpea is regulated by the formation of a functional MBW complex to regulate the biosynthesis pathway. Further, the transcript levels of *CaPAR* genes were well correlated with the soluble and insoluble PA content with significant higher content of both forms of PAs in seed coat of desi type. Interestingly, embryo and endosperm of neither desi nor kabuli accumulate substantial amount of PA which suggests cell specificity of PAs in chickpea seeds. Similar to our observation, *MtPAR* expression and accumulation of both soluble and insoluble PAs is restricted to the seed coat of *Medicago truncatula* (Verdier et al. 2012). Our metabolomics data suggest another interesting observation that the seed coat of the desi types accumulates both catechin and epicatechin as soluble forms of PAs. This finding is quite opposite to *Arabidopsis* seeds which exclusively contain only epicatechin unit as *Arabidopsis* lacks the functional LAR gene (Tanner et al. 2003; Routaboul et al. 2006). Although *M. truncatula* harbors a LAR gene in its genome, this plant accumulates almost exclusively epicatechin units (Pang et al. 2007). Our finding suggests that both catechin and epicatechin have a role in tannin accumulation in the seed coat of desi types. It would be interesting to elucidate the role of *CaANR* and *CaLAR* genes in chickpea which is involved in the biosynthesis of epicatechin and catechin polymer, respectively. Though our study is confined to mature seeds, further study is needed for distribution of PAs in different developmental stages. It was inferred in our study that seed color difference is mainly due to different PA content of seed coats. Ectopic expression of *CaPAR* in PA-deficient *A. thaliana* seeds partially restored the PA biosynthesis in *transparent testa* seeds which validated the CaPAR function.

In this study, we set out to broaden the knowledge on regulatory proteins involved in PA biosynthesis. Yeast two-hybrid and BiFC results were consistent with previous findings where CaPARs interact with CaTT8 both in heterologous system and *in planta* for the formation of MBW complexes. Further, the co-transfection analysis suggested that CaPAR interacts with CaTT8 and CaTTG1 to form a complex which then trans-activates the promoters of *CaANR* and *CaLAR*. Moreover, our study provides a link between CaPAR proteins and a previously characterized bHLH (CaTT8; Penmetsa et al. 2016) by protein–protein and DNA protein interactions.

It would be interesting to further dissect the regulatory network of PA biosynthesis in chickpea and to unveil the entire regulon of the MBW complex. Our co-expression data proposed several genes which could participate in the formation and accumulation of PAs along with the MBW complex. The role of *F3H* and *CHI* genes has been reported in the biosynthesis of flavonoids, while MATE transporters are well known for flavonoid transport (Burbulis and Winkel-Shirley 1999; Marinova et al. 2007). An ABA-induced DkbZIP5 protein regulates the expression of *DkMYB4* to induce PA synthesis in persimmon fruit (Akagi et al. 2012). Further, the significance of ABA in the synthesis of the anthocyanin has been discussed in sweet cherry fruit development, which also supported the co-expression AREB proteins. ABA plays a major role in salt, drought, cold and osmotic stress tolerance. ABA-induced anthocyanin synthesis has been a major marker for abiotic stress in plants. Co-expression of ABA-related genes suggests that it may play a role in anthocyanin accumulation in *C. arietinum*. Various co-expressed genes, such as flavonoid-specific glycosyltransferases (*GTs*), phenylalanine ammonia lyase, etc. associatively function with different enzyme complexes in flavonoid biosynthesis (Burbulis and Winkel-Shirley 1999). Therefore, the presented co-expression data hint to several genes which could be directly

or indirectly involved in the flavonoid biosynthesis pathway. Of course, these putative involvements have to be proven in future experiments.

In the effort to modulate PA biosynthesis pathway by biotechnological approaches, the use of regulatory proteins has gained attention during the last years. This is because of their ability to modulate the whole biosynthetic pathway instead of a single step. For example, the PA biosynthesis was successfully enhanced in alfalfa by targeting *MtPAR* (Verdier et al. 2012). Similarly, the knowledge gained in our study could be used for genetic improvement of PA biosynthesis in *C. arietinum*, with CaPAR1 and CaTT8 as primary targets.

Author contribution statement

AP conceived the idea and designed the research. RR, ST, BP and RS performed in silico analysis. RR and JN performed the experiments. RR, RS and AP wrote the manuscript. All the authors have read and approved the manuscript for submission.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-022-03979-z.

Acknowledgements This work was supported by the core grant of National Institute of Plant Genome Research and Department of Biotechnology grant (BT/PR36694/NNT/28/1722/2020) to AP. RR and JN acknowledge Council of Scientific and Industrial Research, Government of India, for Senior Research Fellowships. ST acknowledges Department of Biotechnology for Research Associate Fellowship. The authors are thankful to DBT-eLibrary Consortium (DeLCON) for providing access to e-resources. We acknowledge the Metabolome facility (BT/ INF/22/SP28268/2018) at NIPGR for phytochemical analysis.

Data availability All data supporting the findings of this study are available within the paper and within the supplementary data published online.

Declarations

Conflict of interest Authors declare no conflict of interest.

References

Akagi T, Katayama-Ikegami A, Kobayashi S et al (2012) Seasonal abscisic acid signal and a basic leucine zipper transcription factor, DkbZIP5, regulate proanthocyanidin biosynthesis in persimmon fruit. Plant Physiol 158:1089–1102. https://doi.org/10.1104/pp.111.191205

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

An X-H, Tian Y, Chen K-Q et al (2015) *MdMYB9* and *MdMYB11* are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. Plant Cell Physiol 56:650–662. https://doi.org/10.1093/pcp/pcu205

Appelhagen I, Thiedig K, Nordholt N et al (2014) Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. Planta 240:955–970. https://doi.org/10.1007/s00425-014-2088-0

Bajaj D, Das S, Upadhyaya HD, Ranjan R, Badoni S, Kumar V, Tripathi S, Gowda CL, Sharma S, Singh S, Tyagi AK (2015) A genome-wide combinatorial strategy dissects complex genetic architecture of seed coat color in chickpea. Front Plant Sci 17:6–979. https://doi.org/10.3389/fpls.2015.00979

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. Proc Intern AAAI Conference on Web and Social Media 3(1): 361–362. https://ojs.aaai.org/index.php/ICWSM/article/view/13937

Baudry A, Heim MA, Dubreucq B et al (2004) TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 39:366–380. https://doi.org/10.1111/j.1365-313X.2004.02138.x

Bogs J, Jaffé FW, Takos AM et al (2007) The grapevine transcription factor *VvMYBPA1* regulates proanthocyanidin synthesis during fruit development. Plant Physiol 143:1347–1361. https://doi.org/10.1104/pp.106.093203

Bondonno NP, Dalgaard F, Kyrø C et al (2019) Flavonoid intake is associated with lower mortality in the Danish diet cancer and health cohort. Nat Commun 10:3651. https://doi.org/10.1038/s41467-019-11622-x

Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. Proc Natl Acad Sci USA 96:12929–12934. https://doi.org/10.1073/pnas.96.22.12929

Carmona-Gutierrez D, Zimmermann A, Kainz K et al (2019) The flavonoid 4,4'-dimethoxychalcone promotes autophagy-dependent longevity across species. Nat Commun 10:651. https://doi.org/10.1038/s41467-019-08555-w

Chang X, Xie S, Wei L et al (2020) Origins and stepwise expansion of R2R3-MYB transcription factors for the terrestrial adaptation of plants. Front Plant Sci 11:575360. https://doi.org/10.3389/fpls.2020.575360

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743. https://doi.org/10.1046/j.1365-313x.1998.00343.x

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190. https://doi.org/10.1101/gr.849004

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol 133:462–469. https://doi.org/10.1104/pp.103.027979

Deluc L, Barrieu F, Marchive C et al (2006) Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant Physiol 140:499–511. https://doi.org/10.1104/pp.105.067231

Deluc L, Bogs J, Walker AR et al (2008) The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol 147:2041–2053. https://doi.org/10.1104/pp.108.118919

Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Dubos C, Stracke R, Grotewold E et al (2010) MYB transcription factors in *Arabidopsis*. Trends Plant Sci 15:573–581. https://doi.org/10.1016/j.tplants.2010.06.005

Fan D, Wang X, Tang X et al (2018) Histone H3K9 demethylase JMJ25 epigenetically modulates anthocyanin biosynthesis in poplar. Plant J 96:1121–1136. https://doi.org/10.1111/tpj.14092

Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18:53–63. https://doi.org/10.1093/dnares/dsq028

Gesell A, Yoshida K, Tran LT, Constabel CP (2014) Characterization of an apple TT2-type R2R3 MYB transcription factor functionally similar to the poplar proanthocyanidin regulator PtMYB134. Planta 240:497–511. https://doi.org/10.1007/s00425-014-2098-y

Haak M, Vinke S, Keller W et al (2018) High quality de novo transcriptome assembly of *Croton tiglium*. Front Mol Biosci 5:62. https://doi.org/10.3389/fmolb.2018.00062

Hartmann U, Sagasser M, Mehrtens F et al (2005) Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol 57:155–171. https://doi.org/10.1007/s11103-004-6910-0

He Q, Jones DC, Li W et al (2016) Genome-wide identification of R2R3-MYB genes and expression analyses during abiotic stress in *Gossypium raimondii*. Sci Rep 6:22980. https://doi.org/10.1038/srep22980

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic Acids Res 27:297–300. https://doi.org/10.1093/nar/27.1.297

Jin H, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol 41:577–585. https://doi.org/10.1023/a:1006319732410

Katiyar A, Smita S, Lenka SK et al (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. BMC Genomics 13:544. https://doi.org/10.1186/1471-2164-13-544

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Khandal H, Gupta SK, Dwivedi V et al (2020) Root-specific expression of chickpea cytokinin oxidase/dehydrogenase 6 leads to enhanced root growth, drought tolerance and yield without compromising nodulation. Plant Biotechnol J 18:2225–2240. https://doi.org/10.1111/pbi.13378

Koncz C, Schell J (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol Gen Genet 204:383–396. https://doi.org/10.1007/BF00331014

Koyama K, Numata M, Nakajima I et al (2014) Functional characterization of a new grapevine MYB transcription factor and regulation of proanthocyanidin biosynthesis in grapes. J Exp Bot 65:4433–4449. https://doi.org/10.1093/jxb/eru213

Kranz H, Scholz K, Weisshaar B (2000) c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. Plant J 21:231–235. https://doi.org/10.1046/j.1365-313x.2000.00666.x

Kumar K, Purayannur S, Kaladhar VC et al (2018) mQTL-seq and classical mapping implicates the role of an AT-HOOK MOTIF containing nuclear localized (AHL) family gene in *Ascochyta* blight resistance of chickpea. Plant Cell Environ 41:2128–2140. https://doi.org/10.1111/pce.13177

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594. https://doi.org/10.1080/10635150590947131

Li C, Pei J, Yan X et al (2021) A poplar B-box protein PtrBBX23 modulates the accumulation of anthocyanins and proanthocyanidins in response to high light. Plant Cell Environ 44:3015–3033. https://doi.org/10.1111/pce.14127

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu X, Yu W, Zhang X et al (2017) Identification and expression analysis under abiotic stress of the R2R3-MYB genes in *Ginkgo biloba* L. Physiol Mol Biol Plants 23:503–516. https://doi.org/10.1007/s12298-017-0436-9

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2-\Delta\Delta CT$ method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Mao Z, Jiang H, Wang S et al (2021) The MdHY5-MdWRKY41-MdMYB transcription factor cascade regulates the anthocyanin and proanthocyanidin biosynthesis in red-fleshed apple. Plant Sci 306:110848. https://doi.org/10.1016/j.plantsci.2021.110848

Marinova K, Pourcel L, Weder B et al (2007) The *Arabidopsis* MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19:2023–2038. https://doi.org/10.1105/tpc.106.046029

Martin K, Kopperud K, Chakrabarty R et al (2009) Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. Plant J 59:150–162. https://doi.org/10.1111/j.1365-313X.2009.03850.x

Naik J, Rajput R, Pucker B et al (2021) The R2R3-MYB transcription factor MtMYB134 orchestrates flavonol biosynthesis in *Medicago truncatula*. Plant Mol Biol 106:157–172. https://doi.org/10.1007/s11103-021-01135-x

Nakagawa T, Kurose T, Hino T et al (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng 104:34–41. https://doi.org/10.1263/jbb.104.34

Nesi N, Jond C, Debeaujon I et al (2001) The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13:2099–2114. https://doi.org/10.1105/tpc.010098

Ogata K, Kanei-Ishii C, Sasaki M et al (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. Nat Struct Biol 3:178–187. https://doi.org/10.1038/nsb0296-178

Pang Y, Peel GJ, Wright E et al (2007) Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. Plant Physiol 145:601–615. https://doi.org/10.1104/pp.107.107326

Pang Y, Abeysinghe ISB, He J et al (2013) Functional characterization of proanthocyanidin pathway enzymes from tea and their application for metabolic engineering. Plant Physiol 161:1103–1116. https://doi.org/10.1104/pp.112.212050

Parween S, Nawaz K, Roy R et al (2015) An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). Sci Rep 5:12806. https://doi.org/10.1038/srep12806

Pipas JM, Levine AJ (2001) Role of T antigen interactions with p53 in tumorigenesis. Semin Cancer Biol 11:23–30. https://doi.org/10.1006/scbi.2000.0343

Pucker B, Holtgräwe D, Rosleff Sörensen T et al (2016) A de novo genome sequence assembly of the *Arabidopsis thaliana* accession niederzenz-1 displays presence/absence variation and strong synteny. PLoS ONE 11:e0164321. https://doi.org/10.1371/journal.pone.0164321

Pucker B, Pandey A, Weisshaar B, Stracke R (2020) The R2R3-MYB gene family in banana (*Musa acuminata*): genome-wide identification, classification and expression patterns. PLoS ONE 15:e0239275. https://doi.org/10.1371/journal.pone.0239275

Quattrocchio F, Baudry A, Lepiniec L, Grotewold E (2006) The regulation of flavonoid biosynthesis. In: Grotewold E (ed) The science of flavonoids. Springer, New York, NY, pp 97–122

Rajkumar MS, Gupta K, Khemka NK, Garg R, Jain M (2020) DNA methylation reprogramming during seed development and its functional relevance in seed size/weight determination in chickpea. Commun Biol 3:1–3. https://doi.org/10.1038/s42003-020-1059-1

Rajput R, Naik J, Stracke R, Pandey A (2022) Interplay between R2R3 MYB-type activators and repressors regulates proanthocyanidin biosynthesis in banana (*Musa acuminata*). New Phytol. https://doi.org/10.1111/nph.18382

Ravaglia D, Espley RV, Henry-Kirk RA et al (2013) Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biol 13:68. https://doi.org/10.1186/1471-2229-13-68

Routaboul J-M, Kerhoas L, Debeaujon I et al (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta 224:96–107. https://doi.org/10.1007/s00425-005-0197-5

Schwinn KE, Ngo H, Kenel F et al (2016) The onion (*Allium cepa* L.) R2R3-MYB gene MYB1 regulates anthocyanin biosynthesis. Front Plant Sci 7:1865. https://doi.org/10.3389/fpls.2016.01865

Seo J, Gordish-Dressman H, Hoffman EP (2006) An interactive power analysis tool for microarray hypothesis testing and generation. Bioinformatics 22:808–814

Shelton D, Stranne M, Mikkelsen L et al (2012) Transcription factors of lotus: regulation of isoflavonoid biosynthesis requires coordinated changes in transcription factor activity. Plant Physiol 159:531–547. https://doi.org/10.1104/pp.112.194753

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr Opin Plant Biol 4:447–456

Stracke R, Ishihara H, Huep G et al (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J 50:660–677. https://doi.org/10.1111/j.1365-313X.2007.03078.x

Stracke R, Jahns O, Keck M et al (2010) Analysis of production of flavonol glycosides-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. New Phytol 188:985–1000. https://doi.org/10.1111/j.1469-8137.2010.03421.x

Stracke R, Holtgräwe D, Schneider J et al (2014) Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). BMC Plant Biol 14:249. https://doi.org/10.1186/s12870-014-0249-8

Stracke R, Thiedig K, Kuhlmann M, Weisshaar B (2016) Analyzing synthetic promoters using *Arabidopsis* protoplasts. Methods Mol Biol 1482:67–81. https://doi.org/10.1007/978-1-4939-6396-6_5

Terrier N, Torregrosa L, Ageorges A et al (2009) Ectopic expression of *VvMybPA2* promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiol 149:1028–1041. https://doi.org/10.1104/pp.108.131862

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680. https://doi.org/10.1093/nar/22.22.4673

Trezzini GF, Horrichs A, Somssich IE (1993) Isolation of putative defense-related genes from *Arabidopsis thaliana* and expression in fungal elicitor-treated cells. Plant Mol Biol 21:385–389. https://doi.org/10.1007/BF00019954

Vale M, Rodrigues J, Badim H et al (2021) Exogenous application of non-mature miRNA-encoded miPEP164c inhibits proanthocyanidin synthesis and stimulates anthocyanin accumulation in grape berry cells. Front Plant Sci 12:706679. https://doi.org/10.3389/fpls.2021.706679

Varma Penmetsa R, Carrasquilla-Garcia N, Bergmann EM et al (2016) Multiple post-domestication origins of kabuli chickpea through allelic variation in a diversification-associated transcription factor. New Phytol 211:1440–1451. https://doi.org/10.1111/nph.14010

Varshney RK, Song C, Saxena RK et al (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240–246. https://doi.org/10.1038/nbt.2491

Verdier J, Zhao J, Torres-Jerez I et al (2012) *MtPAR* MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. Proc Natl Acad Sci USA 109:1766–1771. https://doi.org/10.1073/pnas.1120916109

Wang ZY, Kenigsbuch D, Sun L et al (1997) A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis Lhcb* gene. Plant Cell 9:491–507. https://doi.org/10.1105/tpc.9.4.491

Wang N, Xu H, Jiang S et al (2017) *MYB12* and *MYB22* play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. niedzwetzkyana). Plant J 90:276–292. https://doi.org/10.1111/tpj.13487

Wang N, Qu C, Jiang S et al (2018) The proanthocyanidin-specific transcription factor *MdMYBPA1* initiates anthocyanin synthesis under low-temperature conditions in red-fleshed apples. Plant J 96:39–55. https://doi.org/10.1111/tpj.14013

Wilkins O, Nahal H, Foong J et al (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. Plant Physiol 149:981–993. https://doi.org/10.1104/pp.108.132795

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126:485–493. https://doi.org/10.1104/pp.126.2.485

Zhou H, Lin-Wang K, Liao L et al (2015) Peach MYB7 activates transcription of the proanthocyanidin pathway gene encoding leucoanthocyanidin reductase, but not anthocyanidin reductase. Front Plant Sci 6:908. https://doi.org/10.3389/fpls.2015.00908

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J 40:22–34. https://doi.org/10.1111/j.1365-313X.2004.02183.x