frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 04 November 2021
doi: 10.3389/fevo.2021.707708

The *BAHD* Gene Family in Cacao (*Theobroma cacao*, Malvaceae): Genome-Wide Identification and Expression Analysis

Abdullah[1†], *Sahar Faraji*[2], *Parviz Heidari*[3*] *and Péter Poczai*[4,5*]

[1] Faculty of Biological Sciences, Department of Biochemistry, Quaid-I-Azam University, Islamabad, Pakistan, [2] Faculty of Crop Sciences, Department of Plant Breeding, Sari Agricultural Sciences and Natural Resources University (SANRU), Sari, Iran, [3] Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran, [4] Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland, [5] Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

OPEN ACCESS

Edited by:
Jiasheng Wu,
Zhejiang Agriculture and Forestry University, China

Reviewed by:
Sayaka Miura,
Temple University, United States
Zhi Gang Meng,
Biotechnology Research Institute of Chinese Agricultural Sciences, China
Roxana Yockteng,
Colombian Agricultural Research Corporation (CORPOICA), Colombia

***Correspondence:**
Parviz Heidari
heidarip@shahroodut.ac.ir
Péter Poczai
peter.poczai@helsinki.fi

†ORCID:
Abdullah
orcid.org/0000-0003-1628-8478

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *10 May 2021*
Accepted: *20 September 2021*
Published: *04 November 2021*

Citation:
Abdullah, Faraji S, Heidari P and Poczai P (2021) The BAHD Gene Family in Cacao (Theobroma cacao, Malvaceae): Genome-Wide Identification and Expression Analysis. Front. Ecol. Evol. 9:707708. doi: 10.3389/fevo.2021.707708

The benzyl alcohol O-acetyl transferase, anthocyanin O-hydroxycinnamoyl transferase, N-hydroxycinnamoyl anthranilate benzoyl transferase, and deacetylvindoline 4-O-acetyltransferase (BAHD) enzymes play a critical role in regulating plant metabolites and affecting cell stability. In the present study, members of the *BAHD* gene family were recognized in the genome of *Theobroma cacao* and characterized using various bioinformatics tools. We found 27 non-redundant putative *tcBAHD* genes in cacao for the first time. Our findings indicate that *tcBAHD* genes are diverse based on sequence structure, physiochemical properties, and function. When analyzed with BAHDs of *Gossypium raimondii* and *Corchorus capsularis* clustered into four main groups. According to phylogenetic analysis, *BAHD* genes probably evolved drastically after their divergence. The divergence time of duplication events with purifying selection pressure was predicted to range from 1.82 to 15.50 MYA. Pocket analysis revealed that serine amino acid is more common in the binding site than other residuals, reflecting its key role in regulating the activity of tcBAHDs. Furthermore, cis-acting elements related to the responsiveness of stress and hormone, particularly ABA and MeJA, were frequently observed in the promoter region of *tcBAHD* genes. RNA-seq analysis further illustrated that *tcBAHD13* and *tcBAHD26* are involved in response to *Phytophthora megakarya* fungi. In conclusion, it is likely that evolutionary processes, such as duplication events, have caused high diversity in the structure and function of *tcBAHD* genes.

Keywords: *BAHD* superfamily, *in silico* analysis, phylogenetics, *Phytophthora megakarya*, biotic stresses, *Theobroma cacao*, Malvaceae

INTRODUCTION

Theobroma cacao L. is an economically important species of the plant family Malvaceae (Purseglove, 1968) due to its use in chocolate production, cosmetics, and confectionery (Litz et al., 2020). The chocolate of cacao contains 45–55% fats, and its quality is determined by the aroma (Mustiga et al., 2019), which differs among varieties due to the presence and quantity of specific

compounds such as ethyl phenylacetate, ethyl octanoate, phenylethyl alcohol, 3-methylbutanal, and 2-heptanol (Castro-Alayo et al., 2019). The cacao tree grows in up to fifty countries in the humid tropics, providing an important source of income to these economies (Motamayor et al., 2013). However, the high humidity of the growing regions predisposed this plant to various fungal diseases (McElroy et al., 2018). For example, the species of *Phytophthora* fungus (*Phytophthora palmivora*, *Phytophthora megakarya*, and *Phytophthora capsici*) cause black rod/pod rot, which leads to a 20–30% loss in yield and 10% mortality of cacao (Bridgemohan and Mohammed, 2019).

The benzylalcohol O-acetyl transferase, anthocyanin O-hydroxycinnamoyl transferase, N-hydroxycinnamoyl anthranilate benzoyl transferase, and deacetylvindoline 4-O-acetyltransferase (*BAHD*) superfamily (St-Pierre, 2000) is composed of enzymes with two common domains (HXXXD and DFGWG), and similar amino acid sequences (Molina and Kosma, 2015). The HXXXD motif exists in the reaction center and contributes to catalysis, whereas the DFGWG motif exists far from the active site but is crucial for normal function (El-Sharkawy et al., 2005). This gene family plays a vital role in the biosynthesis of lipids and catalyzes acyl transfer reactions between CoA-activated hydroxycinnamic acid derivatives and hydroxylated aliphatics (Molina and Kosma, 2015). Specifically, these genes transfer CoA conjugates (e.g., malonyl-acetyl-, β-phenylalanine, anthraniloyl, tiglyl, and benzoyl-groups) to acceptor compounds for modification (D'Auria, 2006; Bontpart et al., 2015), whereas they use hydroxyl (OH) or amine (NH_2) group as an acceptor of different CoA donor substrates in the *O*- or *N*-acylation reaction (Bontpart et al., 2015). They are also involved in synthesizing a variety of polymers and secondary metabolites like volatiles, lignin, cutin, suberin, pigments, and defense-related compounds (D'Auria, 2006). Several BAHD enzymes have been characterized, including the following: isoflavone malonyltransferases (GmIMaT1 and GmIMaT3), which can differentially modify isoflavone glucosides under various stresses in soybean (Ahmad et al., 2017); glycosyltransferase; and malonyltransferase (GmMT7), which is also important for isoflavonoidin in soybean seeds (Dhaubhadel et al., 2008). Similarly, nucleocytoplasmic-localized acyltransferases (MtMaT1, 2, 3) catalyze the malonylation of 7-O-glycosidic (iso) flavones in *Medicago truncatula* (Yu et al., 2008).

Several *BAHD* genes that catalyze the formation of diverse metabolites have been characterized in plant species. Whole-genome analyses provide information about the genetic basis of response to abiotic and biotic stresses (Motamayor et al., 2013). In particular, genome-wide studies of the *BAHD* family have been reported in various species, including *Populus* (Yu et al., 2009), *Arabidopsis* (Yu et al., 2009), soybean (Ahmad et al., 2020b), *Cynara cardunculus* (Moglia et al., 2016), and in some species of Rosaceae (Zhang et al., 2019; Liu et al., 2020a). The chromosome-level genome assembly of *T. cacao* (Motamayor et al., 2013; Argout et al., 2017) provides resources for the characterization of gene families, which has led to the identification and characterization of *WRKY* (Dayanne et al., 2017), *NAC* (Shen et al., 2020), *GASA* (Abdullah et al., 2021a), and *MGT* (Heidari et al., 2021a) in cacao. However, genome-wide characterization of the *BAHD* gene family has not been reported to date.

The aim of the current study is to identify and characterize *BAHD* genes of *T. cacao*, since *BAHDs* play an important role in fat biosynthesis, and chocolate an important product of the cacao seed, comprises up to 55% fats. Here, we provide the first insight into the chromosome-wide distribution of *BAHD* genes, chemical properties, cis-regulatory elements of promoter regions, subcellular localization, and protein structure in cacao. We also aim to explore the possible role of *BAHDs* against *P. megakarya* that causes black rot.

MATERIALS AND METHODS

Identification of *BAHD* Genes in the *T. cacao* Genome and the Analysis of *BAHD* Conserved Domains

In the present study, the *BAHD* family characteristic domain (Pfam: PF02458) was used as a query using the BLAST tool in Ensembl database with an expected value of E^{-10} to identify *BAHD* genes in the *T. cacao* genome (*T. cacao* Belizian Criollo B97-61/B2) (Argout et al., 2017) and some identified *BAHD* genes of cacao for further confirmation and validations of previously identified genes. The same procedure was employed to identify *BAHD* genes in *Gossypium raimondii* and *Corchorus capsularis* for phylogenetic analysis. All protein sequences were further analyzed to confirm the presence of conserved domains (Pfam: PF02458) using the Pfam server (http://pfam.xfam.org/) and the Conserved Domains Database (CDD: https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) of the National Center for Biotechnology (NCBI). The presence of two other important domains (HXXXD and DFGWG) was confirmed by performing multiple alignments using ClustalW (Larkin et al., 2007) and visualizing in MEGA X (Kumar et al., 2018). Finally, all the sequences that lacked these domains were not included for downstream analyses following the previous study (Liu et al., 2020a) as these are important for the function of *BAHD* enzymes. We also retrieved coding DNA sequences (CDS), genomic sequences, promoter sequences (1,500 bp upstream of the gene), and protein sequences to study the sequence structure of *BAHD* genes.

Chromosome Mapping and Characterization of Physiochemical Properties

The position of each *BAHD* gene and chromosome number has been recorded. We renamed all genes based on the chromosome number and position, as shown in **Supplementary Table 1**. A phenogram was constructed to show the position of each gene on the chromosome along with duplicated genes using TBtools (Chen et al., 2020). We also determined physiochemical properties of proteins including protein length, isoelectric point (pI), and molecular weight (MW) using the ExPASy tool (Gasteiger et al., 2005). In addition, we predicted the subcellular localization of *BAHD* genes using the BUSCA webserver (Savojardo et al., 2018).

Promoter Site Analysis

We retrieved regions 1,500 bp upstream of BAHD genes and considered these to be promoter sites. PlantCare (Lescot et al., 2002) was used to analyze sequences of these promoter regions to study *cis*-regulatory elements. Each *cis*-regulatory element was classified into either hormone responsive elements (REs), stress REs, growth REs, or light REs based on its annotation in the PlantCare database.

Phylogenetic Inference and Analyses of Conserved Proteins Motif

A maximum likelihood tree was constructed for *BAHD* genes, first in cacao alone and then combined with *BAHD* genes of two other species (*Gossypium raimondii* and *Corchorus capsularis*). The maximum likelihood tree was constructed using IQ-tree (Nguyen et al., 2015) with default parameters and the best fit model JTT+I+G4 based on predictions of ModelFinder (Kalyaanamoorthy et al., 2017). The visualization of both the trees was improved using the interactive tree of life (Letunic and Bork, 2019) and MEGA X (Kumar et al., 2018). The distribution of conserved protein motifs was elucidated in BAHD proteins using MEME v5.3.0 server (http://meme-suite.org/tools/meme) (Bailey et al., 2009), which searched for 10 conserved motifs with a minimum width of motif 6 and a maximum width of motif 30.

Gene Duplications and Synteny Analyses

The *BAHD* genes of cacao were pairwise aligned in Geneious R8.1 (Kearse et al., 2012), and gene pairs that had similarity of 85% or higher were considered duplicated genes following previous studies (Zheng et al., 2010; Musavizadeh et al., 2021). Duplicated genes that occurred within 200 kb region were considered tandemly duplicated genes, whereas those that were separated >200 kb region or located on different chromosomes were considered segmentally duplicated genes following a recent study (Ahmad et al., 2020a). The rate of synonymous (Ks) and non-synonymous substitutions (Ka) and events of gene duplication were determined using DnaSP v.6 (Rozas et al., 2017). The selection pressure on duplicated genes was determined based on the ratio of Ka/Ks and interpreted as negative (<1), neutral (=1), and positive (>1) (Lawrie et al., 2013). The divergence time of duplication was calculated by a synonymous mutation rate of λ substitutions per synonymous site per year as $T = (Ks/2\lambda)$ $(\lambda = 6.5 \times 10^{-9}) \times 10^{-6}$ (Yang et al., 2008). We also analyzed synteny relationships of cacao *BAHD* genes with *G. raimondii* and *C. capsularis* and drawn at chromosome level using Circos software (Krzywinski et al., 2009).

Three-Dimensional Protein Modeling and Molecular Docking

The three-dimensional structure of BAHD proteins was estimated by the protein homology/analogy Recognition Engine Version 2.0 (Phyre2) server (Kelley et al., 2015). The predicted structure of proteins was validated using the Ramachandran plot (Lovell et al., 2003) following a previous study (Abdullah et al., 2021a). The Beta Cavity webserver (Kim et al., 2015) was used to estimate molecular voids and pockets in proteins. We also used the ProSA server (Wiederstein and Sippl, 2007) to estimate errors in protein structure and validate 3D-modeled proteins. The P2Rank in PrankWeb software (Jendele et al., 2019) and CASTp tool (Tian et al., 2018) were used for docking analysis of the ligand-binding regions in the modeled proteins. Finally, the results were analyzed using PyMOL (DeLano, 2002).

Expression Analysis of *BAHD* Genes Under Biotic Stress

The available already processed RNAseq data of biotic stress related to *P. megakarya* infection is available in GEO DataSets under accession number GSE116041 for tolerant and susceptible cultivars of cacao. The data have been reported in the National Center for Biotechnology Information (NCBI) by Pokou et al. (2019) after doing RNA sequencing of both cultivars after inoculating with fungus at 0, 6, 24, and 72 h. They trimmed the raw reads by trimmomatic (Bolger et al., 2014) and mapped the high-quality trimmed reads by HISAT2 (Kim et al., 2015) to the reference genome of cacao (Criollo genome v2.0) (Argout et al., 2017). The differential expression patterns of genes were determined using DESeq2 with the default setting. The complete details can be seen at Pokou et al. (2019). We downloaded the processed RNAseq data of the aforementioned method and analyzed that to extract the data of BAHD using their gene IDs. The heatmaps of all expressed BAHD genes were drawn by the TBtools package (Chen et al., 2020) after log2 transformation for 0, 6, 24, and 72 h fungus inoculation. In the heatmaps, the control condition was 0 h (before infection).

Coexpression Network of *BAHD* Genes

A coexpression network of tcBAHDs was constructed by String database (Szklarczyk et al., 2019) using their orthologs in *Arabidopsis thaliana*. The network was created with an interaction score of more than 0.30 and a number of interactors was set to >20 in the first shell and >5 in the second shell. Finally, the network was provided by Cytoscap software (Franz et al., 2016).

RESULTS

Identification of *tcBAHD* Genes and Their Distributions on Chromosomes Within Genomes

We spotted 27 *BAHD* genes in cacao genomes (*tcBAHD*) with functional domains of HXXXD and DFGWG (**Figure 1A**) distributed across nine of ten chromosomes (**Figure 1B**). The gene ID was renamed from *tcBAHD1* to *tcBAHD27* based on the existence on the chromosome and the position starting from chromosome 1. If, two or more genes were located on the same chromosome, the gene located first was renamed as first (**Supplementary Table 1**). Among 27 genes, eight were located on chromosome V, four each on chromosomes III and chromosome IX, three on chromosome II, two each on chromosome IV, VI, and VIII, and one each on chromosome I and X. No *BAHD* gene was found on chromosome VII. This data revealed the unequal distribution of *tcBAHD* genes across the cacao genome (**Supplementary Table 1**, **Figure 1B**).

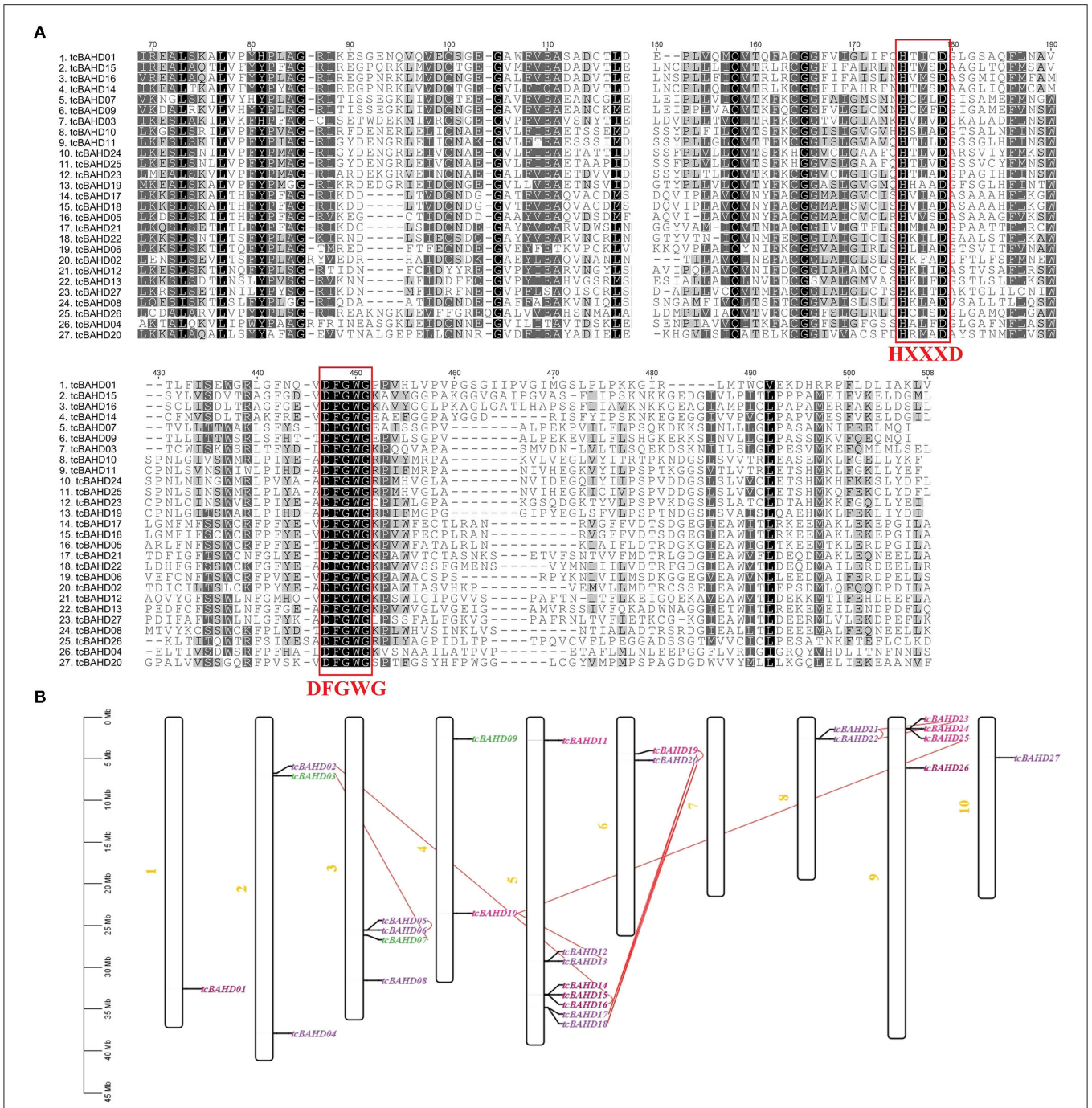

FIGURE 1 | Multiple alignment and phenogram of tcBAHD proteins. **(A)** the multiple alignment of tcBAHD proteins indicates the presence of HXXXD and DFGWG domain. **(B)** location of each *tcBAHD* gene on the chromosome of *Theobroma cacao*. The color indicates the phylogenetic grouping of the gene and the link between genes shows duplicate pairs.

Physiochemical Characterization of tcBAHD Proteins

The molecular weight of BAHDs ranged from 46.83 kDa (tcBAHD01) to 55.74 kDa (tcBAHD14) with a length of 364 (tcBAHD11) to 504 (tcBAHD14) amino acids. The isoelectric point (pI), as an indicator to determine the optimal pH, varied among tcBAHD proteins from 4.69 (tcBAHD07) to 8.72 (tcBAHD25). Overall, 14 proteins predicted with pI <6.5, highlighting that some BAHD enzymes are alkaline and some are acidic in nature. Subcellular localization analysis showed the localization of these enzymes in the cytoplasm, chloroplast, mitochondria, and organellar membrane

(**Supplementary Table 1**). Overall, these results illustrate that tcBAHDs are diverse based on their sequence and physiochemical properties.

Evolutionary Analyses of tcBAHD Proteins

Maximum likelihood analysis demonstrated that 65 *BAHD* sequences of three species were clustered into four groups, including 27 sequences of cacao, 18 of *G. raimondii*, and 20 of *C. capsularis* (**Figure 2**). Groups II and IV were further divided into three and four subgroups, respectively. The tcBAHD proteins of cacao were found distributed in all four groups. Group I comprised three tcBAHDs, group II included five tcBAHDs, one each in II-a and II-b, and three in II-c, group III had six tcBAHDs, and group IV included 13 tcBAHDs one each in IV-a and IV-c, and other eleven in IV-d. Most of the cacao sequences showed sister relationships with the sequences of *C. capsularis* rather than *G. raimondii*. Each group contained enzymes of the same functions or of diverse functions, such as group I contained all benzyl alcohol O-benzoyltransferase, whereas group II-a contained putative 10-deacetylbaccatin III 10-O-acetyltransferase (tcBAH01), II-b contained putative omega-hydroxypalmitate O-feruloyl transferase (tcBAHD26), and II-c had all omega-hydroxypalmitate O-feruloyl transferase BAHD enzymes (tcBAHD03, tcBAHD07, and tcBAHD09) as shown in **Supplementary Table 1**. Group IV was highly diverse, in which IV-a included putative shikimate O-hydroxycinnamoyltransferase (tcBAHD20), IV-c comprised putative brassino steroid-related acyltransferase 1 (tcBAHD04), and IV-d comprised putative vinorine synthase (tcBAHD06, 08, 12, 13, 22, 27), putative salutaridinol 7-O-acetyltransferase (tcBAHD05, 17, 18), putative acetyl-CoA-benzylalcohol acetyltransferase (tcBAHD02), and putative acylsugar acyltransferase 3 (tcBAHD21). The phylogenetic relationship was also correlated with the gain and loss of protein motifs and introns. Hence, a separate tree of tcBAHD was constructed (**Figure 3A**). The analysis of protein motifs also revealed high similarities among the sequences that cluster together (**Figure 3**). Ten conserved motifs were recognized in the protein sequence of tcBAHDs (**Figure 3B**). The sequences of group I, II-c, and III contained the same motif patterns and included all other motifs, except motif 7 which was only found in group IV-b and IV-d and was absent in the sequences of all other groups. In group II-a, tcBAHD01 lacked four motifs (motifs 7, 8, 9, and 10), whereas in group II-b tcBAHD26 lacked three motifs (motifs 7, 8, and 10). In group IV, nine sequences contained all 10 motifs. Hence, the four sequences that lacked some protein motifs are tcBAHD20 of IV-a (lacked motifs 7, 9, and 10), tcBAHD04 of IV-b (lacked motifs 9 and 10), and tcBAHD08 and tcBAHD13 of IV-d lacked motif 10 and 9, respectively. The conserved motifs were more distributed in the region of the conserved transferase domain (**Table 1**). The gain and loss of introns showed different results across the four clusters of the maximum likelihood tree (**Figure 3C**). The gene of group I and IV showed high similarities i.e., all three genes of group I contained one intron, whereas single genes of group IV-a and IV-c contained one and two intron(s), respectively. All the 11 genes of group IV-d lacked introns. The genes of group II and III showed some diversity as the number of introns varied from 0 to 2 within the same cluster, such as single gene of group II-a contains one intron, single gene of II-b lacked intron, and genes of II-c had either one or two introns. The five genes of group III varied not only in the number of introns (0–2 introns) but also in the pattern of distribution (**Figure 2C** and **Supplementary Table 1**).

BAHD Genes Duplication and Synteny Analyses

The paralogous relationships among *tcBAHD* genes were analyzed along with orthologous relationships by comparing them with the *BAHD* sequences of *G. raimondii* and *C. capsularis*. The duplication events were recorded for 14 pairs of genes, ten of which led to the generation of gene with different functions based on predicted physicochemical properties and their reported annotation (**Table 2**). These results showed that gene duplications are responsible for the expansion and diversification of function in *BAHD* genes. Evidence of purifying selection pressure on these genes was seen by estimating the ratio of non-synonymous and synonymous substitutions in DnaSP v.6. The purifying selection pressure indicates that these genes are expressed regularly under high selection pressure. Hence avoiding incorporation of any new amino acid that may either cause malfunctioning or completely disturb the protein structure. The divergence time analyses showed that duplication events mainly occurred recently and ranged from 1.82 to 15.50 MYA (**Table 2**). The intraspecies synteny of *BAHDs* was drawn between cacao and *G. raimondii* and between cacao and *C. capsularis* (**Figure 4**). The 27 *tcBAHDs* in cacao showed eight syntenic block relationships with *BAHDs* in *G. raimondii* (**Figure 4A**) and 12 syntenic block relationships with *BAHDs* in *C. capsularis* (**Figure 4B**). These results showed that *tcBAHDs* have more syntenic relationships with *BAHDs* of *C. capsularis* than *G. raimondii*.

Pocket Analysis of tcBAHD Proteins

In the present study, the predicted 3D structure analysis of tcBAHDs showed diverse structures (**Supplementary Figure 1**). Pocket sites related to activation or binding site were highlighted in the structure of tcBAHDs (**Figure 5A**). The amino acid residues serine (SER), glycine (GLY), proline (PRO), lysine (LYS), threonine (THR), cysteine (CYS), and arginine (ARG) were more commonly recognized as the critical binding sites in the pocket sites of tcBAHDs (**Figures 5A,B**). In particular, SER amino acid was more abundant in the binding site of proteins, indicating that it may have potential roles that affect the function of tcBAHDs.

Promoter Regions and Structure Analyses of tcBAHD

We identified several key responsive elements in the promoter region of *tcBAHDs*. The most prominent responsive elements included those related to stress (45%), hormone (27%), light (21%), and growth (7%) (**Figure 6A**). Elements related to DNA and protein-binding sites were also recorded

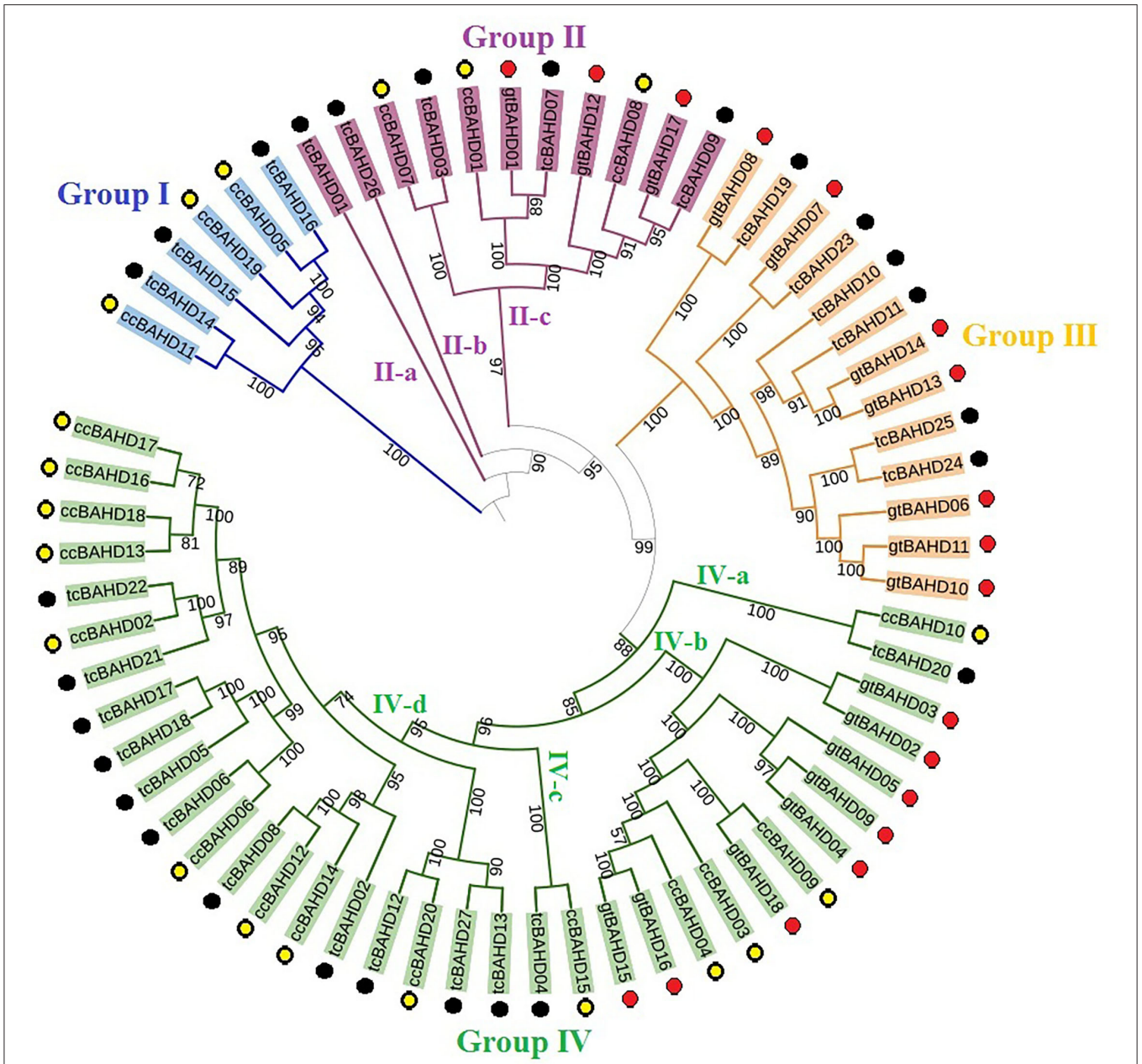

FIGURE 2 | Phylogeny analysis of 65 BAHDs in three species including *Theobroma cacao*, *Gossypium raimondii,* and *Corchorus capsularis*. Overall, 27 tcBAHDs (black circles), 20 BAHDs of *C. capsularis* (yellow circles), and 18 BAHDs of *G. raimondii* (red circles) were classified into four main groups. The percentage of bootstrap values is provided in the branches.

(**Supplementary Table 2**). Regulatory sites were found for numerous hormones, such as salicylic acid, auxin, gibberellin, MeJA, and ABA (**Supplementary Table 2**, **Figure 6B**). We found that *tcBAHD* genes may be more induced by ABA and MeJA based on the distribution of cis-acting elements in their promoter site. Similarly, regulatory elements were identified for drought, anoxic inducibility, elicitor, seed-specific regulation, anaerobic induction, low temperature, circadian control, and plant defense/stress (**Supplementary Table 2**, **Figure 6C**).

Expression Analyses of *tcBAHDs* in Biotic Stress (Response to *P. megakarya*)

The role of *tcBAHDs* was also elucidated against *P. megakarya* using RNA-seq data of cacao at 0 hours (0 h), 6, 24, and 72 h in fungal resistant cultivars (Scavina; SCA6) and susceptible cultivars (Nanay; NA32) (**Figure 7**). In the susceptible cultivars, *tcBAHD*01 was up-regulated after 6 h compared to 0 h (as a control condition), while after 24 h, *tcBAHD*25 was more induced (**Figure 7A**).

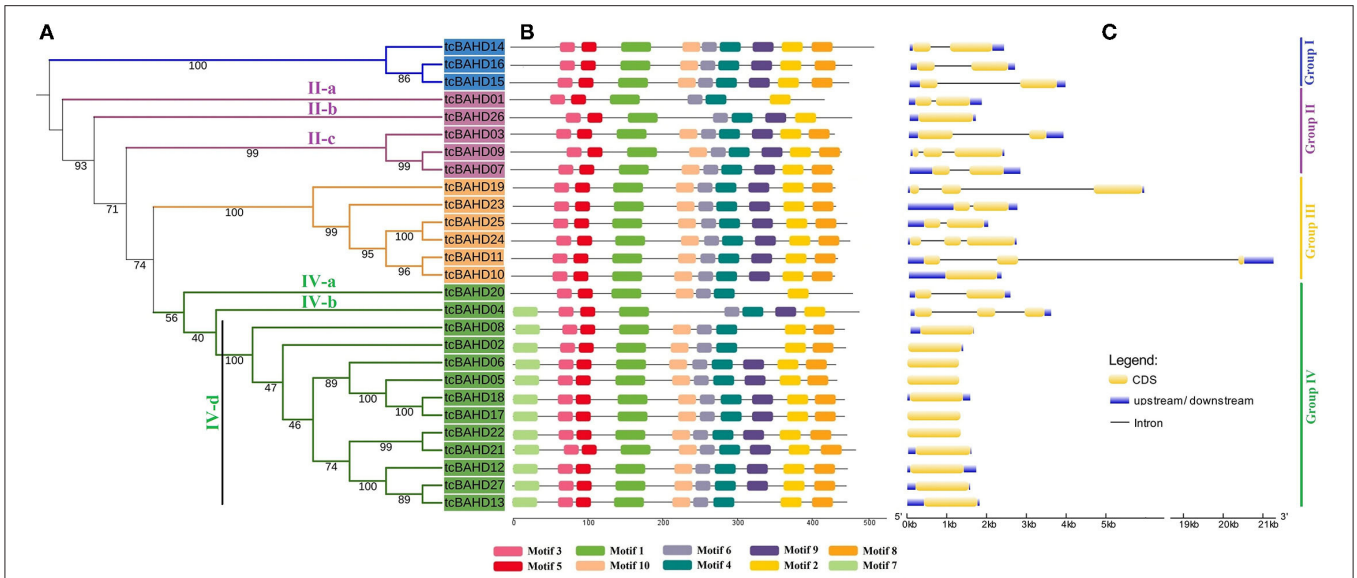

FIGURE 3 | Evolutionary analysis of tcBAHD proteins. Maximum likelihood analysis of tcBAHD proteins **(A)**, distribution of conserved motifs into tcBAHD proteins **(B)**, and gene structure of *tcBAHDs* **(C)**.

TABLE 1 | The conserved protein motifs predicted in tcBAHD proteins of cacao.

Motif no.	*E*-value	Amino acid sequence of motif	Width (aa)	Domain
Motif 1	4.0e-44	IQVTKFKCGGFAJGLCLSHTJADGSAALQFLNSWAEVARGL	41	Transferase
Motif 2	3.2e-29	PNLGISSWCRFPFYEADFGWGKPVWVGPA	29	Transferase
Motif 3	9.1e-21	LKESLSKTLVPFYPLAGRLKE	21	Transferase
Motif 4	4.2e-19	DQPTKLLIAVBGRSRLNPPLPSGYIGNVI	29	Transferase
Motif 5	2.9e-17	IDCNDEGVLFVEAZVBCTLDE	21	Transferase
Motif 6	6.0e-15	KPSRFEALTAFIWRCRTKARK	21	Transferase
Motif 7	1.7e-14	MEVQIISRETIKPSSPTPHHLRTFKLSLLDQLAP	34	Transferase
Motif 8	1.5e-13	LPSTKDGGGIEAWITLDESEMKIFEKDLE	29	Transferase
Motif 9	5.2e-11	PLSDLVKLIRZAIKEMDDEYLRSAIDYME	29	Transferase
Motif 10	1.4e-08	EKSVTRRFVFTADKJATLKAKAKED	25	–

In the fungal resistant cultivars, *tcBAHD13* showed up-regulation after 6 h, and 24 h compared to 0 h; after 72 h, *tcBAHD26* was more up-regulated in response to *P. megakarya* infection (**Figure 7B**). The expression profile also confirmed that *tcBAHDs* have diverse functions as well as physicochemical properties.

Coexpression Analyses of *tcBAHDs*

Co-expression analysis revealed that orthologs of *tcBAHD* genes interact with genes involved in the phenylpropanoid pathway, lignin metabolic process, coumarin biosynthetic process, flavonoid biosynthetic process, response to karrikin and abiotic stress (**Figure 8** and **Supplementary Table 3**). The hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase (HCT), as an ortholog of tcBAHDs, showed a high interaction score in co-expression network with 4-coumarate: CoA ligase (4CL), and phenylalanine ammonia-lyases (PALs), which are both involved in phenylpropanoid metabolic process and response to UV (**Figure 8**). These findings reveal that BAHDs are associated with several metabolic pathways, which may increase the resistance of plants to abiotic stresses. In addition, the endoplasmic reticulum was identified as a cellular component for the activity of *BAHDs* and their interactors (**Supplementary Table 3**).

DISCUSSION

Transfer of acetyl to cellular metabolites can affect their activity and stability. *BAHD* is an important plant gene family that affects the acetylation of many metabolites (D'Auria, 2006). In the present study, 27 non-redundant putative *tcBAHD* genes were recognized in the cacao genome for the first time. These

TABLE 2 | Predicted Ka/Ks values for the duplicated gene pairs of *tcBAHD* in cacao genome.

Gene 1	Function	Gene 2	Function	Ka/Ks	Divergence time (MYA)
tcBAHD02	Acetyl-CoA-benzylalcohol acetyltransferase	*tcBAHD15*	Benzyl alcohol O-benzoyltransferase	0.292	15.50
tcBAHD03	Putative Omega-hydroxypalmitate O-feruloyl transferase	*tcBAHD07*	Omega-hydroxypalmitate O-feruloyl transferase	0.397	10.44
tcBAHD05	Salutaridinol 7-O-acetyltransferase	*tcBAHD06*	Vinorine synthase	0.387	9.74
tcBAHD10	Shikimate O-hydroxycinnamoyltransferase	*tcBAHD13*	Vinorine synthase	0.475	4.43
tcBAHD10	Shikimate O-hydroxycinnamoyltransferase	*tcBAHD25*	Shikimate O-hydroxycinnamoyltransferase	0.368	8.11
tcBAHD15	Benzyl alcohol O-benzoyltransferase	*tcBAHD16*	Benzyl alcohol O-benzoyltransferase	0.323	9.42
tcBAHD17	Salutaridinol 7-O-acetyltransferase	*tcBAHD19*	Shikimate O-hydroxycinnamoyltransferase	0.323	6.86
tcBAHD17	Salutaridinol 7-O-acetyltransferase	*tcBAHD20*	Shikimate O-hydroxycinnamoyltransferase	0.315	6.54
tcBAHD18	Salutaridinol 7-O-acetyltransferase	*tcBAHD19*	Shikimate O-hydroxycinnamoyltransferase	0.332	5.42
tcBAHD18	Salutaridinol 7-O-acetyltransferase	*tcBAHD20*	Shikimate O-hydroxycinnamoyltransferase	0.375	4.83
tcBAHD19	Shikimate O-hydroxycinnamoyltransferase	*tcBAHD20*	Shikimate O-hydroxycinnamoyltransferase	0.385	4.54
tcBAHD21	Vinorine synthase	*tcBAHD22*	Vinorine synthase	0.344	1.82
tcBAHD21	Acylsugar acyltransferase 3	*tcBAHD23*	Shikimate O-hydroxycinnamoyltransferase	0.608	3.14
tcBAHD22	Vinorine synthase	*tcBAHD23*	Shikimate O-hydroxycinnamoyltransferase	0.533	3.42

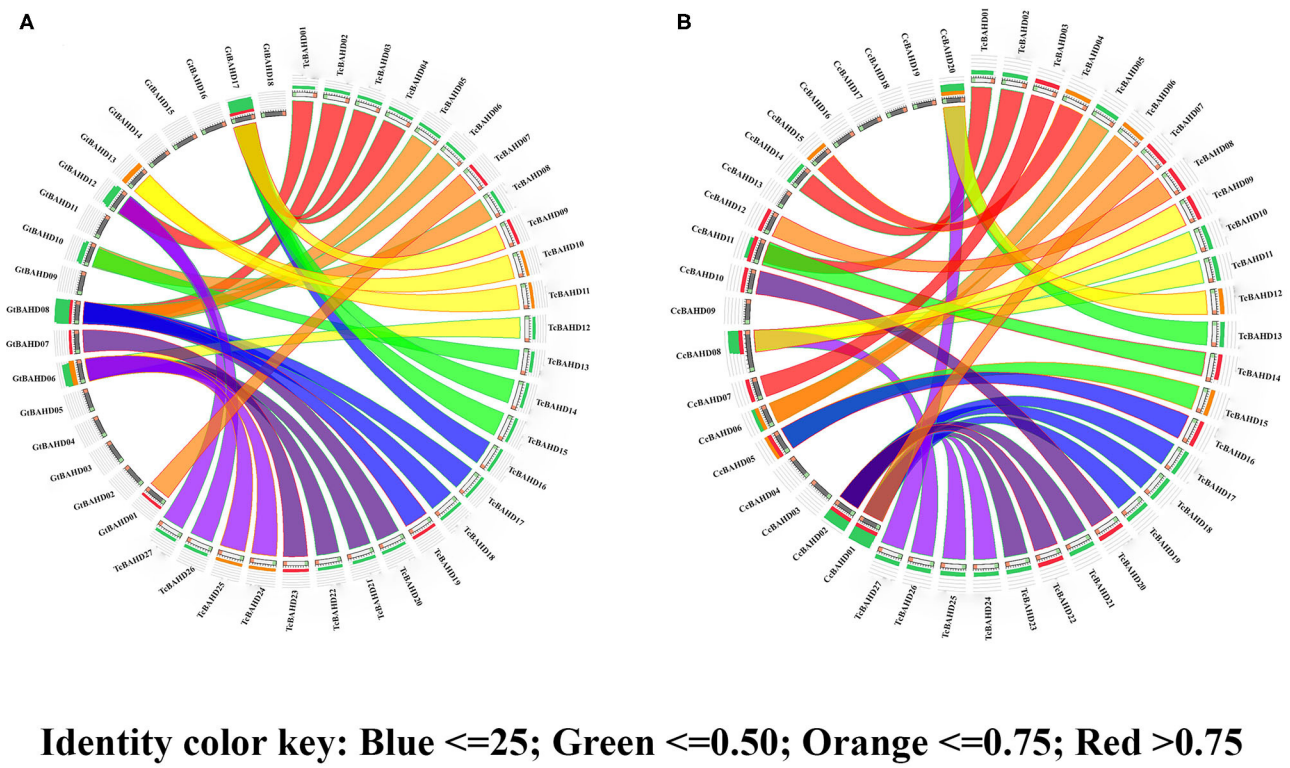

FIGURE 4 | Synteny analysis of BAHD genes. The syntenic blocks of *Theobroma cacao* BAHD genes are compared with *Gossypium raimondii* **(A)** and *Corchorus capsularis* **(B)**. The percentage identity of syntenic blocks shows under the color spectrum that red to blue, the %identity is reducing. The gray stack bars with green and red tips indicate genes of *Gossypium raimondii* **(A)** and *Corchorus capsularis* **(B)**, whereas the white stack bars with orange and green tips indicate genes of *T. cacao*. The presence of red, orange, or green stack bars on *T. cacao* and the presence of one or two types of stack bars on genes of *G. raimondii* and *C. capsularis* are to show the cover rate between syntenic blocks with the genes of other species.

genes are lowered in number as compared to previous reports of the Rosaceae family, in which 69–141 genes were reported (Zhang et al., 2019; Liu et al., 2020a). However, the number of genes within a gene family can vary among species (Rezaee et al., 2020; Song et al., 2020; Abdullah et al., 2021a; Faraji et al., 2021). The identified *tcBAHDs* showed high diversity based

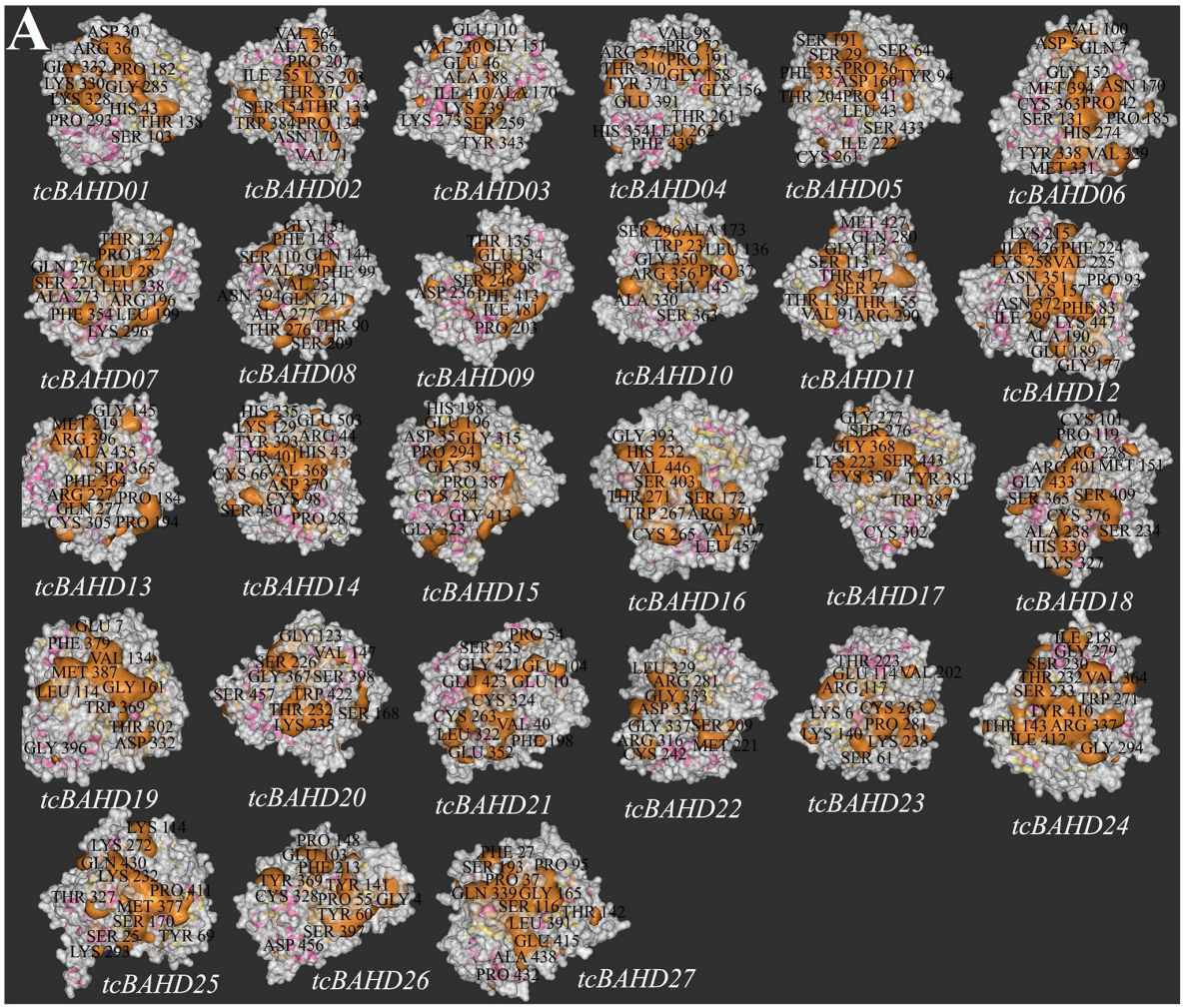

FIGURE 5 | Pocket sites analysis of the tcBAHD proteins. **(A)** Docking analysis of the major protein pocket sites into the structure of tcBAHDs. **(B)** Frequency of amino acids in the predicted pocket sites.

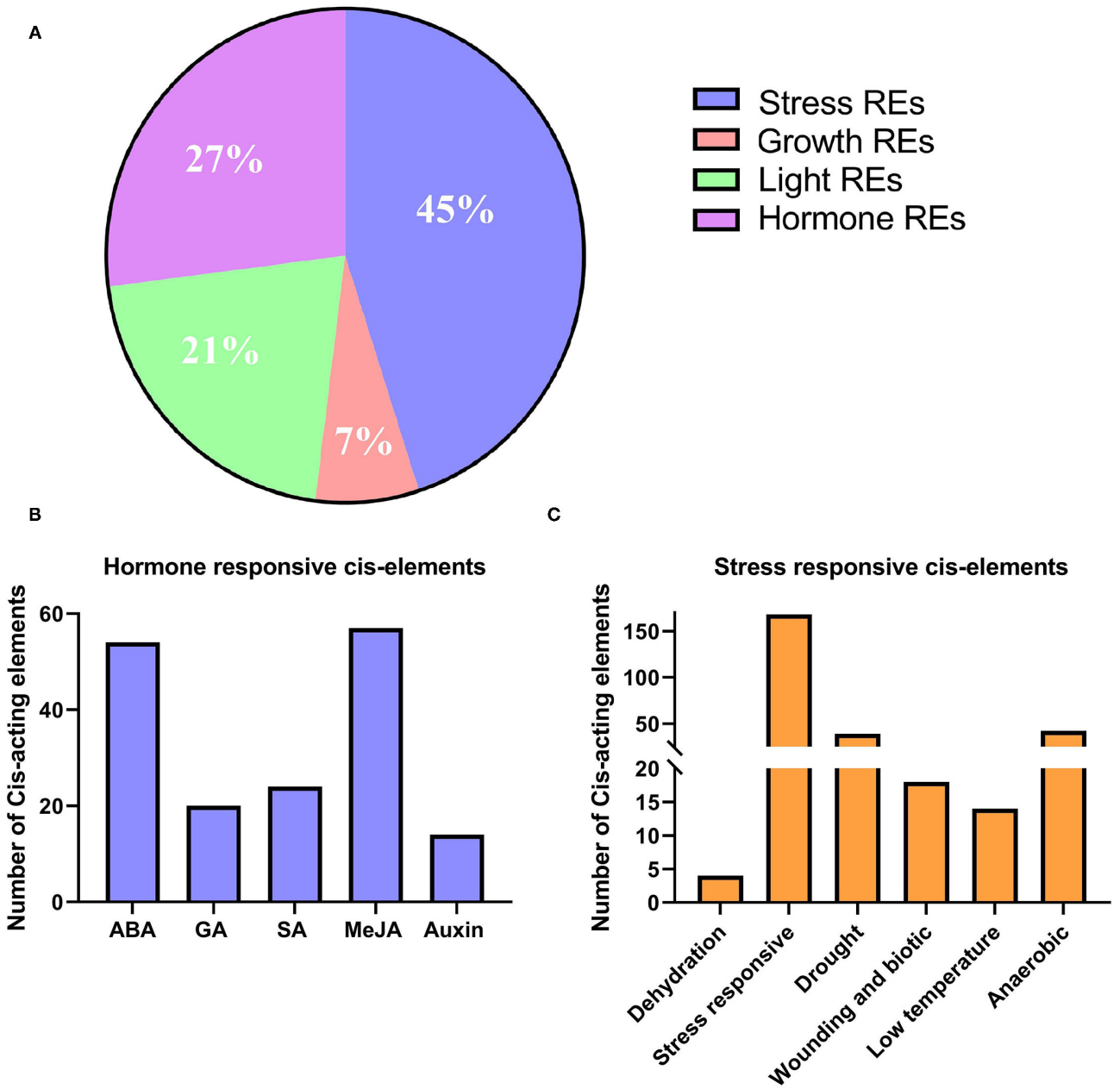

FIGURE 6 | Distribution of cis-regulatory elements into promoter regions of *tcBAHD* genes. **(A)** Classification of identified regulatory elements based on function and their response to hormone, light, stress, and growth. **(B)** Distribution of different types of hormone-related *cis*-regulatory elements. **(C)** *Cis*-regulatory elements are related to various types of stresses.

on sequence structure, physicochemical properties, functions, and distribution across the chromosome, which is in agreement with previous reports of the *BAHD* gene family (Moglia et al., 2016; Zhang et al., 2019; Ahmad et al., 2020b; Liu et al., 2020a). The annotation retrieved from the Ensembl indicates the identified tcBAHD enzymes have high diversity in the function of tcBAHD. The benzyl alcohol benzoyl transferase synthesizes the minor constituent of floral aroma benzyl benzoate and other volatile esters in esters, and also provides support during leaf damage and phytopathogenic bacteria stress (D'Auria et al., 2002). The omega-hydroxypalmitate O-feruloyl transferase synthesizes suberin aromatics and also reinforce barrier against the pathogen (Balestrini et al., 2020). Brassino steroid-related acyltransferase is important for plant development and regulation of various biological pathways including the development of flowers and seeds (Singh and Savaldi-Goldstein, 2015) and also protect against various biotic and abiotic stresses (Krishna, 2003). Shikimate

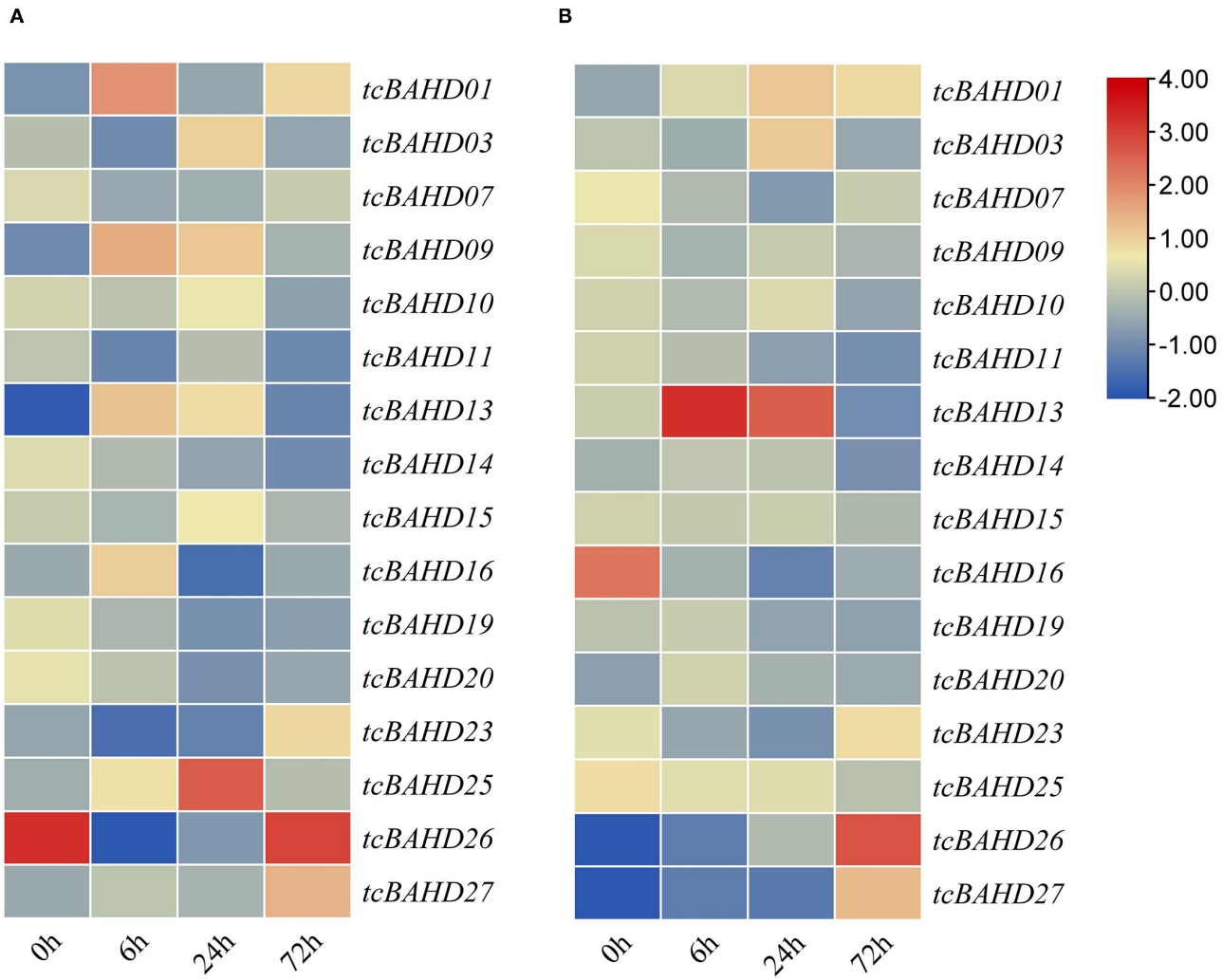

FIGURE 7 | Expression analyses of *tcBAHD* genes in cacao plants under inoculation with *P. megakarya* after 0, 6,24, and 72 h. Nanay (NA-32) susceptible cultivar **(A)**, Scavina (SCA6) tolerant cultivar **(B)**.

O-hydroxycinnamoyl transferase accepts p-coumaroyl-CoA and caffeoyl-CoA as substrates and transfers the acyl group on both quinate and shikimate acceptors (Levsh et al., 2016) and is involved specifically in lignin and phenylpropanoid biosynthesis to support plant growth (Hoffmann et al., 2004). These examples indicate the diverse role of the identified tcBAHDs in cacao growth and development. The recognized BAHDs in cacao, *G. raimondii*, and *C. capsularis* clustered into four groups that each contained enzymes with similar or diverse functions. Previous studies also revealed that phylogenetic clustering in the same group is not an indication of the same function (Nawaz et al., 2019; Abdullah et al., 2021a). Interestingly, unlike *G. raimondii*, the *BAHD* genes of *C. capsularis* and *T. cacao* were closely related in each cluster of the tree. This finding agrees with previous phylogenetic studies of the family Malvaceae which indicate cacao among the earlier diverged species while *Gossypium* is recently diverged (Abdullah et al., 2019, 2020, 2021b). The *tcBAHD* genes also differed in terms of intron number and protein motifs. These variations may be responsible for the diverse functions, as they can affect the function of homologous genes and protein-protein interactions (Heidari et al., 2020; Faraji et al., 2021). Hence, it has been suggested that *BAHD* genes rapidly evolved after divergence (Yu et al., 2009).

Tandem and segmental duplication are helpful for domestication, survival, and resistance to biotic and abiotic stresses in plants as they generate structural and functional diversity within genes (Liu et al., 2020b; Schilling et al., 2020; Zan et al., 2020). We identified 14 duplication events within *BAHD* genes, which mainly led to the generation of genes with diverse functions (**Table 2**). This genes duplication may also play a significant role in the evolution and domestication of cacao. The purifying selection on duplicated genes indicates that these genes play important

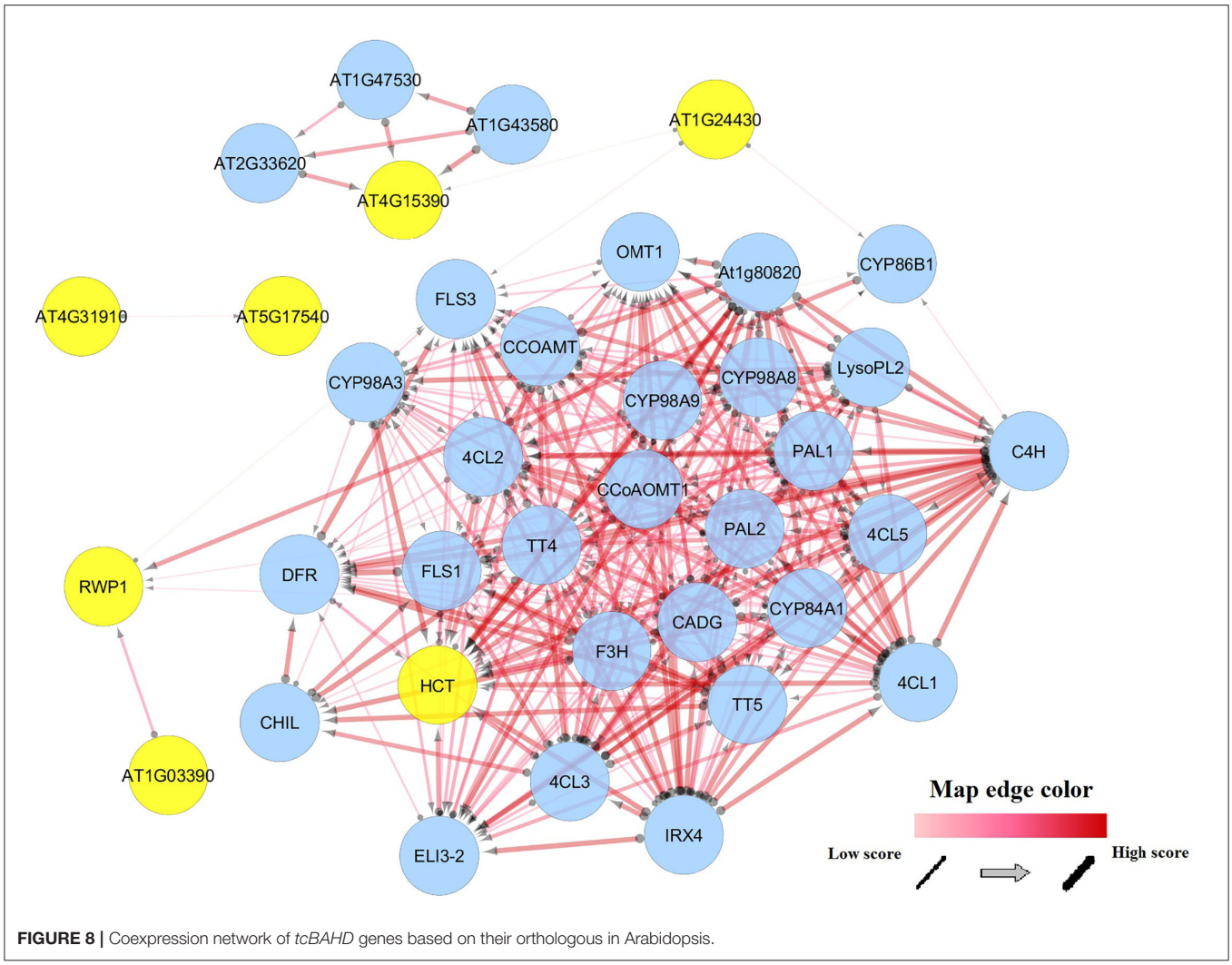

FIGURE 8 | Coexpression network of *tcBAHD* genes based on their orthologous in Arabidopsis.

functions in cacao growth and development. Hence, they are expressed regularly and avoid deleterious mutations that cause malfunctions/structure modifications that can terminate/decrease the function of these genes (Page and Holmes, 2009; Cvijović et al., 2018). Moreover, the prediction of subcellular localization revealed that tcBAHDs localize in the cytoplasm, chloroplast, and mitochondria, which further supports the diverse function of these enzymes within the different cell compartments, important for the regulation of cacao.

Surface pocket analysis of tcBAHDs is considered important because this provide insight into the key binding site of protein structures that affecting the enzymatic activity and protein-protein interaction (Stank et al., 2016). Serine, glycine, and proline were highly observed amino acids at pocket sites, revealing that tcBAHDs may also respond to adverse conditions, including abiotic and biotic stresses (Faraji et al., 2020; Heidari et al., 2021b). Serine was observed in the pocket site of most BAHDs suggesting its key role in regulating the activity of tcBAHDs and cellular pathways belonging to this gene family. Furthermore, a high amount of *cis*-acting elements related to stress-response in the promoter region of *tcBAHD* genes suggests that members of this gene family may be induced by transcription factors related to stress stimuli, as observed in other studies (Ahmadizadeh and Heidari, 2014; Heidari et al., 2019). *Cis*-acting elements related to ABA and MeJA hormones are frequently observed in promoter sites, indicating that *tcBAHD* genes are more induced by hormones associated with response to stress stimuli. Previous studies also reported that *BAHD acyltransferases* are involved in diverse pathways related to regulation of plant structure and function, cell stability, and the production of secondary metabolites (Luo et al., 2007; Grienenberger et al., 2009; Li et al., 2018; Kusano et al., 2019). Besides, the coexpression network illustrated that BAHDs are involved in the biosynthesis of secondary metabolites, which may relate to response to abiotic stress. Hence, these may also be important in resistance to biotic and abiotic stresses. We further studied the expression pattern of *tcBAHD* genes

in resistant (Scavina; SCA6) and susceptible (Nanay; NA32) cultivars of cacao in response to fungal infection (*P. megakarya*). Our results revealed two *BAHD* genes, *tcBAHD13* (a vinorine synthase) and *tcBAHD26* (an Omega-hydroxypalmitate O-feruloyl transferase) that were upregulated specifically in resistant cultivars, indicating that these two genes are potentially involved in the *P. megakarya* fungi response. Although, these results were not validated from qualitative PCR (qPCR), previous studies reported high consistency between the result of RNA-seq and qPCR, i.e., in the *GASA* gene family in soybean (Ahmad et al., 2019) and apple (Fan et al., 2017), extensin genes in tomato (Ding et al., 2020), papain-like cysteine proteases in rice (Niño et al., 2020) and cotton (Zhang et al., 2019), and auxin/indole-3-acetic acid in pepper (Waseem et al., 2018). Nevertheless, a functional study is required to draw a complete conclusion.

CONCLUSIONS

In the present study, we identified and characterized 27 *tcBAHD* genes in the cacao genome using bioinformatics tools. Our findings indicated that tcBAHDs have a high degree of structural diversity and a wide range of functions. Various duplication events can be attributed to evolutionary process that produce this increased diversity. Further investigations are required to confirm the role of *tcBAHDs* in cacao growth and response to biotic/abiotic stresses.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

A and PH: manuscript drafting. A and SF: data analyses and data curation. A, SF, PH, and PP: data interpretation. A, PP, and PH: conceptualization. PH and PP: review and editing of the first draft and supervision. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.707708/full#supplementary-material

Supplementary Figure 1 | 3D structure analyses of tcBAHD proteins.

Supplementary Table 1 | List of the identified tcBAHD genes and their characteristics in the cacao genome.

Supplementary Table 2 | Promoter important cis-elements engaged in various developmental and stress-responsive pathways in the tcBAHD genes.

Supplementary Table 3 | List of significant GO terms based on the co-expression network of orthologs of *tcBAHD* genes in the Arabidopsis.

REFERENCES

Abdullah, Faraji, S., Mehmood, F., Malik, H. M. T., Ahmed, I., Heidari, P., et al. (2021a). The GASA gene family in cacao (*Theobroma cacao*, Malvaceae): genome wide identification and expression analysis. *Agronomy* 11:1425. doi: 10.3390/agronomy11071425

Abdullah, Mehmood, F., Shahzadi, I., Ali, Z., Islam, M., Naeem, M., et al. (2021b). Correlations among oligonucleotide repeats, nucleotide substitutions and insertion-deletion mutations in chloroplast genomes of plant family Malvaceae. *J. Syst. Evol.* 59, 388–402. doi: 10.1111/jse.12585

Abdullah, Mehmood, F., Shahzadi, I., Waseem, S., Mirza, B., Ahmed, I., et al. (2020). Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics* 112, 581–591. doi: 10.1016/j.ygeno.2019.04.010

Abdullah, Shahzadi, I., Mehmood, F., Ali, Z., Malik, M. S., Waseem, S., et al. (2019). Comparative analyses of chloroplast genomes among three *Firmiana* species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Gene* 19:100199. doi: 10.1016/j.plgene.2019.100199

Ahmad, B., Yao, J., Zhang, S., Li, X., Zhang, X., Yadav, V., et al. (2020a). Genome-wide characterization and expression profiling of *GASA* genes during different stages of seed development in grapevine (*Vitis vinifera* L.) predict their involvement in seed development. *Int. J. Mol. Sci.* 21, 1–16. doi: 10.3390/ijms21031088

Ahmad, M. Z., Li, P., Wang, J., Rehman, N. U., and Zhao, J. (2017). Isoflavone malonyltransferases GmIMaT1 and GmIMaT3 differently modify isoflavone glucosides in soybean (*Glycine max*) under various stresses. *Front. Plant Sci.* 8:735. doi: 10.3389/fpls.2017.00735

Ahmad, M. Z., Sana, A., Jamil, A., Nasir, J. A., Ahmed, S., Hameed, M. U., et al. (2019). A genome-wide approach to the comprehensive analysis of *GASA* gene family in *Glycine max*. *Plant Mol. Biol.* 100, 607–620. doi: 10.1007/s11103-019-00883-1

Ahmad, M. Z., Zeng, X., Dong, Q., Manan, S., Jin, H., Li, P., et al. (2020b). Global dissection of the *BAHD* acyltransferase gene family in soybean: expression profiling, metabolic functions, and evolution. doi: 10.21203/rs.2.21482/v1

Ahmadizadeh, M., and Heidari, P. (2014). Bioinformatics study of transcription factors involved in cold stress. *Biharean Biol.* 8, 83–86.

Argout, X., Martin, G., Droc, G., Fouet, O., Labadie, K., Rivals, E., et al. (2017). The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* 18:730. doi: 10.1186/s12864-017-4120-9

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Balestrini, R., Ghignone, S., Quiroga, G., Fiorilli, V., Romano, I., and Gambino, G. (2020). Long-term impact of chemical and alternative fungicides applied to Grapevine cv Nebbiolo on Berry Transcriptome. *Int. J. Mol. Sci.* 21:6067. doi: 10.3390/ijms21176067

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bontpart, T., Cheynier, V., Ageorges, A., and Terrier, N. (2015). BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *New Phytol.* 208, 695–707. doi: 10.1111/nph.13498

Bridgemohan, P., and Mohammed, M. (2019). The ecophysiology of abiotic and biotic stress on the pollination and fertilization of cacao (*Theobroma cacao* L.; formerly Sterculiaceae family). *Abiotic Biotic Stress Plants* 524, 141–157. doi: 10.5772/intechopen.84528

Castro-Alayo, E. M., Idrogo-Vásquez, G., Siche, R., and Cardenas-Toro, F. P. (2019). Formation of aromatic compounds precursors during fermentation of Criollo and Forastero cocoa. *Heliyon* 5:e01157. doi: 10.1016/j.heliyon.2019.e01157

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cvijović, I., Good, B. H., and Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics* 209:235. doi: 10.1534/genetics.118.301058

D'Auria, J. C. (2006). Acyltransferases in plants: a good time to be BAHD. *Curr. Opin. Plant Biol.* 9, 331–340. doi: 10.1016/j.pbi.2006.03.016

D'Auria, J. C., Chen, F., and Pichersky, E. (2002). Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*. *Plant Physiol.* 130:466. doi: 10.1104/pp.006460

Dayanne, S. M. D. A., Oliveira Jordão Do Amaral, D., Del-Bem, L. E., Bronze Dos Santos, E., Santana Silva Raner, J., Peres Gramacho, K., et al. (2017). Genome-wide identification and characterization of cacao WRKY transcription factors and analysis of their expression in response to witches' broom disease. *PLoS ONE* 12:e0187346. doi: 10.1371/journal.pone.0187346

DeLano, W. L. (2002). Pymol: an open-source molecular graphics tool. *CCP4 Newsl. protein Crystallogr.* 40, 82–92.

Dhaubhadel, S., Farhangkhoee, M., and Chapman, R. (2008). Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. *J. Exp. Bot.* 59, 981–994. doi: 10.1093/jxb/ern046

Ding, Q., Yang, X., Pi, Y., Li, Z., Xue, J., Chen, H., et al. (2020). Genome-wide identification and expression analysis of extensin genes in tomato. *Genomics* 112, 4348–4360. doi: 10.1016/j.ygeno.2020.07.029

El-Sharkawy, I., Manríquez, D., Flores, F. B., Regad, F., Bouzayen, M., Latché, A., et al. (2005). Functional characterization of a melon *alcohol acyl-transferase* gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. *Plant Mol. Biol.* 59, 345–362. doi: 10.1007/s11103-005-8884-y

Fan, S., Zhang, D., Zhang, L., Gao, C., Xin, M., Tahir, M. M., et al. (2017). Comprehensive analysis of *GASA* family members in the *Malus domestica* genome: identification, characterization, and their expressions in response to apple flower induction. *BMC Genomics* 18:827. doi: 10.1186/s12864-017-4213-5

Faraji, S., Ahmadizadeh, M., and Heidari, P. (2021). Genome-wide comparative analysis of *Mg transporter* gene family between *Triticum turgidum* and *Camelina sativa*. *Biometals* 34, 639–660. doi: 10.1007/s10534-021-00301-4

Faraji, S., Filiz, E., Kazemitabar, S. K., Vannozzi, A., Palumbo, F., Barcaccia, G., et al. (2020). The AP2/ERF gene family in *Triticum durum*: genome-wide identification and expression analysis under drought and salinity stresses. *Genes* 11:1464. doi: 10.3390/genes11121464

Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., and Bader, G. D. (2016). Cytoscape. js: a graph theory library for visualisation and analysis. *Bioinformatics* 32, 309–311. doi: 10.1093/bioinformatics/btv557

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, 571–607.

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., et al. (2009). A *BAHD* acyltransferase is expressed in the tapetum of *Arabidopsis* anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* 58, 246–259. doi: 10.1111/j.1365-313X.2008.03773.x

Heidari, P., Abdullah, F.araji, S., and Poczai, P. (2021a). Magnesium transporter Gene family: genome-wide identification and characterization in *Theobroma cacao*, *Corchorus capsularis* and *Gossypium hirsutum* of family Malvaceae. *Agronomy* 11:1651. doi: 10.3390/agronomy11081651

Heidari, P., Ahmadizadeh, M., Izanlo, F., and Nussbaumer, T. (2019). *In silico* study of the CESA and CSL gene family in *Arabidopsis thaliana* and *Oryza sativa*: focus on post-translation modifications. *Plant Gene* 19:100189. doi: 10.1016/j.plgene.2019.100189

Heidari, P., Faraji, S., Ahmadizadeh, M., Ahmar, S., and Mora-Poblete, F. (2021b). New insights into structure and function of TIFY genes in *Zea mays* and *Solanum lycopersicum*: a genome-wide comprehensive analysis. *Front. Genet.* 12:534. doi: 10.3389/fgene.2021.657970

Heidari, P., Mazloomi, F., Nussbaumer, T., and Barcaccia, G. (2020). Insights into the *SAM synthetase* gene family and its roles in tomato seedlings under abiotic stresses and hormone treatments. *Plants* 9:586. doi: 10.3390/plants9050586

Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., et al. (2004). Silencing of *hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase* affects phenylpropanoid biosynthesis. *Plant Cell* 16, 1446–1465. doi: 10.1105/tpc.020297

Jendele, L., Krivak, R., Skoda, P., Novotny, M., and Hoksza, D. (2019). PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Res.* 47, W345–W349. doi: 10.1093/nar/gkz424

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Kim, J.-K., Cho, Y., Lee, M., Laskowski, R. A., Ryu, S. E., Sugihara, K., et al. (2015). BetaCavityWeb: a webserver for molecular voids and channels. *Nucleic Acids Res.* 43, W413–W418. doi: 10.1093/nar/gkv360

Krishna, P. (2003). Brassinosteroid-mediated stress responses. *J. Plant Growth Regul.* 22, 289–297. doi: 10.1007/s00344-003-0058-z

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kusano, H., Li, H., Minami, H., Kato, Y., Tabata, H., and Yazaki, K. (2019). Evolutionary developments in plant specialized metabolism, exemplified by two *transferase* families. *Front. Plant Sci.* 10:794. doi: 10.3389/fpls.2019.00794

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lawrie, D. S., Messer, P. W., Hershberg, R., and Petrov, D. A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet.* 9:e1003527. doi: 10.1371/journal.pgen.1003527

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Levsh, O., Chiang, Y., Tung, C., Noel, J., Wang, Y., and Weng, J. (2016). Dynamic conformational states dictate selectivity toward the native substrate in a substrate-permissive *acyltransferase. Biochemistry* 55, 6314–6326. doi: 10.1021/acs.biochem.6b00887

Li, G., Jones, K. C., Eudes, A., Pidatala, V. R., Sun, J., Xu, F., et al. (2018). Overexpression of a rice *BAHD* acyltransferase gene in switchgrass (*Panicum virgatum* L.) enhances saccharification. *BMC Biotechnol.* 18:54. doi: 10.1186/s12896-018-0464-8

Litz, R. E., Pliego-Alfaro, F., and Hormaza, J. I. (eds.). (2020). *Biotechnology of Fruit and Nut Crops*. CABI. Available online at: https://www.cabi.org/cabebooks/ebook/20083015400

Liu, C., Qiao, X., Li, Q., Zeng, W., Wei, S., Wang, X., et al. (2020a). Genome-wide comparative analysis of the BAHD superfamily in seven Rosaceae species and expression analysis in pear (*Pyrus bretschneideri*). *BMC Plant Biol.* 20:14. doi: 10.1186/s12870-019-2230-z

Liu, C., Wu, Y., Liu, Y., Yang, L., Dong, R., Jiang, L., et al. (2020b). Genome-wide analysis of tandem duplicated genes and their contribution to stress resistance in pigeonpea (*Cajanus cajan*). *Genomics* 113(1 Pt 2), 728–735. doi: 10.1016/j.ygeno.2020.10.003

Lovell, S. C., Davis, I. W., Arendall, W. B. III, De Bakker, P. I. W., Word, J. M., Prisant M. G., et al. (2003). Structure validation by Cα geometry:

φ, ψ and Cβ deviation. *Proteins Struct. Funct. Bioinform.* 50, 437–450. doi: 10.1002/prot.10286

Luo, J., Nishiyama, Y., Fuell, C., Taguchi, G., Elliott, K., Hill, L., et al. (2007). Convergent evolution in the BAHD family of acyl *transferases*: identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J.* 50, 678–695. doi: 10.1111/j.1365-313X.2007.03079.x

McElroy, M. S., Navarro, A. J. R., Mustiga, G., Stack, C., Gezan, S., Peña, G., et al. (2018). Prediction of Cacao (*Theobroma cacao*) resistance to *Moniliophthora* spp. Diseases via genome-wide association analysis and genomic selection. *Front. Plant Sci.* 9;343. doi: 10.3389/fpls.2018.00343

Moglia, A., Acquadro, A., Eljounaidi, K., Milani, A. M., Cagliero, C., Rubiolo, P., et al. (2016). Genome-wide identification of bahd acyltransferases and *in vivo* characterization of HQT-like enzymes involved in caffeoylquinic acid synthesis in globe artichoke. *Front. Plant Sci.* 7:1424. doi: 10.3389/fpls.2016.01424

Molina, I., and Kosma, D. (2015). Role of HXXXD-motif/BAHD acyltransferases in the biosynthesis of extracellular lipids. *Plant Cell Rep.* 34, 587–601. doi: 10.1007/s00299-014-1721-5

Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone, D., Cornejo, O., et al. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* 14:r53. doi: 10.1186/gb-2013-14-6-r53

Musavizadeh, Z., Najafi-Zarrini, H., Kazemitabar, S. K., Hashemi, S. H., Faraji, S., Barcaccia, G., et al. (2021). Genome-wide analysis of potassium channel genes in rice: expression of the *OsAKT* and *OsKAT* genes under salt stress. *Genes* 12:784. doi: 10.3390/genes12050784

Mustiga, G. M., Morrissey, J., Stack, J. C., DuVal, A., Royaert, S., Jansen, J., et al. (2019). Identification of climate and genetic factors that control fat content and fatty acid composition of *Theobroma cacao* L. *Beans. Front. Plant Sci.* 10:1159. doi: 10.3389/fpls.2019.01159

Nawaz, Z., Kakar, K. U., Ullah, R., Yu, S., Zhang, J., Shu, Q. Y., et al. (2019). Genome-wide identification, evolution and expression analysis of cyclic nucleotide-gated channels in tobacco (*Nicotiana tabacum* L.). *Genomics* 111, 142–158. doi: 10.1016/j.ygeno.2018.01.010

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Niño, M. C., Kang, K. K., and Cho, Y. G. (2020). Genome-wide transcriptional response of papain-like cysteine protease-mediated resistance against *Xanthomonas oryzae* pv. *oryzae* in rice. *Plant Cell Rep.* 39, 457–472. doi: 10.1007/s00299-019-02502-1

Page, R. D., and Holmes, E. C. (2009). *Molecular Evolution: A Phylogenetic Approach*. John Wiley & Sons.

Pokou, D. N., Fister, A. S., Winters, N., Tahi, M., Klotioloma, C., Sebastian, A., et al. (2019). Resistant and susceptible cacao genotypes exhibit defense gene polymorphism and unique early responses to *Phytophthora megakarya* inoculation. *Plant Mol. Biol.* 99, 499–516. doi: 10.1007/s11103-019-00832-y

Purseglove, J. W. (1968). *Tropical crops: dicotyledons 1 and 2. Trop. Crop. Dicotyledons 1 2*. Available online at: https://www.cabdirect.org/cabdirect/abstract/19691701550 (accessed September 16, 2018).

Rezaee, S., Ahmadizadeh, M., and Heidari, P. (2020). Genome-wide characterization, expression profiling, and post- transcriptional study of *GASA* gene family. *Gene Reports* 20:100795. doi: 10.1016/j.genrep.2020.100795

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Savojardo, C., Martelli, P. L., Fariselli, P., Profiti, G., and Casadio, R. (2018). BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* 46, W459–W466. doi: 10.1093/nar/gky320

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC -type MADS -box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122

Shen, S., Zhang, Q., Shi, Y., Sun, Z., Zhang, Q., Hou, S., et al. (2020). Genome-wide analysis of the *NAC* domain transcription factor gene family in *Theobroma cacao*. *Genes* 11:35. doi: 10.3390/genes11010035

Singh, A. P., and Savaldi-Goldstein, S. (2015). Growth control: brassinosteroid activity gets context. *J. Exp. Bot.* 66, 1123–1132. doi: 10.1093/jxb/erv026

Song, X., Li, E., Song, H., Du, G., Li, S., Zhu, H., et al. (2020). Genome-wide identification and characterization of nonspecific *lipid transfer protein (nsLTP)* genes in *Arachis duranensis*. *Genomics* 112, 4332–4341. doi: 10.1016/j.ygeno.2020.07.034

Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016). Protein binding pocket dynamics. *Acc. Chem. Res.* 49, 809–815. doi: 10.1021/acs.accounts.5b00516

St-Pierre, B. (2000). Evolution of acyltransferase genes: origin and diversification of the *BAHD* superfamily of acyltransferases involved in secondary metabolism. in *Recent Advances in Phytochemistry* (Elsevier Inc.) 285–315. doi: 10.1016/S0079-9920(00)80010-6

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131

Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 46, W363–W367. doi: 10.1093/nar/gky473

Waseem, M., Ahmad, F., Habib, S., and Li, Z. (2018). Genome-wide identification of the auxin/indole-3-acetic acid (*Aux/IAA*) gene family in pepper, its characterisation, and comprehensive expression profiling under environmental and phytohormones stress. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-30468-9

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi: 10.1093/nar/gkm290

Yang, S., Zhang, X., Yue, J. -X., Tian, D., and Chen, J. -Q. (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genomics* 280, 187–198. doi: 10.1007/s00438-008-0355-0

Yu, X.-H., Chen, M.-H., and Liu, C.-J. (2008). Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7- *O*-glycosidic (iso)flavones in *Medicago truncatula*. *Plant J.* 55, 382–396. doi: 10.1111/j.1365-313X.2008.03509.x

Yu, X. H., Gou, J. Y., and Liu, C. J. (2009). *BAHD* superfamily of acyl-CoA dependent acyltransferases in Populus and Arabidopsis: bioinformatics and gene expression. *Plant Mol. Biol.* 70, 421–442. doi: 10.1007/s11103-009-9482-1

Zan, T., Li, L., Xie, T., Zhang, L., and Li, X. (2020). Genome-wide identification and abiotic stress response patterns of abscisic acid stress ripening protein family members in *Triticum aestivum* L. *Genomics* 112, 3794–3802. doi: 10.1016/j.ygeno.2020.04.007

Zhang, T., Huo, T., Ding, A., Hao, R., Wang, J., Cheng, T., et al. (2019). Genome-wide identification, characterization, expression and enzyme activity analysis of coniferyl alcohol acetyltransferase genes involved in eugenol biosynthesis in *Prunus mume*. *PLoS ONE* 14:e0223974. doi: 10.1371/journal.pone.0223974

Zheng, L., Ying, Y., Wang, L., Wang, F., Whelan, J., and Shou, H. (2010). Identification of a novel iron regulated basic helix-loop-helix protein involved in Fe homeostasis in *Oryza sativa*. *BMC Plant Biol.* 10:166. doi: 10.1186/1471-2229-10-166

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.