*Bulletin of Biological and Allied Sciences Research*
***ISSN:* 2521-0092**
www.bbasr.org
DOI: https://doi.org/10.64013/bbasr.v2026i1.133
*Bull. Biol. All. Sci. Res., Volume, **11**: 133*
***Original Research Article*** ***Open Access***

# GENOME-WIDE EVALUATION OF THE *YABBY* TRANSCRIPTION FACTOR GENE FAMILY IN *THEOBROMA CACAO* AND COMPARATIVE ANALYSIS WITH *ARABIDOPSIS THALIANA*

**HAYYAT Q[1*], AMJAD MS[2], ZOHAIB MA[3]**

[1]*Department o Plant Breeding and Genetics, Faculty of Agriculture Sciences, University of Punjab, Lahore, Pakistan*
[2]*Department of Entomology, Faculty of Agriculture and Environmental Sciences, The Islamia University of Bahawalpur, Pakistan*
[3]*Department of Agronomy, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan*
**Correspondence Author Email Address: qaisarhayyat2@gmail.com*

*(Received, 4th February 2026, Accepted 17th July 2026, Published 28th July 2026)*

**Abstract** *YABBY gene family consist of plant transcription factors, which also share a DNA-binding domain with a number of other activities including style length regulation in flowering plants and polarity establishment of lateral organs. The genome of Theobroma cacao was utilized to discover members of the YABBY gene family via some online web techniques. Structure of the gene, location on the chromosome, protein motifs, phylogenetic investigation, synteny and dual synteny analysis, cis-regulatory elements, subcellular localization and functional enrichment, and protein–protein interaction networks were analyzed. In this case, 7 YABBY genes were identified at random on all 10 chromosomes. Cacao YABBY proteins formed five sub-categories (AtINO, AtCRC, AtYAB5, AtAFO and AtYAB2) based on the familiar Arabidopsis classification. Comparison of the synteny of cacao and Arabidopsis showed that a conserved orthologous relationship was present between the two species, indicating the evolutionary presence of YABBY genes in dicot species. Ka/Ks analysis showed that TcYABBY genes have mostly evolved under severe purifying pressure. Promoter analysis revealed the presence of multiple hormone- and stress-responsive cis-regulatory elements, suggesting potential regulatory roles in growth and environmental responses. Functional enrichment and interaction analyses further supported the involvement of TcYABBY proteins in developmental and metabolic regulatory pathways. Comprehensively, this research paper offers an extensive framework of YABBY gene evolution and functional potential of cacao and is an important source of future functional characterization and breeding-related studies.*

[*Citation:* Hayayt, Q., Amjad, M.S., Zohaib, M.A. (2026). Genome-wide evaluation of the *YABBY* transcription factor gene family in *Theobroma cacao* and comparative analysis with *Arabidopsis thaliana*. *Bull. Biol. All. Sci. Res.* **11***: 133.* https://doi.org/10.64013/bbasr.v2026i1.133]
**Keywords*:*** *Theobroma cacao; YABBY transcription factors; Genome-wide analysis; Phylogenetic analysis; Synteny and gene duplication; Arabidopsis thaliana*

## Introduction

The regulation of genes depends on a number of biological activities in plants including development, stress response and hormone signaling, among others, and transcription factors (TFs) are one of the most important regulators. The most distinct of these is the *YABBY* transcription factor family, which is a distinctive plant family with a conserved N-terminal C2C2 -like zinc-finger domain and a C-terminal *YABBY* (*HMG-like*) DNA-binding domain. The possession of all *YABBY* genes is mainly in determining the abaxial cell fate of lateral organs, floral and leaf polarity, lamina growth, as well as the development of floral organs(Zhang et al., 2020). *Arabidopsis thaliana* has been reported to have a total of six *YABBY* genes that have been sorted into unique subfamilies, which are *FILAMENTOUS FLOWER (FIL/YAB1), YAB2, YAB3, YAB5, CRABS CLAW (CRC),* and *INNER NO OUTER (INO)*. Functional studies have demonstrated their roles in leaf development, carpel and nectary formation, and ovule integument development. Comparative analyses across plant species indicate that *YABBY* genes are evolutionarily conserved but have undergone lineage-specific expansion and diversification(Han et al., 2025). *Theobroma cacao* is a valuable perennial tree crop with significant economic value and is the major producer of cocoa and chocolate. Despite the availability of a high-quality genome sequence, transcription factor families in cacao remain underexplored. The structural changes of the *YABBY* gene family in cacao can be examined on a genome-wide scale, offering new suggestions on how the gene family can be utilized in agronomic traits through study of its development, evolution, and conservation(Li et al., 2019; Zhao et al., 2023). The present study aims to (i) identify *YABBY* genes in the cacao genome, (ii) analyze their phylogenetic relationships and structural features, (iii) investigate chromosomal distribution and duplication events, (iv) assess evolutionary selection pressure, (v) examine

regulatory elements and functional enrichment, and (vi) compare cacao *YABBY* genes with their *Arabidopsis* counterparts through synteny and interaction analysis(Liu et al., 2025).

***Materials and Methods***

**Cacao Identification of the YABBY genes.**

*YABBY* PSTrFs (Plant-specific transcription factors) gene amino acid sequences were taken from Pfam (http://pfam.xfam.org), i.e., PF04690. 164 Amino acid sequences of the *YABBY* gene were separated from the *Arabidopsis thaliana* (Accession No. A0A1P8APE2)(Buttar et al., 2020; Hao et al., 2022). Using this sequence, further genes containing coding for *YABBY* proteins were discovered using the database of the proteome at the Ensemble Plant website (https://plants.ensembl.org) utilizing the BLAST-P (protein alignment search tool) program. Ensemble Plant site provided *TcYABBY* amino acid sequences, which were inputted in motif finder (https://www.genome.jp) and national center of biotechnology information CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using already set parameters. The proteins lacking real *YABBY* conserved domains in their sequence were excluded. The subjective studies reduced the protein sequences absent in the conserved domain of *YABBY* proteins. The sequences of known *YABBY* proteins in *Arabidopsis thaliana* served as queries to screen the candidate *YABBY* genes in cacao as provided by sequence similarity searches. The presence of the *YABBY* domain was additional confirmation of the candidate proteins through domain prediction tools. The redundant sequences were eliminated, and the rest of the genes were named as *TcYABBY1* to *TcYABBY7* based on the position of genes on the chromosome(Ren et al., 2025).

**Phylogenetic analysis**

*The* amino acid sequence alignment and the phylogenetic analysis were done using the MEGA 11 v11.0.10 (Tamura et al., 2021) program using the neighbor-joining (NJ) method and bootstrapping held constant at 1000 replications with pairwise deletion. For phylogenetic exploration, 7 cacao *YABBY* proteins and 6 *Arabidopsis YABBY* protein sequences were utilized(Xi et al., 2023).

**Investigation of chemical and physical characteristics of *TcYABBY* proteins**

The online *ProtParam* web tool was used to depict amino acid length, isoelectric values, and molecular weight of the molecules(Mariyam et al., 2021). *ProtParam* tool was used to predict the protein length (amino acid residues), molecular weight, and theoretical pI of the *YABBY* proteins (http://web.expasy.org/protparam). The data in terms of the gene ID, the location of the gene on the chromosome, and the sequence of the gene and protein were obtained on plants Ensemble. The *YABBY* genes were renamed to the original position in the order that was found on the respective website. The online web tool *WoLFPSORT* was used to identify subcellular localization of *TcYABBY* genes (https://wolfpsort.hgc.jp) (Table 1)(Akbudak et al., 2020).

**Gene structure and conserved motif analysis**

To predict the arrangement of exons and introns, CDS and genomic data of *TcYABBY* genes were taken from the online website of Ensemble Plants. Furthermore, other full genomic, gff, and protein files of the cacao genome were also taken from Ensemble Plants (Sen, 2022). Sequences of these genes were inserted into Gene Structure Display Server (GSDS v2.0) (http://gsds.cbi.pku.edu.cn) to draw sequences were further used to draw the gene structure using Gene Structure Display Server (GSDS v2.0) (available at http://gsds.cbi.pku.edu.cn). MEME suite website was utilized to figure out the number of motifs using Multiple Em for Motif Elicitation (MEME) software with a total no. of motifs customized to 10 (https://meme-suite.org/meme/tools/meme) (Adak et al., 2023). The width of the motifs was customized between 6 and 50, including a few other parameters.

**Chromosomal localization and duplication analysis**

*TcYABBY* genes were physically mapped on cacao chromosomes. Based on the data of the GFF annotation file of the T. cacao genome, all *YABBY* genes were mapped on their corresponding chromosomes, and their positions were visualized using TBtools. The occurrences of gene duplication were detected based on sequence similarity and the location of the gene(Sen, 2022).

**Synteny analysis with *Arabidopsis***

Comparative synteny analysis was performed between cacao and *Arabidopsis* genomes to identify orthologous *YABBY* gene pairs and conserved genomic blocks. The Multiple Collinearity Scan toolkit (MCScanX) was used for a check of duplication events of the genes, using parameters that were already set. Arabidopsis was also subjected to dual synteny with cacao. The syntenic maps were developed using Advanced Circos within the TB tools to illustrate the syntenic relationship between paralogous gene pairs YABBY obtained in the cacao(Chen et al., 2025a).

**Evolutionary selection pressure (Ka/Ks) analysis**

Nonsynonymous (Ka) and Synonymous (Ks) replacement values for the duplicated *TcYABBY* gene pairs were calculated, and the ratios Ka/Ks should be used to estimate the rate of selection(Gain et al., 2024).

**Cis-regulatory element analysis**

The upstream sequences of *TcYABBY* genes were checked for the presence of potential cis-acting regulatory element using PlantCARE database(https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). For every *TcYABBY* gene, a promoter sequence of 2, 000 bp upstream from the transcription start site was taken from the Ensembl Plants database and sent to the PlantCARE web server. The analysis was done using default

parameters. *PlantCARE* was utilized to locate and assign cis-regulatory elements that are related to hormonal regulation, light regulation, stress responses, and developmental processes. The output consisted of the position, strand orientation, and functional annotation of each recognized cis element. The identified cis-regulatory elements were first checked manually, then grouped according to their known biological functions, and finally presented in a table form to show their similarity/difference among the *TcYABBY* genes. The findings were also displayed visually with the help of TBtools(Gao et al., 2017; Yilmaz et al., 2025).

**Subcellular localization prediction**

The subcellular localization of the proteins encoded by the *TcYABBY* genes was predicted by the online WoLF PSORT web tool(https://wolfpsort.hgc.jp) and visualized as a heatmap using TBtools (Horton et al., 2006).

**Functional enrichment and protein–protein interaction analysis**

Gene Ontology and pathway enrichment analyses were performed to consider the biological functions of *TcYABBY* genes. The protein sequences of the *TcYABBY* genes were submitted to the BUSCA (Bologna Unified Subcellular Component Annotator) web server for the first time (https://busca.biocomp.unibo.it/) to get corresponding Gene Ontology identifiers(Nadeau et al., 2021). The gene IDs that were found were then used as input for an enrichment analysis via the ShinyGO v0.77 online platform (https://bioinformatics.sdstate.edu/go/).

*ShinyGO* was utilized to detect the significantly enriched GO terms that represented biological processes, molecular functions, and cellular components. Default settings were applied, and the evaluation of statistical significance was determined based on the false discovery rate (FDR)- adjusted p-values. Enrichment results were illustrated as dot plots showing the fold enrichment values and the number of genes participating in them. Protein and protein interaction networks were generated through the *STRING* database (http://string-db.org/) using a combination of predicted and known interaction evidence, such as gene neighborhood, gene fusion, and gene co-occurrence, etc. Additional evidence sources included text mining, co-expression, and protein homology, which were also considered. These integrated interaction channels were used to infer functional associations among *TcYABBY* proteins(Fowler et al., 2011; Pathan et al., 2015).

**Results**

**Identification and classification of *TcYABBY* genes**

Seven *YABBY* genes were identified in the cacao genome and designated *TcYABBY*1 to *TcYABBY*7. All identified proteins contained the conserved *YABBY* domain, confirming their classification as *YABBY* transcription factors. Phylogenetic analysis grouped these genes into distinct clusters, indicating diversification within the cacao *YABBY* family. To identify 7 *TcYABBY* genes, the sequence of the *YABBY* domain was blasted against the entire genome sequence of Cacao that was retrieved from the Ensemble database. The first proteomic BLAST resulted in the discovery of a total of 11 *YABBY* proteins. Among these proteins, only those were retained for further analysis that contained the complete *YABBY* domain, while remaining of the proteins which have incomplete or identical protein forms were excluded. Finally, for further investigational study, 7 fully differentially sequenced *TcYABBY* genes were selected (Table 1). The length of amino acids of *TcYABBY* genes fell within the range of 180 to 230 amino acids in length, while the molecular weight was between 19. 09 and 24. 777 kD, with *TcYABBY1* as the shortest and *TcYABBY6* being the longest protein. (Table 1)(Eshed et al., 2004). The pI value of the detected proteins ranged from 8.13 to 9.08. The subcellular localization of these 7 *YABBY* genes was determined by using the online web tool *WoLFPSORT* (https://wolfpsort.hgc.jp/).

***Table 1. Accession information and basic characteristics of TcYABBY genes in Theobroma cacao***

| Transcript ID | Gene Name | Chr ID | Start | End | Strand | AA | Exon | Intron | Weight (KDs) | pI | Instability Index | Aliphatic Index | Grand Average of Hydropathicity (Gravy) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Thecc.01G380200.1.v2.1** | TcYABBY1 | Chr1 | 37921030 | 37923860 | + | 181 | 8 | 7 | 20.43919 | 8.96 | 45.24 | 73.81 | -0.551 |
| **Thecc.03G257400.1.v2.1** | TcYABBY2 | Chr3 | 31935208 | 31938340 | - | 211 | 7 | 6 | 23.47669 | 7.71 | 46.7 | 81.33 | -0.325 |
| **Thecc.09G267900.1.v2.1** | TcYABBY3 | Chr9 | 36312754 | 36320397 | - | 182 | 6 | 5 | 20.35698 | 8.42 | 47.53 | 65.44 | -0.526 |
| **Thecc.09G128600.1.v2.1** | TcYABBY4 | Chr9 | 7045052 | 7049007 | - | 184 | 7 | 6 | 20.78662 | 8.13 | 50.08 | 74.24 | -0.406 |
| **Thecc.07G117600.1.v2.1** | TcYABBY5 | Chr7 | 7884417 | 7889767 | - | 212 | 7 | 6 | 23.45482 | 8.24 | 47.81 | 76.32 | -0.297 |
| **Thecc.02G254000.1.v2.1** | TcYABBY6 | Chr2 | 33732346 | 33735235 | - | 222 | 7 | 6 | 24.77014 | 8.13 | 59.96 | 69.46 | -0.517 |
| **Thecc.02G326600.1.v2.1** | TcYABBY7 | Chr2 | 41006601 | 41008631 | - | 173 | 7 | 6 | 19.09077 | 9.08 | 64.91 | 63.7 | -0.558 |

**Phylogenetic relationship across plant species**

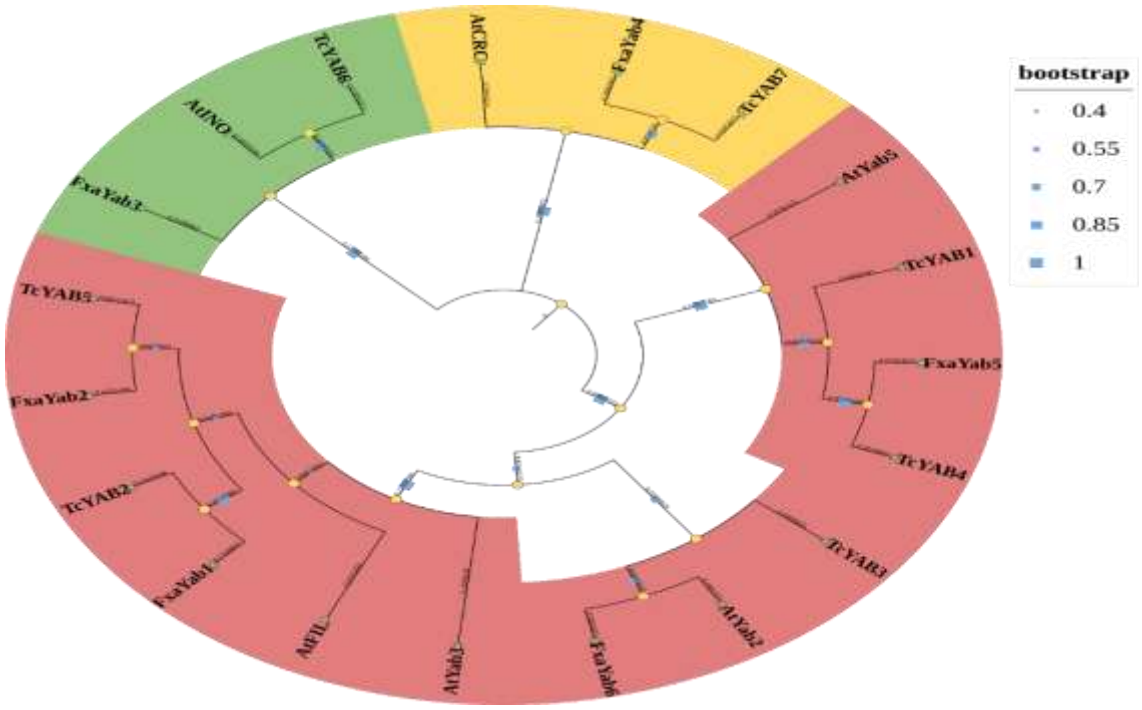


**Figure 1 Comparison of YABBY phylogenetic analysis of transcription factor proteins of Theobroma cacao, Arabidopsis thaliana and other related plant species. The Neighbor-Joining (NJ) method was used to provide the construction of a circular phylogenetic**

A circular phylogenetic tree with Neighbor-Joining (NJ) settings was created using MEGA 11 software by aligning full-length protein sequences, including *YABBY* proteins from cacao, *Arabidopsis*, and other species, i.e., *Fragaria anassa*, revealing that *TcYABBY* genes clustered with known *YABBY* subfamilies(Fitzpatrick et al., 2017). The findings illustrated that 7 *TcYABBY* proteins were scattered in 5 subgroups, which were called *AtINO, AtCRC, AtYAB5, AtAFO/AtYAB3, AtYAB2* (Figure 1).

**Gene structure organization**

The arrangement of introns and exons is very helpful in proofreading the evolutionary relationship between genes and different organisms. The total number of introns and exons serves as an evolutionary ID for different gene families. The phylogenetic analysis along with the intron and exon structure showed that the pattern was quite similar between the phylogenetic clades(Brown et al., 1989). The number of introns changed from 5 to 7 in cacao (Table 1). All 7 *TcYABBY* genes contain introns and exons with a little variation in quantity. *TcYABB1* consisted of 7 introns and 8 exons (14.29%), *TcYABB3* contained 5 introns and 6 exons (14.29%), whereas *TcYABB2*, *TcYABB4*, *TcYABB5*, *TcYABB6*, and *TcYABB7* each harbored 6 introns and 7 exons (71.43%) (Table 1 and Figure 2).

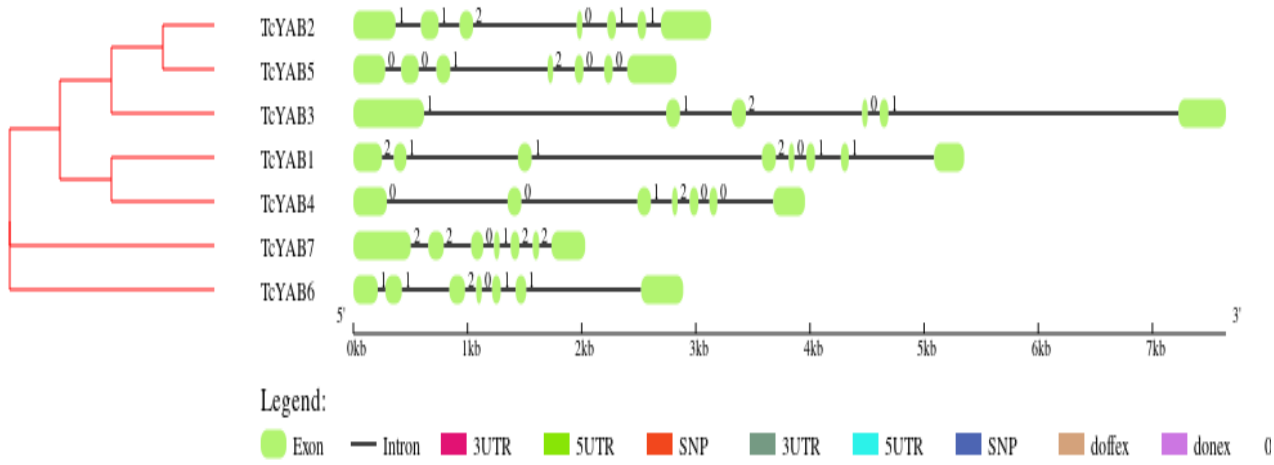


***Figure 2 Theobroma cacao Gene structure organization of TcYABBY genes. The exon-intron structure of TcYABBY genes is drawn based on their phylogenetic data. The exons are illustrated by green boxes, introns by black lines, and untranslated regions (UTRs)***

**Conserved motifs and domain architecture**

Motif analysis identified multiple conserved motifs shared among *TcYABBY* proteins. The *YABBY* domain was consistently present across all members, while clade-specific motifs suggested functional divergence. Motif Finder, an online web tool, was used to distinguish and recognize the arrangements of different motifs. All 7 *YABBY* protein sequences were computed for motif analysis. Total no of different motifs was found to be 10(Gushchina et al., 2018). *YABBY* domains were constantly present in all the identified *TcYABBY* proteins. Two other motifs, named Hmg_box and Hmg_box2, were also present in all 7 *YABBY* proteins. HMG box motifs are recognized to bind DNA in the minor groove and to curve the DNA to accurate severe angles. It was noted that genes located in the same group contained motifs whose sequence patterns were alike, from which we can assume that the highly consistent motifs have a specific function related to that specified group. All the cacai *YABBY* genes contain the same *YABBY* domain, but its arrangement was a little varied in some genes. Evolutionary gene expansion might be the cause of the *YABBY* domains having similarity in the motif patterns in different groups (Figure 3).

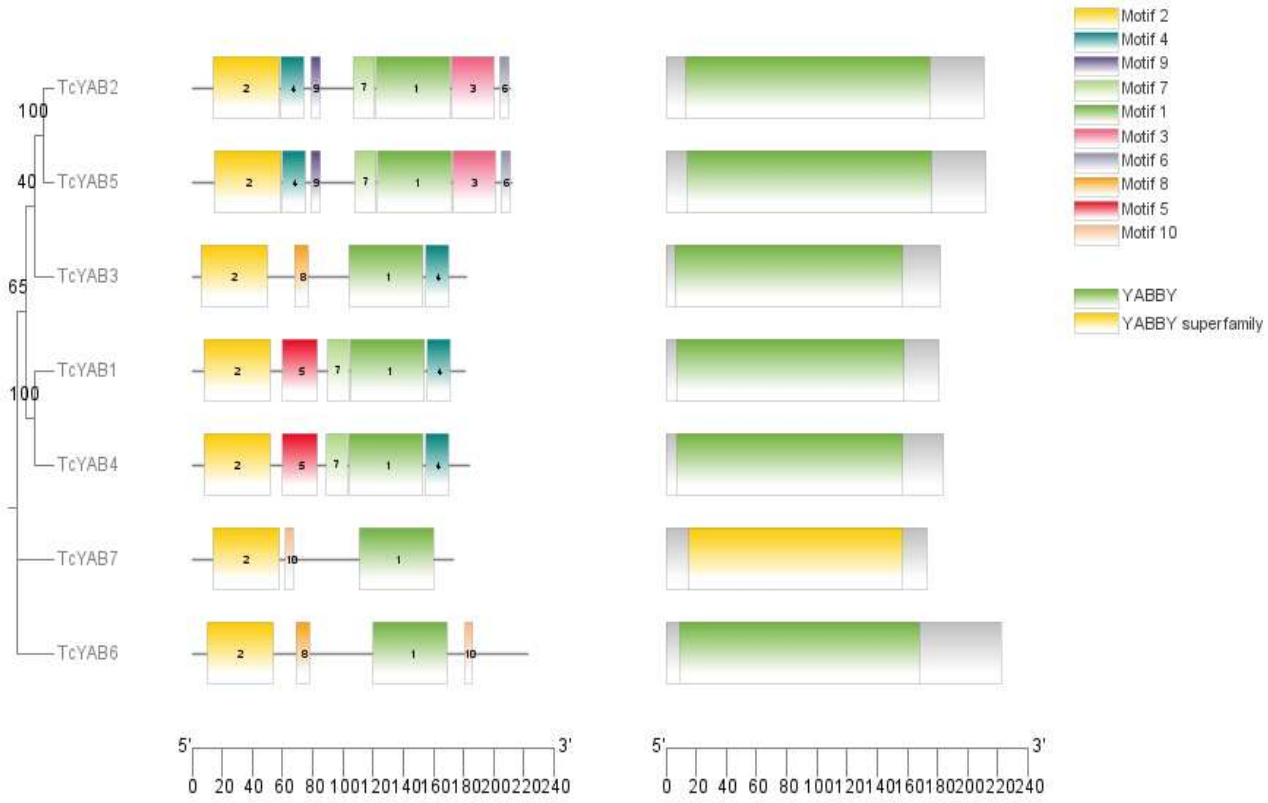


**Figure 3 Conserved motif composition and domain architecture of TcYABBY proteins in Theobroma cacao. On the left is the phylogenetic relationship among TcYABBY proteins, and the middle panel shows the distribution of conserved motifs identified using MEM**

**Chromosomal distribution of *TcYABBY* genes**

*TcYABBY* genes were unevenly distributed across cacao chromosomes. Some chromosomes harbored single genes, while others contained multiple members, indicating non-random genomic distribution. Analysis of the gene mapping of cacao *YABBY* genes revealed that *TcYABBY* genes were located unevenly on different chromosomes. Chromosomal distribution analysis given that *TcYABBY* genes were unevenly divided across the genome. A total of two *TcYABBY* genes (*TcYABBY6* and *TcYABBY7*) were located on chromosome 2, while chromosome 9 harbored two genes **(***TcYABBY3* and *TcYABBY4***)**. In contrast, chromosome 1 **(***TcYABBY1***),** chromosome 3 (*TcYABBY2*), chromosome 7 (*TcYABBY5*), and chromosome 8 (*TcYABBY8*) each contained a single *TcYABBY* gene. No *TcYABBY* genes were detected on the remaining chromosomes(Obe et al., 2002). Furthermore, synteny and duplication analysis revealed that *TcYABBY* genes exhibited several segmental duplication relationships across different chromosomes. The duplicated gene pairs were distributed across chromosomes 2, 7, and 9, indicating that the expansion of the *TcYABBY* gene family was mainly driven by segmental duplication events, while no clear tandem duplication events were observed (Figure 4).

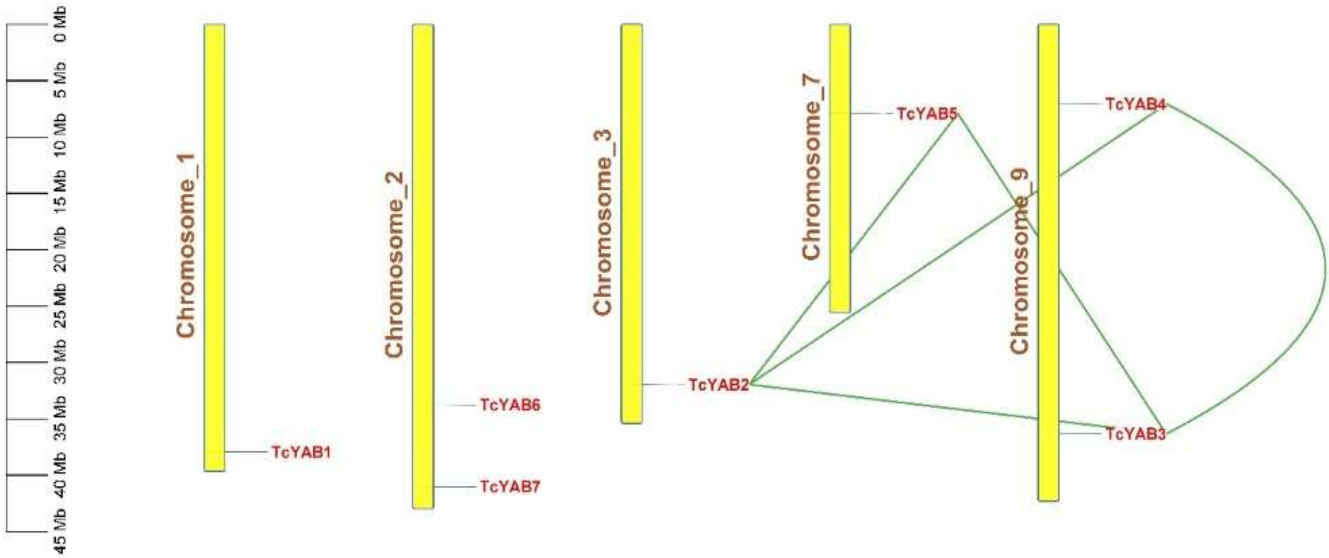


***Figure 4 Chromosomal distribution and duplication analysis of TcYABBY genes in Theobroma cacao. The locations of the TcYABBY genes physically on the cacao chromosomes are depicted by vertical yellow bars with scale markers indicating chromosomal length***

**Synteny between T. cacao and *A. thaliana***

Dual synteny analysis between *Theobroma cacao* and *Arabidopsis thaliana* was conducted using TBtools to investigate the evolutionary conservation of the *TcYABBY* gene family (Figure 5). The synteny map revealed that *TcYABBY* genes were distributed across all ten chromosomes of *T. cacao* and showed clear collinear relationships with corresponding regions in the *Arabidopsis* genome. Multiple inter-genomic syntenic links were observed between T. cacao chromosomes and *Arabidopsis* chromosomes, indicating that several *TcYABBY* genes share conserved orthologous relationships with their Arabidopsis counterparts(Prewitt et al., 2021). The presence of these conserved syntenic blocks suggests that the *YABBY* gene family is evolutionarily conserved between *T. cacao* and *A. thaliana* and that these genes likely originated from common ancestral genomic regions before species divergence.

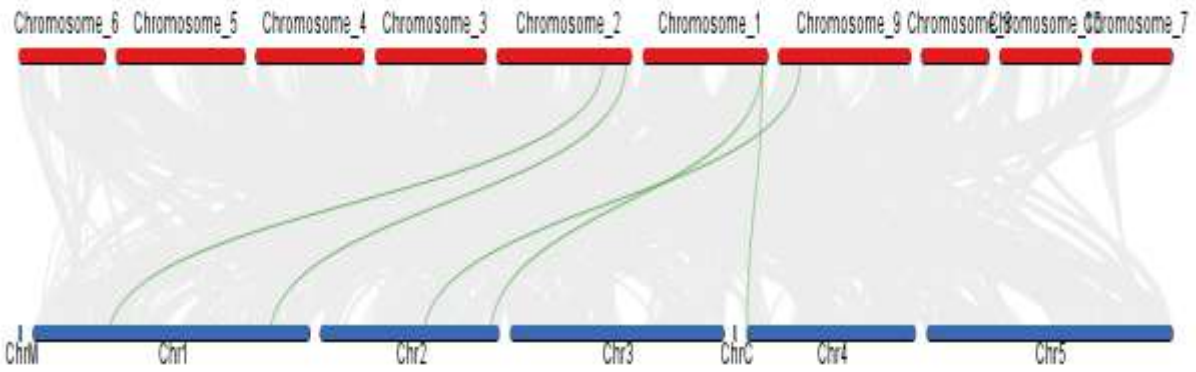


***Figure 5 Dual synteny analysis of YABBY genes between Theobroma cacao and Arabidopsis thaliana. Red bars represent cacao chromosomes, while blue bars represent chromosomes of Arabidopsis. Green lines denote collinear orthologs between the two species***

**Duplication and collinearity within cacao**

A Circos plot was constructed to visualize the chromosomal distribution and duplication relationships of *TcYABBY* genes in *Theobroma cacao* (Fig. 6). The *TcYABBY* genes were distributed across multiple chromosomes, indicating a non-uniform chromosomal organization of the *YABBY* gene family. Several inter-chromosomal links were observed among *TcYABBY* genes, as represented by red connecting lines, suggesting the presence of segmental duplication events. Notably, *TcYABBY3* and *TcYABBY4* showed multiple syntenic connections with other *TcYABBY* members, highlighting their involvement in duplicated genomic blocks. In contrast, no closely adjacent gene pairs were detected on the same chromosome, indicating the absence of tandem duplication(Brown et al., 2008). Overall, the Circos analysis demonstrates that the expansion of the *TcYABBY* gene family in *T. cacao* is mainly attributed to segmental duplication events distributed across different chromosomes (Figure 6).

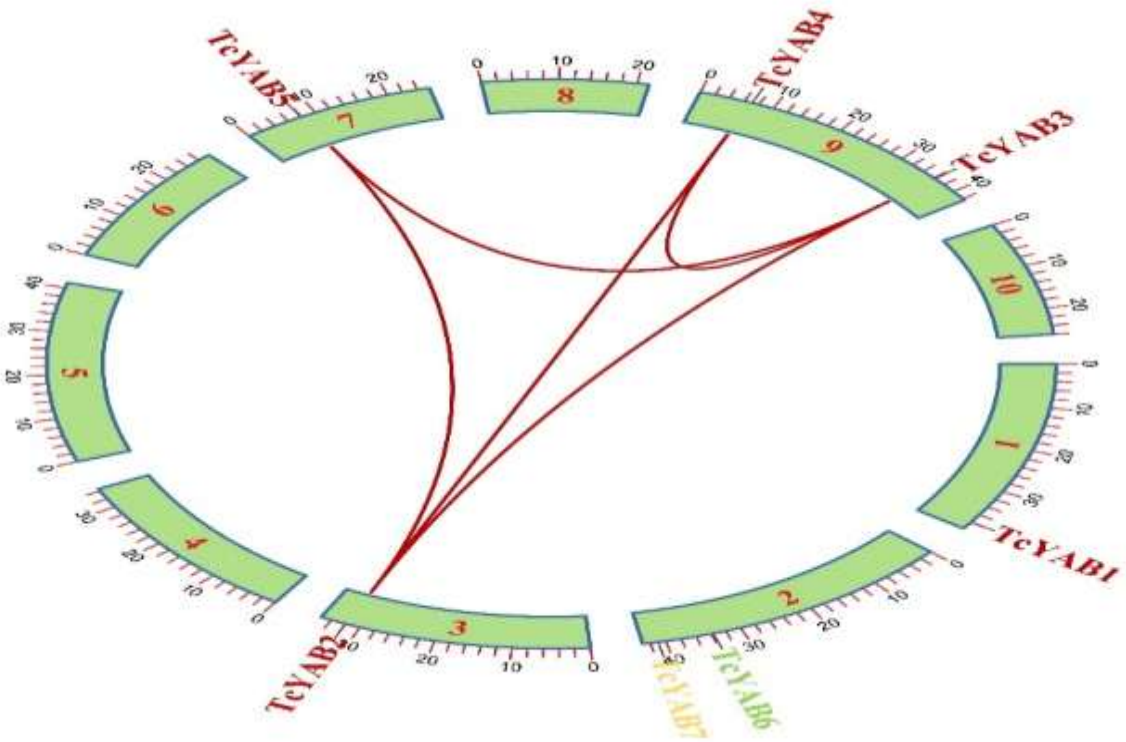


***Figure 6. A Circos diagram is a graphical representation of the duplication and collinearity relationships between TcYABBY gene pairs of Theobroma cacao. The chromosomes are shown in a round layout, and the red lines between them illustrate segmental.***

**Evolutionary selection pressure**

The duplication and evolutionary divergence of *TcYABBY* genes were studied through the computation of Ka, Ks, and Ka/Ks ratios with the help of the simple Ka/Ks calculator that is part of the *TBtools* package (Figure 7). Here, Ka stands for the number of nonsynonymous substitutions per nonsynonymous site, and Ks represents the number of synonymous substitutions per synonymous site. Meanwhile, the Ka/Ks ratio indicates the magnitude of the selective pressure on the duplicated gene pairs. The Ka/Ks ratios for the *TcYABBY* paralogous gene pairs that were analyzed varied between 0. 128 and 0. 384. Specifically, the *TcYABBY3_TcYABBY4* and *TcYABBY3_TcYABBY5* pairs exhibited the lowest Ka/Ks values (0.128), while the *TcYABBY2_TcYABBY3* pair showed the highest ratio (0.384). The corresponding Ks values varied from 0.995 to 2.245, with the highest Ks observed for the *TcYABBY3_TcYABBY4* and *TcYABBY3_TcYABBY5* gene pairs. Notably, all analyzed *TcYABBY* paralogous pairs displayed Ka/Ks ratios less than 1, indicating that these genes have undergone strong purifying (negative) selection following duplication. This suggests that duplicated *TcYABBY* genes are evolutionarily conserved and that functional divergence has occurred under selective constraints rather than positive selection(Back, 1994).

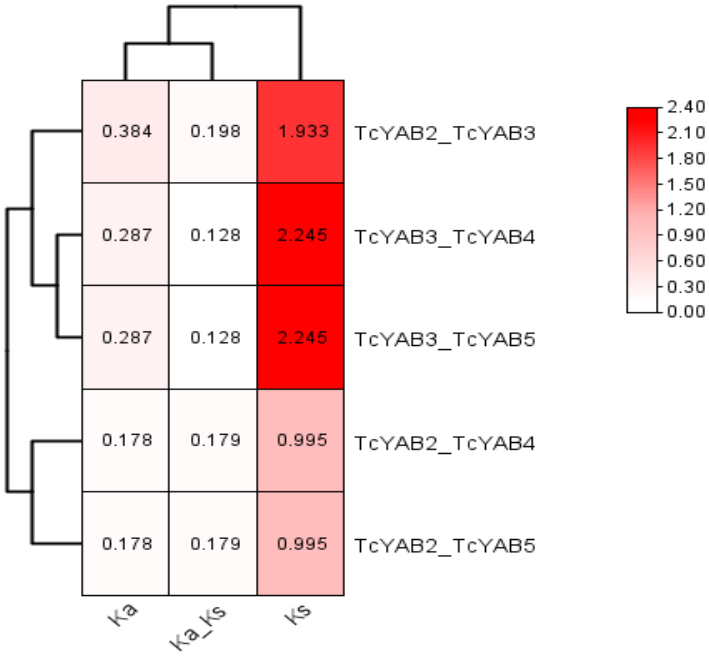


**Figure 7 Analysis of evolutionary selection pressure on duplicated TcYABBY gene pairs. Heatmap clustering displays the calculated Ka, Ks, and Ka/Ks ratios using TBtools. All gene pairs show Ka/Ks ratios less than 1.**

**Cis-regulatory element analysis**

The binding sites of the transcription factors in the promoter constitutions have been thought to be the key molecular players of the gene expression

regulation mechanisms, as they affect the positioning of the cis-regulatory elements. To explore the possible regulatory pathways of the *TcYABBY* gene family, the insilico examination of the upstream promoter regions was carried out. A huge range of cis-acting regulatory elements related to light responsiveness, hormone signaling, stress responses, and developmental regulation were found in the sequences of *TcYABBY* genes (Table 2, Figure 8). *TcYABBY* gene promoter areas were identified to be highly enriched with a variety of light-responsive elements, including *G-box*, *Box 4*, *GT1-motif*, *MRE*, *TCT-motif*, and *TCCC-motif*, which implies that *TcYABBY* genes might play a role in light-mediated transcriptional regulation. Along with that, several phytohormone-responsive elements were found, such as *ABRE* (*abscisic acid responsiveness*), *TGACG*-motif and *CGTCA*-motif (methyl jasmonate responsiveness), *TATC-box* and *P-box* (*gibberellin responsiveness*), and *TCA*-element (*salicylic acid responsiveness*); hence, there is an indication of the involvement of *TcYABBY* genes in hormone-regulated signaling pathways(Wittkopp and Kalay, 2012). In addition, a variety of stress-related cis-elements were observed, including *ARE* (*anaerobic induction*), *MBS* (*drought inducibility*), *LTR* (*low-temperature responsiveness*), *WUN-motif* (*wound response*), and *TC-rich* repeats associated with defense and stress regulation. Elements involved in growth and tissue-specific regulation, such as *CAT-box* (*meristem expression*), *O2-site* (*zein metabolism*), *RY-element* (*seed-specific regulation*), *GCN4-motif* (*endosperm expression*), *MSA-like* element *(cell-cycle regulation*), and *AT-rich* elements related to DNA-binding protein activity, were also identified among *TcYABBY* promoters(Wang et al., 2007). Overall, the diverse composition and variable distribution of cis-regulatory elements among *TcYABBY* genes suggest their functional diversification and complex transcriptional regulation in *Theobroma cacao*.


***Figure 8 Cis, regulatory element distribution around promoter regions of TcYABBY genes. The heatmap shows the number and locations of various cis, acting regulatory elements that are linked to light responsiveness, hormone signaling, stress responses.***

***Table 2* The spatio-temporal functional distribution of the patterns of *cis-regulatory element* of the *YABBY* gene in different tissues and organs through biological growth of the plant**

| Sr # | Cis-Elements | Function | References |
|---|---|---|---|
| **1** | AAGAA-motif | Transcriptional regulation and responsiveness to the stress regulatory element, which causes cis-acting. | (Yamaguchi-Shinozaki and Shinozaki, 2005) |
| **2** | ABRE | Abscisic acid (ABA) receptor Cis-acting factor | (Xiao-Li et al., 2011) |
| **3** | CAAT-box | Oncolytic promoter elements. Crucially prevalent transcription initiation promoter elements, as well as promoter efficiency. | (Papadakis et al., 2004) |

| | | | |
|---|---|---|---|
| **4** | G-Box | Cis-acting regulatory factor in light responsiveness and hormone signaling. | (Farrera-Sal et al., 2020) |
| **5** | TATA-box | Promoter element that is core and is situated close to the point at which transcription will initiate; an aspect that is a must-have to get bound by RNA polymerase. | (Smale et al., 1998) |
| **6** | W box | WRKY binding site for transcription factors; defense response and pathogen signal. | (Lenhard et al., 2012) |
| **7** | ERE | ERE, which is a part of the ethylene signaling pathway. | (Saukkonen and Hemminki, 2004) |
| **8** | GT1-motif | The cis-acting regulatory element is light-responsive. | (Kawakami et al., 2024) |
| **9** | MYB | MYB transcription factor binding site in the process of stress response and development. | (Nettelbeck et al., 2002) |
| **10** | MYB-like sequence | As a part of the regulatory aspect linked to transcription instigated by MYB and response to stress. | (Wang et al., 2021) |
| **11** | STRE | Stress element that is sensitive to general stress. | (Roy, 2016) |
| **12** | WRE3 | Wound-induced regulatory component of injury and defense signaling. | (Chen et al., 2025b) |
| **13** | Box 4 | Light-responsive DNA module that is conserved. | (Harshith et al., 2024) |
| **14** | GATA-motif | Thiol-lysine dipeptide light-sensitive motif of photoregulation and chlorophyll synthesis. | (Lee and Seo, 2022) |
| **15** | GC-motif | Gene transcription and light responsiveness regulatory element. | (Shrestha and Huang, 2022) |
| **16** | P-box | Element in the growth control triggered by gibberellin. | (Ye et al., 2022) |
| **17** | TCA | The regulatory element controlling salicylic acid responsiveness that is cis-acting. | (Tang et al., 2025) |
| **18** | ARE | Cis-acting factor in anaerobic induction. | (Lowry and Zitomer, 1984) |
| **19** | MYC | MYC transcription factor binding site in stress and jasmonic acid signaling. | (Stirnberg and Happe, 2004) |
| **20** | TCCC-motif | Cis-acting regulatory element that is responsive to light. | (Geffers et al., 2000) |
| **21** | O2-site | Zein metabolism Cis-acting element by which the seed storage protein is regulated. | (Von Harsdorf et al., 1997) |
| **22** | MBS | MYB binding site implicated in inducibility to drought. | (Bowen and Hassan, 1993) |
| **23** | Myb | MYB transcription factor binding site on stress and developmental regulation. | (Lee and Kaplan, 1992) |
| **24** | Myb-binding site | MYB-dependent transcription regulation Cis-acting regulatory element. | (Ter Linde and Steensma, 2003) |
| **25** | TC-rich repeats | Cis-regulatory component of defensive and stress responsiveness. | (Guillermo et al., 2017) |

**Subcellular localization prediction**

The subcellular localization of *TcYABBY* proteins was predicted to gain insight into their potential functional roles within the cell (Figure 9). The analysis revealed that *TcYABBY* proteins are predominantly localized in the nucleus, consistent with their role as transcription factors. Strong nuclear localization signals were observed particularly for *TcYABBY2*, *TcYABBY7*, *TcYABBY1*, and *TcYABBY5*, as indicated by higher scores in the nuclear compartment. In addition to nuclear localization, several *TcYABBY* proteins also showed predicted localization in cytoplasmic and plastid compartments, suggesting potential multifunctional roles. *TcYABBY3*, *TcYABBY4*, and *TcYABBY6* exhibited relatively broader localization patterns, with detectable signals in the cytoplasm, plastids, vacuole, and mitochondria, indicating a more diverse subcellular distribution. Minor localization signals were also observed in the endoplasmic reticulum and chloroplast compartments, though these were comparatively weaker(Dönnes and Höglund, 2004). Hierarchical clustering of localization patterns grouped *TcYABBY* proteins with similar subcellular distributions together, reflecting functional similarity among closely related members. Overall, the results suggest that while nuclear localization is dominant for *TcYABBY* proteins, certain members may also function in other cellular compartments, highlighting potential diversification of regulatory roles within the *TcYABBY* gene family.

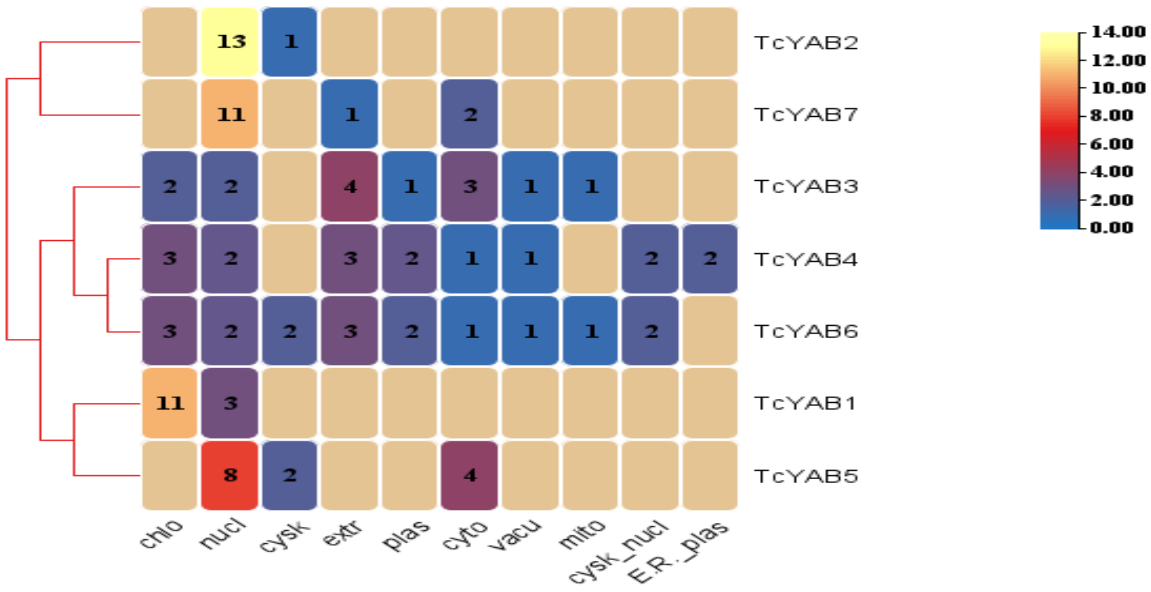


**Figure 9 Predicted subcellular localization of TcYABBY proteins. Heatmap visualization shows the likelihood of protein localizations in different cellular compartments such as the nucleus, cytoplasm, plastids, vacuole, mitochondria, and endoplasmic reticulum**

**Functional enrichment analysis**

Gene Ontology (GO) enrichment analysis was carried out to investigate the potential functional roles of *TcYABBY* genes in *Theobroma cacao*. It was found, through the results, that *TcYABBY* genes play a major role in processes biologically related to their functioning, including nucleotide metabolism, adenosine/AMP deaminase activity, purine nucleoside phosphorylase activity, and general nucleotide metabolic processes, showing the highest fold enrichment values. In addition, enrichment was observed for pathways associated with pyrimidine metabolism and purine biosynthesis, indicating the involvement of *TcYABBY* genes in nucleic acid–related metabolic regulation(Wu et al., 2010) (Figure 10). GO terms related to molecular functions, such as metal binding, zinc ion binding, metal-dependent hydrolase activity, and hydrolase activity, were also significantly enriched. These results suggest that *TcYABBY* proteins may participate in enzymatic and regulatory functions requiring metal cofactors. Overall, the GO enrichment analysis indicates that *TcYABBY* genes in *T. cacao* are functionally associated with nucleotide metabolism, metal ion binding, and catalytic activities, highlighting their potential roles in metabolic regulation and cellular processes(Yadav and Srivastava, 2018).

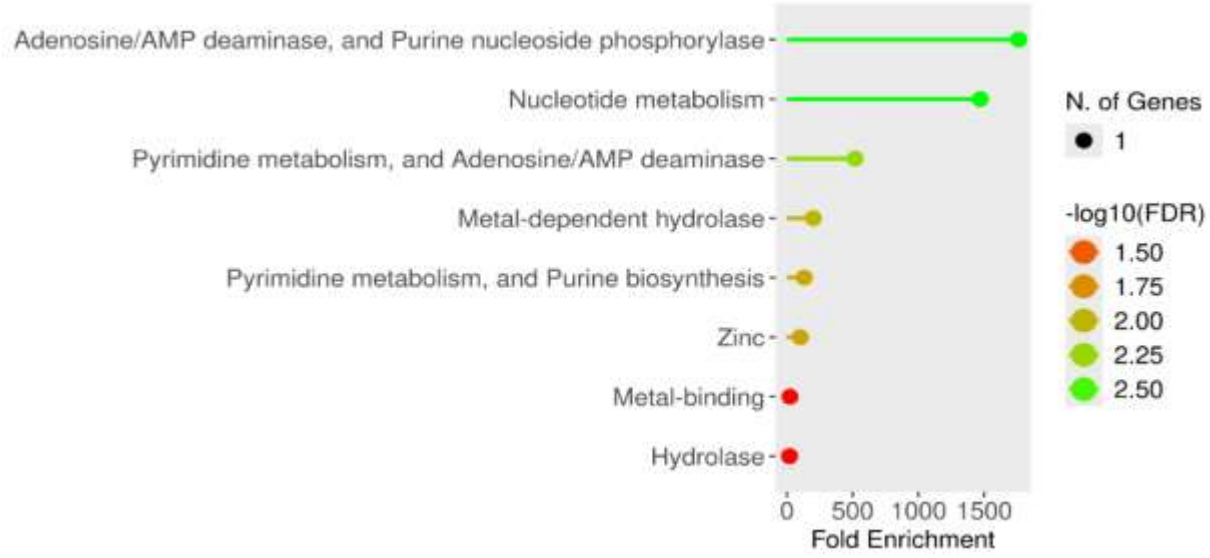


***Figure 10 Gene Ontology (GO) enrichment analysis of TcYABBY genes. The dot plot points to significantly enriched biological processes and molecular functions such as nucleotide metabolism, metal binding, hydrolase activity, and metabolic regulation pathway***

**Protein–protein interaction network**

Protein–protein interaction (PPI) analysis was conducted to explore the functional interaction landscape of *TcYABBY* proteins in *Theobroma cacao* (Figure 11). The PPI network revealed that *TcYABBY* proteins form an interconnected interaction network, indicating their involvement in coordinated regulatory processes. Among the *TcYABBY* members, *TcYABBY7* appeared as a central hub protein, showing interactions with multiple *TcYABBY* proteins as well as several non-*YABBY* interacting partners(Zuo et al., 2009). The network also identified interactions between *TcYABBY5*, *TcYABBY6*, and *TcYABBY7*, suggesting potential cooperative or related functional roles among these proteins. In addition, *TcYABBY* proteins were found to interact with several *TCM*-labeled proteins (e.g., *TCM_017424*, *TCM_024217*, *TCM_033940*, and *TCM_021794*), which are likely associated with regulatory or metabolic functions. The presence of both intra-family (*YABBY–YABBY*) and inter-family interactions highlights the integration of *TcYABBY* proteins into broader cellular networks. Overall, the PPI analysis suggests that *TcYABBY* proteins participate in complex regulatory and metabolic interaction networks, supporting their potential roles in transcriptional regulation and coordination of diverse biological processes in *T. cacao*.

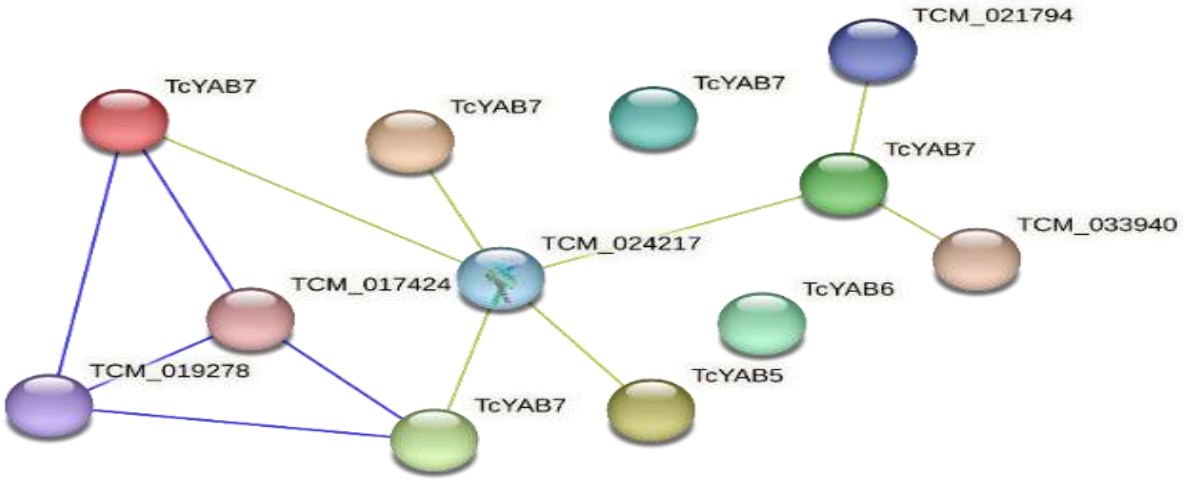


**Figure 11 Protein protein interaction (PPI) network of TcYABBY proteins. Network analysis reveals that there are interactions between TcYABBY proteins and other associated regulatory proteins. TcYABBY7 is identified to play a significant role as a central hub protein, implying its potential major role in transcriptional and metabolic regulatory networks.**

**Discussion**

The *YABBY* gene family represents a class of plant-specific transcription factors that are evolutionarily conserved and critically involved in organ polarity, tissue differentiation, and reproductive development, and the present genome-wide analysis provides the first comprehensive characterization of this gene family in *Theobroma cacao*. The identification of seven *TcYABBY* genes, grouped into five conserved subfamilies consistent with *Arabidopsis thaliana*, reflects strong evolutionary conservation across dicot species while also indicating limited lineage-specific expansion in cacao. The conserved exon–intron organization and shared *YABBY* and *HMG-box* motifs across all *TcYABBY* proteins highlight strong structural constraints, supporting their functional indispensability. Chromosomal localization and synteny analyses revealed that *TcYABBY* genes are unevenly distributed across cacao chromosomes and that segmental duplication, rather than tandem duplication, has been the primary driver of gene family expansion, a pattern typical of transcription factor families maintaining dosage balance. The predominance of purifying selection acting on duplicated *TcYABBY* gene pairs, as indicated by Ka/Ks ratios below unity, further underscores the evolutionary pressure to preserve functional integrity following duplication events. Promoter analysis uncovered a rich diversity of cis-regulatory elements related to light responsiveness, phytohormone signaling, stress responses, and tissue-specific regulation, suggesting that *TcYABBY* genes act as integrators of developmental and environmental cues(Argout et al., 2011). Subcellular localization predictions confirmed the dominant nuclear localization of *TcYABBY* proteins, consistent with their transcriptional regulatory roles, while secondary localization in other cellular compartments for certain members suggests potential functional diversification. Functional enrichment analysis further linked *TcYABBY* genes to nucleotide metabolism, metal binding, and catalytic activities, expanding their potential biological roles beyond classical morphogenetic functions. Additionally, protein–

protein interaction network analysis identified *TcYABBY7* as a central hub interacting with both *YABBY* and non-*YABBY* proteins, indicating that *TcYABBY* proteins operate within complex regulatory and metabolic networks(Karimizadeh et al., 2019; Kumar et al., 2016). Overall, these findings demonstrate that *TcYABBY* genes are highly conserved yet functionally diversified regulators in cacao, providing valuable insights into their evolutionary dynamics and laying a robust foundation for future functional studies and breeding-oriented applications(Laukens et al., 2015).

**Conclusion**

In conclusion, this research presents a comprehensive genomic properties of the *YABBY* transcription factor gene family in Theobroma cacao and provides important observations into their framework organization, evolutionary background, regulatory complexity, and biological potential. Seven *TcYABBY* genes were characterized and divided into conserved subfamilies consistent with Arabidopsis thaliana, highlighting the strong evolutionary conservation of *YABBY* genes among dicotyledonous plants. Structural and motif analyses demonstrated a high level of conservation, underscoring functional importance of *YABBY* domain, while chromosomal localization, synteny, and duplication analyses demonstrated that large scale duplication and purifying selection have played key roles in shaping the *TcYABBY* gene family. The quantity of hormone-, light-, and stress-responsive cis-regulatory elements, together with predominant nuclear localization and enriched functional pathways, suggests that *TcYABBY* genes act as integrative regulators of developmental and environmental signaling. Furthermore, PPI analyses revealed that *TcYABBY* proteins, particularly *TcYABBY7*, are embedded within complex regulatory networks, supporting their central role in transcriptional control and metabolic coordination. Overall, these findings advance our awareness of *YABBY* gene evolution and function in cacao and provide a valuable genomic resource for future functional validation, molecular breeding, and crop improvement strategies aimed at enhancing cacao development and stress resilience.

**References**

Adak, S., Agarwal, T., Das, P., Ray, S., & Lahiri Majumder, A. (2023). Characterization of myo-inositol oxygenase from rice (OsMIOX): influence of salinity stress in different indica rice cultivars. *Physiology and Molecular Biology of Plants*, **29**(7), 927-945. https://doi.org/10.1007/s12298-023-01340-6

Akbudak, M. A., Yildiz, S., & Filiz, E. (2020). Pathogenesis related protein-1 (PR-1) genes in tomato (Solanum lycopersicum L.): Bioinformatics analyses and expression profiles in response to drought stress. *Genomics*, **112**(6), 4089-4099. https://doi.org/10.1016/j.ygeno.2020.07.004

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M. J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., & Maximova, S. N. (2011). The genome of Theobroma cacao. *Nature genetics*, **43**(2), 101-108. https://doi.org/https://doi.org/10.1038/ng.736

Back, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. Proceedings of the first IEEE conference on evolutionary computation. IEEE World Congress on Computational Intelligence,

Bowen, S., & Hassan, H. (1993). Characterization of cis-acting regulatory mutations causing anaerobic expression of the sodA gene in Escherichia coli. *Archives of biochemistry and biophysics*, **302**(2), 372-379. https://doi.org/10.1006/abbi.1993.1226

Brown, J. S., Sautter, R. T., Olano, C. T., Borrone, J. W., Kuhn, D. N., Motamayor, J., & Schnell, R. J. (2008). A composite linkage map from three crosses between commercial clones of cacao, Theobroma cacao L. *Tropical Plant Biology*, **1**(2), 120-130. https://doi.org/10.1007/s12042-008-9011-4

Brown, J. W., Daniels, C. J., Reeve, J. N., & Konisky, J. (1989). Gene structure, organization, and expression in archaebacteria. *CRC Critical reviews in microbiology*, *16*(4), 287-337. https://doi.org/10.3109/10408418909105479

Buttar, Z. A., Yang, Y., Sharif, R., Nan Wu, S., Xie, Y., & Wang, C. (2020). Genome wide identification, characterization, and expression analysis of YABBY-gene family in wheat (Triticum aestivum L.) *Agronomy*, **10**(8), 1189. https://doi.org/10.3390/agronomy10081189

Chen, M., Liang, X., Yang, Y., Wang, Y., Wei, J., Duan, Y., Peng, H., Duan, Y., Huang, Y., & Zou, W. (2025). MpYABBY2 promotes ABA accumulation and influences MpMYB11/MpbHLH79 complex to regulate anthocyanin biosynthesis in Malus 'Pinkspire'fruits under cold stress. *Postharvest Biology and Technology*, **222**, 113341. https://doi.org/10.1016/j.postharvbio.2024.113341

Chen, Y., Jin, G., Lu, J., Lou, Y., Jiménez-Alemán, G. H., & Li, R. (2025). Wound-induced transcriptional dynamics in rice. *Crop Health*, **3**(1), 15. https://doi.org/10.1007/s44297-025-00055-2

Dönnes, P., & Höglund, A. (2004). Predicting protein subcellular localization: past, present, and future. *Genomics, proteomics & bioinformatics*, **2**(4), 209-215. https://doi.org/10.1016/S1672-0229(04)02027-3

Eshed, Y., Izhaki, A., Baum, S. F., Floyd, S. K., & Bowman, J. L. (2004). Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and

YABBY activities. https://doi.org/10.1242/dev.01186

Farrera-Sal, M., Fillat, C., & Alemany, R. (2020). Effect of transgene location, transcriptional control elements and transgene features in armed oncolytic adenoviruses. *Cancers*, **12**(4), 1034. https://doi.org/10.3390/cancers12041034

Fitzpatrick, C. R., Gehant, L., Kotanen, P. M., & Johnson, M. T. (2017). Phylogenetic relatedness, phenotypic similarity and plant–soil feedbacks. *Journal of Ecology*, *105*(3), 786-800. https://doi.org/10.1111/1365-2745.12709

Fowler, D. M., Araya, C. L., Gerard, W., & Fields, S. (2011). Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics*, **27**(24), 3430-3431. https://doi.org/10.1093/bioinformatics/btr577

Gain, H., De, S., & Banerjee, J. (2024). Expressional vagaries of OsCAMTA genes under differential abiotic stresses supported with protein–protein interaction study and prediction of miRNA target sites. *Plant Biotechnology Reports*, **18**(5), 637-658. https://doi.org/10.1007/s11816-024-00899-0

Gao, X., Zhang, Y., He, Z., & Fu, X. (2017). 1Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China; 2Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. http://dx.doi.org/10.1016/B978-0-12-811562-6.00004-9

Geffers, R., Cerff, R., & Hehl, R. (2000). Anaerobiosis-specific interaction of tobacco nuclear factors with cis-regulatory sequences in the maize GapC4 promoter. *Plant Molecular Biology*, **43**(1), 11-21. https://doi.org/10.1023/A:1006419232075

Guillermo, B.-V., Marisela, R.-D., Rosalba, T.-R., Reginaldo, B.-S., & Martín-Ernesto, T.-H. (2017). Physiological function of rhamnogalacturonan lyase genes based in the analysis of cis-acting elements located in the promoter region. *Research Journal of Biotechnology Vol*, **12**, 6. https://doi.org/10.1016/j.jplph.2018.08.002

Gushchina, L. V., Kwiatkowski, T. A., Bhattacharya, S., & Weisleder, N. L. (2018). Conserved structural and functional aspects of the tripartite motif gene family point towards therapeutic applications in multiple diseases. *Pharmacology & therapeutics*, **185**, 12-25. https://doi.org/10.1016/j.pharmthera.2017.10.020

Han, K., Lai, M., Zhao, T., Yang, X., An, X., & Chen, Z. (2025). Plant YABBY transcription factors: a review of gene expression, biological functions, and prospects. *Critical Reviews in Biotechnology*, *45*(1), 214-235. https://doi.org/10.1080/07388551.2024.2344576

Hao, L., Zhang, J., Shi, S., Li, P., Li, D., Zhang, T., & Guo, H. (2022). Identification and expression profiles of the YABBY transcription factors in wheat. *PeerJ*, **10**, e12855. https://doi.org/10.7717/peerj.12855

Harshith, C. Y., Pal, A., Chakraborty, M., Nair, A., Raju, S., & Shivaprasad, P. V. (2024). Wound-induced small-peptide-mediated signaling cascade, regulated by OsPSKR, dictates balance between growth and defense in rice. *Cell reports*, **43**(7). https://doi.org/10.1016/j.celrep.2024.114515

Horton, P., Park, K.-J., Obayashi, T., & Nakai, K. (2006). Protein subcellular localization prediction with WoLF PSORT. Proceedings of the 4th Asia-Pacific bioinformatics conference,

Karimizadeh, E., Sharifi-Zarchi, A., Nikaein, H., Salehi, S., Salamatian, B., Elmi, N., Gharibdoost, F., & Mahmoudi, M. (2019). Analysis of gene expression profiles and protein-protein interaction networks in multiple tissues of systemic sclerosis. *BMC medical genomics*, **12**(1), 199. https://doi.org/10.1186/s12920-019-0632-2

Kawakami, H., Ijichi, N., Obama, Y., Matsuda, E., Mitsui, K., Nishikawaji, Y., Watanabe, M., Nagano, S., Taniguchi, N., & Komiya, S. (2024). An optimal promoter regulating cytokine transgene expression is crucial for safe and effective oncolytic virus immunotherapy. *Translational Research*, **273**, 32-45. https://doi.org/10.1016/j.trsl.2024.07.002

Kumar, A., Thotakura, P. L., Tiwary, B. K., & Krishna, R. (2016). Target identification in Fusobacterium nucleatum by subtractive genomics approach and enrichment analysis of host-pathogen protein-protein interactions. *BMC microbiology*, **16**(1), 84. https://doi.org/10.1186/s12866-016-0700-0

Laukens, K., Naulaerts, S., & Berghe, W. V. (2015). Bioinformatics approaches for the functional interpretation of protein lists: from ontology term enrichment to network analysis. *Proteomics*, **15**(5-6), 981-996. https://doi.org/10.1002/pmic.201400296Digital

Lee, J. K., & Kaplan, S. (1992). cis-acting regulatory elements involved in oxygen and light control of puc operon transcription in Rhodobacter sphaeroides. *Journal of bacteriology*, **174**(4), 1146-1157. https://doi.org/10.1128/jb.174.4.1146-1157.1992

Lee, K., & Seo, P. J. (2022). Wound-induced systemic responses and their coordination by electrical signals. *Frontiers in Plant Science*, **13**, 880680. https://doi.org/10.3389/fpls.2022.880680

Lenhard, B., Sandelin, A., & Carninci, P. (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, **13**(4), 233-245. https://doi.org/10.1038/nrg3163

Li, Z., Li, G., Cai, M., Priyadarshani, S. V., Aslam, M., Zhou, Q., Huang, X., Wang, X., Liu, Y., & Qin, Y. (2019). Genome-wide analysis of the YABBY transcription factor family in pineapple and functional identification of AcYABBY4 involvement in salt stress. *International journal of molecular sciences*, **20**(23), 5863. https://doi.org/10.3390/ijms20235863

Liu, T., Sharif, R., Shi, Z., Guo, K., Zhang, Z., Bao, X., & Ali, A. (2025). Transcriptomic analysis reveals the crucial role of YABBY genes family in hormonal induced parthenocarpy in Cucumis sativus L. *BMC Plant Biology*, **25**(1), 45. https://doi.org/10.1186/s12870-024-06018-z

Lowry, C. V., & Zitomer, R. S. (1984). Oxygen regulation of anaerobic and aerobic genes mediated by a common factor in yeast. *Proceedings of the National Academy of Sciences*, **81**(19), 6129-6133. https://doi.org/10.1073/pnas.81.19.6129

Mariyam, Shafiq, M., Haseeb, M., Atif, R. M., Naqvi, S. A. A. A. A., Ali, N., Javed, M. A., Gillani, F., & Haider, M. S. (2021). Genome-wide identification and characterization of a plant-specific Dof transcription factor gene family in olive (Olea europaea) and its comparison with Arabidopsis. *Horticulture, Environment, and Biotechnology*, **62**(6), 949-968. https://doi.org/10.1007/s13580-021-00366-7

Nadeau, R., Byvsheva, A., & Lavallée-Adam, M. (2021). PIGNON: a protein–protein interaction-guided functional enrichment analysis for quantitative proteomics. *BMC bioinformatics*, **22**(1), 302. https://doi.org/10.1186/s12859-021-04042-6

Nettelbeck, D. M., Rivera, A. A., Balagué, C., Alemany, R., & Curiel, D. T. (2002). Novel oncolytic adenoviruses targeted to melanoma: specific viral replication and cytolysis by expression of E1A mutants from the tyrosinase enhancer/promoter. *Cancer research*, **62**(16), 4663-4670.

Obe, G., Pfeiffer, P., Savage, J., Johannes, C., Goedecke, W., Jeppesen, P., Natarajan, A., Martınez-López, W., Folle, G., & Drets, M. (2002). Chromosomal aberrations: formation, identification and distribution. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **504**(1-2), 17-36. https://doi.org/10.1016/S0027-5107(02)00076-3

Papadakis, E., Nicklin, S., Baker, A., & White, S. (2004). Promoters and control elements: designing expression cassettes for gene therapy. *Current gene therapy*, **4**(1), 89-113. https://doi.org/10.2174/1566523044578077

Pathan, M., Keerthikumar, S., Ang, C. S., Gangoda, L., Quek, C. Y., Williamson, N. A., Mouradov, D., Sieber, O. M., Simpson, R. J., & Salim, A. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*, **15**(15), 2597-2601. https://doi.org/10.1002/pmic.201400515

Prewitt, S., Shalit-Kaneh, A., Maximova, S., & Guiltinan, M. (2021). Inter-species functional compatibility of the Theobroma cacao and Arabidopsis FT orthologs: 90 million years of functional conservation of meristem identity genes. *BMC Plant Biology*, **21**(1), 218. https://doi.org/10.1186/s12870-021-02982-y

Ren, X., Chen, S., Wang, X., Wang, L., Qihong, D., Ma, Y., Wang, Y., Zhang, B., Han, Y., & Qin, Y. (2025). Identification of YABBY transcription factors in foxtail millet and functional characterization of SiFILs under GA and drought conditions. *International Journal of Biological Macromolecules*, 148653. https://doi.org/10.1016/j.ijbiomac.2025.148653

Roy, S. (2016). Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant signaling & behavior*, **11**(1), e1117723. https://doi.org/10.1080/15592324.2015.1117723

Saukkonen, K., & Hemminki, A. (2004). Tissue-specific promoters for cancer gene therapy. *Expert opinion on biological therapy*, **4**(5), 683-696. https://doi.org/10.1517/14712598.4.5.683

Sen, S. (2022). Genome-wide TIFY family in Arachis hypogaea in the perspective of legume JAZs. *Journal of Crop Science and Biotechnology*, **25**(4), 465-488. https://doi.org/10.1007/s12892-022-00145-5

Shrestha, K., & Huang, Y. (2022). Genome-wide characterization of the sorghum JAZ gene family and their responses to phytohormone treatments and aphid infestation. *Scientific reports*, **12**(1), 3238. https://doi.org/10.1038/s41598-022-07181-9

Smale, S., Jain, A., Kaufmann, J., Emami, K., Lo, K., & Garraway, I. (1998). The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harbor symposia on quantitative biology,

Stirnberg, M., & Happe, T. (2004). Identification of a cis-acting element controlling anaerobic expression of the HydA gene from Chlamydomonas reinhardtii. In *Biohydrogen III* (pp. 117-127). Elsevier.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, **38**(7), 3022-3027.

Tang, L., Luo, S., Wang, D., Liu, X., Wang, X., Tan, M., He, B., Zhao, Z., & Huang, M. (2025). Identification and characterization of pathogen recognition genes families in Gastrodia elata: Insights into their roles in growth and disease

resistance. https://doi.org/10.21203/rs.3.rs-7424538/v1

Ter Linde, J., & Steensma, H. (2003). Transcriptional regulation of YML083c under aerobic and anaerobic conditions. *Yeast*, **20**(5), 439-454. https://doi.org/10.1002/yea.975

Von Harsdorf, R., Edwards, J. G., Shen, Y.-T., Kudej, R. K., Dietz, R., Leinwand, L. A., Nadal-Ginard, B., & Vatner, S. F. (1997). Identification of a cis-acting regulatory element conferring inducibility of the atrial natriuretic factor gene in acute pressure overload. *The Journal of clinical investigation*, **100**(5), 1294-1304. https://doi.org/10.1172/JCI119643

Wang, X., Niu, Y., & Zheng, Y. (2021). Multiple functions of MYB transcription factors in abiotic stress responses. *International journal of molecular sciences*, **22**(11), 6125. https://doi.org/10.3390/ijms22116125

Wang, Z., Wei, G.-H., Liu, D.-P., & Liang, C.-C. (2007). Unravelling the world of cis-regulatory elements. *Medical & biological engineering & computing*, **45**(8), 709-718. https://doi.org/10.1007/s11517-007-0195-9

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, **13**(1), 59-69. https://doi.org/10.1038/nrg3095

Wu, G., Feng, X., & Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. *Genome biology*, **11**(5), R53. https://doi.org/10.1186/gb-2010-11-5-r53

Xi, R., Liu, H., Chen, Y., Zhuang, H., Han, H., Wang, H., Wang, Q., & Li, N. (2023). Genome-wide characterization of tomato FAD gene family and expression analysis under abiotic stresses. *Plants*, **12**(22), 3818. https://doi.org/10.3390/plants12223818

Xiao-Li, S., Li, Y., Hua, C., Xi, B., Wei, J., Zuo-Jun, J., & Yan-Ming, Z. (2011). Arabidopsis bZIP1 transcription factor binding to ABRE cis-element regulates abscisic acid signal transduction. *Acta Agronomica Sinica*, **37**(4), 612-619. https://doi.org/10.1016/j.tplants.2004.12.012

Yadav, R., & Srivastava, P. (2018). Clustering, pathway enrichment, and protein-protein interaction analysis of gene expression in neurodevelopmental disorders. *Advances in Pharmacological and Pharmaceutical Sciences*, **2018**(1), 3632159. https://doi.org/10.1155/2018/3632159

Yamaguchi-Shinozaki, K., & Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. *Trends in plant science*, **10**(2), 88-94. https://doi.org/10.1016/j.tplants.2004.12.012

Ye, L., Cao, L., Zhao, X., Guo, X., Ye, K., Jiao, S., Wang, Y., He, X., Dong, C., & Hu, B. (2022). Investigation of the JASMONATE ZIM-DOMAIN gene family reveals the canonical JA-signaling pathway in Pineapple. *Biology*, **11**(3), 445.

Yilmaz, A. A., Coşkun, S., Aras, S., & Büyük, İ. (2025). Integrative multi-omics analysis of the FAR1 gene family in Phaseolus vulgaris. *Communications Faculty of Sciences University of Ankara Series C Biology*, **34**(2), 154-175. https://doi.org/10.53447/communc.1657858

Zhang, T., Li, C., Li, D., Liu, Y., & Yang, X. (2020). Roles of YABBY transcription factors in the modulation of morphogenesis, development, and phytohormone and stress responses in plants. *Journal of Plant Research*, **133**(6), 751-763. https://doi.org/10.1007/s10265-020-01227-7

Zhao, S., Zhang, Y., Tan, M., Jiao, J., Zhang, C., Wu, P., Feng, K., & Li, L. (2023). Identification of YABBY transcription factors and their function in ABA and salinity response in Nelumbo nucifera. *Plants*, **12**(2), 380. https://doi.org/10.3390/plants12020380

Zuo, C., Liang, S., Wang, Z., Li, H., & Ma, W. (2009). Enriching protein-protein and functional interaction networks in human embryonic stem cells. *International journal of molecular medicine*, **23**(6), 811-819. https://doi.org/10.3892/ijmm_00000197

**Statements and Declarations**

**Funding**

The authors declare that they have not received any funds, grants, or other type of financial support for the preparation of this study.

**Competing Interests**

The authors declare that they have no financial or none, financial competing interests related to this work.

**Author Contributions**

Conceptualization and study design were performed by Qaisar Hayyat and Muhammad Saood Amjad. Data collection, bioinformatics analysis, and interpretation were carried out by Muhammed Asadullah Zohaib. The writing of the first manuscript draft was done by Qaisar Hayyat has checked and approved the final version of the manuscript.

**Ethics Approval**

Since this research does not involve human subjects, animals, or any biological material that would require both the use of and the ethical approval, the authors confirm that no such approval was necessary.

**Consent to Participate**

Not applicable.

**Consent to Publish**

Not applicable.

**Data Availability**

All data generated or analyzed during this study are included in this published article.

**Declaration of Generative AI and AI-assisted Technologies in the Writing Process**

During the preparation of this work, the author(s) have not used any ChatGPT (OpenAI) tool.