Genome-wide Analysis of Plant Specific YABBY Transcription Factor Gene Family in Watermelon (Citrullus lanatus) and Arabidopsis

Mehr-ul-Nisa\textsuperscript{1}, Muhammad Shafiq\textsuperscript{1}, Mujahid Manzoor*\textsuperscript{2}, Muhammad Bilal\textsuperscript{3}, Tariq Manzoor\textsuperscript{4}, Muhammad Ali\textsuperscript{2}, Malik Muazzam Anees\textsuperscript{1}, Muhammad Rizwan\textsuperscript{4}, Muhammad Zeeshan Haider\textsuperscript{5}, Adnan Sami\textsuperscript{2}, Muhammad Saleem Haider\textsuperscript{6}

\textsuperscript{1}Department of Horticulture, Faculty of Agriculture sciences, University of the Punjab, Lahore, Pakistan
\textsuperscript{2}Department of Entomology, Faculty of Agriculture sciences, University of the Punjab, Lahore, Pakistan
\textsuperscript{3}Department of Agronomy, Faculty of Agriculture sciences, University of the Punjab, Lahore, Pakistan
\textsuperscript{4}Department of Food Science and Technology, Faculty of Agriculture sciences, University of the Punjab, Lahore, Pakistan
\textsuperscript{5}Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab, P.O BOX. 54590, Lahore, Pakistan
\textsuperscript{6}Department of Plant Pathology, Faculty of Agricultural Sciences, University of the Punjab, P.O BOX. 54590, Lahore, Pakistan

*Corresponding Authors Email; mujahid.iags@pu.edu.pk

Abstract

The YABBY gene family is a specific transcription factor for plants and a DNA binding domain that carries out several different functions, such as regulating the length of blooming plant styles and the polarity of lateral organ development. The YABBY gene family members were identified in the watermelon (Citrullus lanatus subsp. vulgaris var. 97103 V1) genome using a set of bioinformatics techniques. Protein motifs, protein architectures, protein sequences, miRNA targets, and tissue-specific expression patterns were all examined. All chromosomes had an uneven distribution of about eight putative YABBY genes. Inner No OuterINO, CRC (Crabs Claw), YAB2, YAB3/AFO, and YAB5 were the five subgroups that the YABBY proteins in watermelon fall within, in accordance with the accepted Arabidopsis categorization which is based on International Standards of Nomenclature. Segmental duplication was more frequent than tandem duplication, and it was predominantly responsible for the growth of the YABBY gene family in watermelon. The results of tissue-specific expression profiling of CIYABBY genes showed that the vast majority of these genes were substantially expressed in roots and seedlings. In this study, cis-regulatory element (CRE) analyses were employed to identify elements in seedlings and roots that are highly responsive to light, wound, drought, auxin, stress, salicylic acid, and abscisic acid (ABA). The findings reveal specific CREs within the promoter regions of genes associated with these responses. Five groups or sub-families have also been identified by comparing the YABBY genes in watermelon and Arabidopsis, however the CRC and YAB2 groups do not share gene pairing among the other groups. This research contributes to a deeper understanding of plant adaptability and stress response mechanisms, with implications for agriculture and plant science.

Keywords: Watermelon, Gene family, Genomic analysis, Specific plant transcription factors, YABBY.

Introduction

Being a sessile organism (which cannot move) by nature, plants are constantly subjected to a variety of external stresses, which impair the biochemical and functional processes occurring within them (Sharif \textit{et al.}, 2020). Several plant transcription factors (TFs) cause the corresponding gene expression in response to such biotic and abiotic stresses (Li \textit{et al.}, 2019; Zhao \textit{et al.}, 2017). A tiny gene family known as the YABBY TFs is unique to seed plants (Yin \textit{et al.}, 2022), typically classified as a subfamily of the zinc-finger super family (Y.-Y. Chen \textit{et al.}, 2020). A collection of TFs with two conserved domains were encoded by the YABBY protein. The N-terminal (Cys2 Cys2) zinc-finger domain and the C-terminal (helix-loop-helix) domain are the two conserved domains (Song, Joshi, DiPiazza, & Joshi, 2020). Both domains retain highly sealed amino acid (AA) residues and participate in the particular binding of DNA (Sawa \textit{et al.}, 1999). The growth and development of lateral organs (Bowman, Smyth, & Meyerowitz,
establishment of adaxial-abaxial polarity (Kumarun, Bowman, & Sundaresan, 2002), expansion of leaves (Eckardt, 2010), development of leaf edges (Finet et al., 2016), and the response to stress are significantly influenced by YABBY TFs (Zhao et al., 2017). The YABBY family was divided into five subfamilies by phylogenetic analysis: the CRABS CLAW (CRC), FILAMENTOUS FLOWER (FIL)/YABBY3 (YAB3), INNER NO OUTER (INO), YABBY2 (YAB2), and YABBY5 (YAB5) (Bowman, 2000; Yamada, Ito, & Kato, 2004). Moreover, CRC helps to create the polarity of developing carpels and nectarines, while FIL, YAB3, YAB2, and YAB5 excessively encourage the development of lateral organs (Siegfried et al., 1999). The remaining INO has a variety of functions that help the ovule’s outer integument, a layer of cells that surrounds the nucleus, grow into the seed coat (Villanueva et al., 1999). Based on their crucial developmental roles, YABBY genes have undergone extensive genomic analysis in plants since it’s crucial to understand how they relate to one another among species or even within a single species. 6 YABBY genes have been found in Arabidopsis thaliana, 9 in Pineapple (Li et al., 2019), 7 in Grapevine (Zhang et al., 2019), 11 CmoyABBYs in Cucurbita moschata, 12 CmaYABBYs in Cucurbita maxima, 11 CpeYABBYs in Cucurbita pepo (Yuan et al., 2020), 12 in Brassica rapa ssp. Chinensis (Hou, Wu, Gao, Zhang, & Hou, 2019), 21 in Triticum aestivum (wheat) (Buttar et al., 2020), 9 in Tuberosum lycopersicum (Tomato) (Huang, Van Houten, Gonzalez, Xiao, & van der Knaap, 2013), 8 in Oryza sativa (Rice) (Toriba et al., 2007), 12 in Gossypium arboreum, 12 G. raimondii and 23 in G. hirsutum (Upland Cotton) (Zhaoen Yang et al., 2018). The Cucurbitaceae family includes the watermelon, Citrullus lanatus subsp. vulgaris var. 97103 V1. It originates from tropical regions of Africa close to the Kalahari Desert (Kyriacou, Leskovar, Colla, & Rouphael, 2018; Naz, Butt, Sultan, Qayyum, & Niaz, 2014). Botanists typically refer to it as a “pepo,” a fruit with a thick skin and fleshy core (Mehra, Pasricha, & Gupta, 2015). Because of its refreshing texture, appealing colour, delicate and sweet flavor, and high water content to alleviate summer thirst, it is typically consumed as a pleasant summer fruit that is well cherished by customers (Romdhane et al., 2017). Watermelon fruits produce 55.3% juice, 31.5% rind, and 10.4% pomace (Maoto, Beswa, & Jideani, 2019; Oberoi & Sogi, 2017). Using various bioinformatics tools, the primary goal of this study was to locate and characterize the genes belonging to the YABBY TFs family in the Watermelon genome. To put it briefly, YABBY genes from the watermelon genome were identified using methodical way i.e. Phylogenetic Analysis, Synteny Analysis and Conserved Motif Analysis etc. Investigations were also conducted into their chromosomal distribution, intron/exon distribution pattern, presence of conserved domains, and cis-regulatory elements. The wide-ranging genome assessment of YABBY genes presents an orientation for cloning and functioning properties in watermelon. Investigations revealed genome-wide distribution and the role of YABBY genes in watermelon with the best-fit comparison to Arabidopsis thaliana.

Materials and methods

Database search and sequence retrieval: By using ID PF04690 (http://pfam.xfam.org/) the AA sequence for the YABBY PSTRFs (Plant specific Transcription factors) was obtained from the peptide genome of Arabidopsis thaliana (Accession No. A0A1P8APE2) and contains 164 amino acids (Mistry et al., 2021). The Cucurbit Genomics Database (CuGenDB) (http://cucurbitgenomics.org/) was used to Blast-p (Protein against Protein search tool) this sequence against the watermelon genome. The obtained amino acid sequences were investigated using the default parameters of the NCBI CDD (Conserved Domain Database) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Lu et al., 2020) and the simple modular architecture research tool (SMART) available at (http://smart.embl-heidelberg.de) (Letunic & Bork, 2018). Any identified proteins that did not include the YABBY conserved domain were ruled out.

Investigation of physio-chemical characteristics of CIYABBY proteins: Through using ProtParam tool (http://web.expasy.org/protparam/), the length (AA residues), molecular weight (MW), and theoretical pl of CIYABBY proteins were estimated (Gasteiger et al., 2005). The (CuGenDB) was utilized to obtain information on gene IDs, chromosomal locations, and protein and gene sequences (http://cucurbitgenomics.org/). According to the order of their physical placements, these CIYABBY genes were given new names such as Cl initials of Scientific Name of Watermelon and Yabby As a transcription factor WoLF PSORT (https://wolfsort.hgc.jp/), an online tool, was used to estimate CIYABBY’s subcellular localization (Horton, Park, Obayashi, & Nakai, 2006).

Gene structure analysis: In order to investigate the intron/exon organization of CIYABBY, the genomic and coding sequences of identified genes were collected from the Cucurbit Genomics Database (CuGenDB) (http://cucurbitgenomics.org/). Additionally, the watermelon genome’s g3 file was acquired from the Cucurbit Genomics Database (CuGenDB). With the use of the Gene Structure Display Server (GSDS v2.0) (Hu et al., 2015) (located at http://gsds.cbi.pku.edu.cn/), these sequences were then used to depict the gene structure.

Multiple sequence alignment and phylogenetic analysis: The YABBY protein’s amino acid sequences were aligned with Clustal W version 2.1 (Thompson, Higgins, & Gibson, 1994), and the phylogeny was built.
using MEGA X v2.0 (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) with neighbour-joining (NJ) and bootstrapping set at 1000 replications with partial deletion. Eight watermelons, thirteen musk melons, seven potatoes, and six Arabidopsis YABBY protein sequences were employed for the phylogenetic study. **CRE and conserved motif analysis:** For each putative CYABBY gene, a sequence 1000 bp upstream of the initiation codon was obtained for the investigation of the promoter regions. Following this, CREs in these sequences were predicted using the PlantCare database. The predicted protein sequences of the CYABBY were used to assess motifs using Multiple EM for Motif Elicitation (MEME) with a maximum number of motifs set at 20 (Bailey, Johnson, Grant, & Noble, 2015). Along with other variables, the default values for motif widths were set to 6 and 50, respectively.

**Gene duplication and synteny analysis of Watermelon (Cucurbita Family):** Using Ks and Ka values, the watermelon YABBY gene family's timing of divergence was calculated. TB Ttools was used to determine the Ka and Ks substitution rates and computed Ka/Ks ratios according to the instructions in the software package manuals. For each pair of paralogous genes, the rates of molecular evolution were determined using the Ka/Ks ratios. T=Ks/2 was used to calculate the time of divergence (T), where λ is equal to 1.5 × 10−8 (P. Wang et al., 2019). With the default settings, the Multiple Collinearity Scan toolkit (MCScanX) was employed to investigate the gene duplication occurrences (Y. Wang, Li, & Paterson, 2013). The syntenic analysis map was created using TbTools' Micro Synteny view software to show the syntenic association of the paralogous YABBY genes isolated from the watermelon (C. Chen et al., 2020). By Using Tb Tools dual syntenic comparison between watermelon and melon, watermelon and musa, watermelon and Arabidopsis, watermelon and bottle gourd was found out.

**Transcriptome analysis:** We gathered previously generated RNA sequence data using high through put sequencing for watermelon plant tissues, including leaves and stems at seedling stage under semi-controlled environment and open field, to examine the organ-specific expression profile of CYABBY at various development stages (Song, Joshi, DiPiazza, & Joshi, 2020). The Reads Per Kilobases per Million mapped reads (RPKM) values from RNA-seq data were log2 converted for expression profiling. Heatmap Illustrator in TBtools displays expression patterns with hierarchical clustering (Chen et al., 2020).

**Putative microRNA and target site analysis:** The CDS sequences of all watermelon CYABBY genes were searched for sequences complementary to miRNAs using psRNATarget (http://plantgrn.noble.org/psRNATarget/analysis?function=3/) (Samad, 2017) using default parameters in order to find miRNAs that potentially target the watermelon CYABBY genes. Later, the NCBI website was searched for these putative microRNAs' functions, and references were noted.

**Results**

**Identification of the YABBY genes in watermelon:** The sequence of the YABBY domain was BLAST searched against the watermelon's entire genome sequence that was received from the (CuGenDB) in order to discover the YABBY genes. There were 27 Proteins found during an initial study. Proteins with a truncated YABBY DNA-binding domain and those produced by the same gene isoform were excluded from the study. Eight distinct CYABBY genes were discovered and further investigated. The highly conserved YABBY domain was present in these non-redundant YABBY protein sequences from watermelons. Eight of the 27 amino acids reported to be present in the highly conserved watermelon YABBY domain sequences were found to be 100% conserved in all YABBY domain sequences (Fig. S1), while the remaining 21 amino acids were found to be varied in all CYABBY proteins. The CYABBY genes code proteins with a MW range of 18.74 to 26.91 kDa and a length of 169 to 242 AA, with CYABBY8 being the shortest and CYABBY1 is the longest protein (Table 1). The discovered proteins have Ip ranging from 7.70 to 9.14.

**Gene structure and recognition of conserved motifs and domains:** The structure of exons and introns serves as the framework for genes and facilitates the study of evolutionary links among genes or species (Koralewski & Krutovsky, 2011). A gene family can be identified by their numbers and dispersion patterns due to evolution. Phylogenetic analysis and a thorough demonstration of the exon-intron architectures of the watermelon YABBY genes showed that the gene structural pattern coincided with the phylogenetic analysis. In watermelon, there were four to six different introns (Fig. S2; Table 1). Four genes, namely CYABBY3, CYABBY5, CYABBY6, and CYABBY7, each include five introns, whereas only two genes, CYABBY8, and CYABBY4, each contain six introns. Only one gene, CYABBY2, contains four introns (Table 1; Fig. S2). While the CYABBY genes in subfamily YAB5 have six introns, all of the CYABBY genes in subfamily INO have five introns. Some of the watermelon's YABBY genes have four introns, similar to the YABBY genes examined in many species, but others have up to six introns (Table 1; Fig. S2). The MEME tool was used to determine the locations of 20 motifs across all of the watermelon YABBY proteins (Fig. 2). All of the CYABBY proteins have the YABBY domain, which was
present in all of them. The observation that the YABBY genes in the same group encode similar motifs leads to the conclusion that these conserved motifs play a crucial role in the activities that are unique to a group or subgroup. With the exception of subgroup YAB5, CIYABBY8 has certain motifs that are distinct from CIYABBY4. The identical YABBY domain is encoded by all 8 watermelon YABBY genes. Similar pattern distribution among different YABBY genes shows that these genes might have evolved as a result of gene expansion.

Table 1: Information about discovered 8 YABBY genes in watermelon genome. Accession Number, Chromosome number and location, gene direction, Amino acid sequence length, Molecular weight, Isoelectric point (Pi-value), No. of Introns and Exons.

<table>
<thead>
<tr>
<th>YABBY gene</th>
<th>Accession Number</th>
<th>Chromosome Number</th>
<th>Chromosome Location (bp)</th>
<th>Direction</th>
<th>NO. of Amino Acids</th>
<th>Molecular Weight (kD)</th>
<th>Pi-Value</th>
<th>No. of Introns</th>
<th>No. of Exons</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIYABBY 1</td>
<td>Cla01155</td>
<td>WCG</td>
<td>1</td>
<td>37680</td>
<td>R</td>
<td>242</td>
<td>26.91</td>
<td>8.73</td>
<td>6</td>
</tr>
<tr>
<td>CIYABBY 2</td>
<td>Cla00567</td>
<td>7</td>
<td>10</td>
<td>35219</td>
<td>R</td>
<td>188</td>
<td>20.77</td>
<td>8.75</td>
<td>4</td>
</tr>
<tr>
<td>CIYABBY 3</td>
<td>Cla00861</td>
<td>0</td>
<td>2</td>
<td>32629</td>
<td>R</td>
<td>170</td>
<td>18.74</td>
<td>9.08</td>
<td>5</td>
</tr>
<tr>
<td>CIYABBY 4</td>
<td>Cla01670</td>
<td>0</td>
<td>11</td>
<td>23955</td>
<td>F</td>
<td>192</td>
<td>21.62</td>
<td>8.82</td>
<td>6</td>
</tr>
<tr>
<td>CIYABBY 5</td>
<td>Cla00080</td>
<td>7</td>
<td>0</td>
<td>18956</td>
<td>R</td>
<td>186</td>
<td>20.70</td>
<td>7.70</td>
<td>5</td>
</tr>
<tr>
<td>CIYABBY 6</td>
<td>Cla01007</td>
<td>3</td>
<td>3</td>
<td>32287</td>
<td>R</td>
<td>169</td>
<td>19.07</td>
<td>8.41</td>
<td>5</td>
</tr>
<tr>
<td>CIYABBY 7</td>
<td>Cla01023</td>
<td>1</td>
<td>5</td>
<td>31061</td>
<td>F</td>
<td>188</td>
<td>20.90</td>
<td>9.14</td>
<td>5</td>
</tr>
<tr>
<td>CIYABBY 8</td>
<td>Cla01222</td>
<td>6</td>
<td>6</td>
<td>20045</td>
<td>R</td>
<td>169</td>
<td>18.82</td>
<td>8.24</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 2. The distribution of 15 motifs on 8 YABBY proteins of Watermelon by using MEME version 4.9.0. The bars represent motifs with different color codes for different types of motifs.
Comparative phylogenetic relatedness of watermelon gene family with Arabidopsis: (Table 2; Fig. 1). The Eight ClYABBY proteins were found to be spread across five groupings, according to the results of phylogenetic study as AtINO, AtCRC, AtYAB5, AtAFO/AtYAB3, and AtYAB2 (Table 2; Fig. 1). Six YABBY proteins, including one from Arabidopsis (AtINO), were found in the group AtINO. The other members of the group were ClYABBY5, ClYABBY6, CmYABBY9, CmYABBY10, and StYABBY2. AtFIL, ClYABBY7, CmYABBY12, CmYABBY11, and CmYABBY7 were all members of the AtCRC group, which consists of five YABBY-like proteins. Seven YABBY proteins were found in the AtYAB5 group, including one from one from the Arabidopsis AtYAB5 plant, two from the watermelon ClYABBY8, ClYABBY4, three from the musk melon CmYABBY4, CmYABBY5, CmYABBY8, and one from the potato StYABBY1. There were nine YABBY-like proteins in the AtAFO, including two from the Arabidopsis AtAFO, AtYAB3, two from watermelon, CmYABBY1, CmYABBY2, three from musk melon, CmYABBY1, CmYABBY2, CmYABBY3 and two from the potato, StYABBY3, StYABBY4. The last group, AtYAB2, contained six YABBY-like proteins, of which one was from the Arabidopsis AtYAB2 plant, one is from the watermelon ClYABBY3, two were from the musk melon CmYABBY6, CmYABBY13 and two were from the potato (StYABBY5, StYABBY7). Individuals belonging to the same clade exhibit the same structure and behavior (Fig 1). Therefore, it was determined that proteins from related clades have comparable functions Table 2.

![Phylogenetic Relationship among YABBY genes of C. lanatus, S. tuberosum, A. thaliana, C. sativus, and C. maxima](image_url)

**Fig 1**: Phylogenetic Relationship among YABBY genes of *C. lanatus*, *S. tuberosum*, *A. thaliana*, *C. sativus*, and *C. maxima* was studied. *C. lanatus* genes are marked with red triangle. The evolutionary history was inferred using the NJ method with 1000 Bootstrap. This analysis involved 34 YABBY genes. Evolutionary analyses were conducted in MEGA 11.

**Table 2**: Gene ontology enrichment analysis of ClYABBY genes their GO functions, Sub-Cellular localization Signal Genes expression, Orthologs in Arabidopsis and their functions.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>GO Function</th>
<th>Sub-Cellular Localization</th>
<th>Gene expression</th>
<th>Ortholog in Arabidopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClYABBY1</td>
<td>DNA-binding transcription factor activity</td>
<td>chromatin assembly or disassembly, regulation of transcription, DNA-templated, stomatal complex</td>
<td>Peroxisomes</td>
<td>YABBY3 Axial regulator</td>
</tr>
</tbody>
</table>

Table 2.
<table>
<thead>
<tr>
<th>miRNA</th>
<th>Targeting Gene</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cla-miR156c</td>
<td>ClYABBY6</td>
<td>Suppress adventitious root development.</td>
<td>(M. Xu et al., 2016a)</td>
</tr>
<tr>
<td>Cla-miR157b</td>
<td>ClYABBY6</td>
<td>Responsible for the temporal expression pattern of most SPL genes.</td>
<td>(M. Xu et al., 2016b)</td>
</tr>
<tr>
<td>Cla-miR159a</td>
<td>ClYABBY6</td>
<td>Mediate strong silencing of GAMYB to enable normal growth.</td>
<td>(M. Xu et al., 2016a)</td>
</tr>
<tr>
<td>Cla-miR162</td>
<td>ClYABBY7</td>
<td>Not Available</td>
<td></td>
</tr>
<tr>
<td>Cla-miRN820</td>
<td>ClYABBY6</td>
<td>Associated with epigenetic modifications</td>
<td>(Nosaka et al., 2013)</td>
</tr>
<tr>
<td>Cla-miRN825</td>
<td>ClYABBY8</td>
<td>Regulate plant defence responses to pathogens</td>
<td>(M. Xu et al., 2016a)</td>
</tr>
<tr>
<td>Cla-miR157a</td>
<td>ClYABBY6</td>
<td>Responsible for the temporal expression pattern of most SPL genes.</td>
<td>(M. Xu et al., 2016b)</td>
</tr>
<tr>
<td>Cla-miR159b</td>
<td>ClYABBY6</td>
<td>Mediate strong silencing of GAMYB to enable normal growth.</td>
<td>(Nosaka et al., 2013)</td>
</tr>
<tr>
<td>Cla-miRN812</td>
<td>ClYABBY1</td>
<td>Not available</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: ClYABBY genes targeting Putative miRNA functions along with their targeted genes.
Location of chromosomes and assessment of gene duplication of watermelon YABBY genes: The distribution of the examined *Citrullus lanatus subsp. vulgaris* var. 97103 YABBY genes on chromosomes showed that different chromosomes included *CIYABBY* genes. The most YABBY genes were found on chromosome 5, and there were 2 on it. There was just one YABBY gene on chromosomes 1, 2, 6, 8, 10, and 11. Contrarily, it was discovered that chromosomes 3, 4, 7, and 9 did not have the YABBY gene (Supplementary Fig. S4). In order to evaluate segmental and tandem duplication of the *CIYABBY* gene family in chromosomal position, synteny analysis for *CIYABBY* genes was also carried out (Fig 2a). While *CIYABBY7* and *CIYABBY6* genes were clustered together on the same chromosome, it was possible that these genes evolved via tandem duplication in the watermelon YABBY genes because three homologs gene pairs were spread unevenly throughout the watermelon genome (Fig. S4). The comparison research shows that the genomic family has expanded due to the lower number of genes that are shared between watermelon and other members of the Cucurbit family. Watermelon and Arabidopsis only share eight genes, but watermelon and bottle gourd share 12 genes, according to a dual syntenic analysis. Additionally, watermelon and melon share 7 genes, whereas watermelon and musa share just 1, revealing their evolutionary ties within the cucurbita family (Fig 2b).

Fig 2(a): Genome-wide synteny analysis of *StYAB* genes showing the dominance of segmental duplication and rare occurrence of tandem duplication. Joining lines showing the duplicated *StYAB* genes in the genome.  
Fig 2(b): Dual synteny analysis of Potato-Arabidopsis, Potato-Tomato, and Potato-Chili. Orange bars represents chromosomes of potato while green bars represent chromosomes of Arabidopsis, Tomato, and Chili respectively. Red lines showing the duplicated genes in the respective genomes.

Additionally, TB Tools software was used to determine the date of gene duplication using pairwise alignment, which produced Ks and Ka values, followed by a manual calculation of Ka/Ks (Fig 3). The ratio of non
synonymous (Ka) to synonymous (Ks) mutation was indicated as Ka/Ks. Ks displays the number of synonymous substitutions per synonymous site, whereas Ka shows the number of non synonymous substitutions per non synonymous site. The CIYABBY5/CIYABBY6 pair had a ratio of 0.36, while the CIYABBY2/CIYABBY1 pair had a ratio of 0.17. The homologs group CIYABBY2/CIYABBY1 tandem gene duplication date was anticipated to be 43.2 Mya, which was the greatest, while the segmental duplication date for the paralogous pair CIYABBY8/CIYABBY4 was predicted to be 30.7 Mya, which was the lowest. Segmental duplication date for the final paralogous pair, CIYABBY5/CIYABBY6, was estimated to be 34.2 Mya (Table S3). In the watermelon, all 3 paralogous group pairs had Ka/Ks ratios larger than 0.3 but lower than 1, which raises the possibility of significant functional divergence following duplication brought on by purifying selection.

**Figure 3:** Time of gene duplication estimated for different paralogous pairs of watermelon CIYAB genes on the basis of Ks and Ka values. Analyses were conducted using Tbtools (Software). Ka/Ks represents the ratio of nonsynonymous (Ka) versus synonymous (Ks) mutations.

**Analysis of CREs:** The presence and arrangement of different CREs at the binding site of TFs on the promoter region influence the spatiotemporal transcriptomic expression of genes. To assess the potential roles of genes, numerous CREs can be analyzed in silico (Bulow & Hehl, 2016) (Jones & Vandepoele, 2020). CRE components with functions like anaerobic induction, ethylene production, sensitivity to light, wound healing, hormone-specific stress, and drought were noticed (Fig. 5; Table S4). Interestingly, the ARE element, which was necessary for anaerobic induction, was present in 6 of the 8 CIYAB genes (CIYABBY1, CIYABBY2, CIYABBY4, CIYABBY6, CIYABBY7, and CIYABBY8). There were also 6 light responsive elements. Only one CIYABBY1 gene displayed the GT1 motif, whereas the G-box has three CIYABBY genes (CIYABBY1, CIYABBY2, and CIYABBY4), the AAAC motif has only one CIYABBY2 gene, the TCT motif includes four CIYABBY genes (CIYABBY2, CIYABBY6, CIYABBY7, and CIYABBY8), the TCCC motif has two CIYABBY genes (CIYABBY7, CIYABBY8), and C-box possess 1 CIYABBY7 gene, Box 4 elements, a piece of a conserved DNA module involved in light responsiveness, were only found in 2 CIYAB genes (CIYABBY4 and CIYABBY7). The ABRE element was present in three CIYAB genes (CIYABBY1, CIYABBY2, and CIYABBY4) and was involved in the abscisic acid response. 4 CIYABBY genes (CIYABBY1, CIYABBY2, CIYABBY5, and CIYABBY8) have the salicylic acid responsive TCA element, while 4 other CIYABBY genes (CIYABBY2, CIYABBY3, CIYABBY6, and CIYABBY7) have the salicylic acid responsive TCA element. Only 2 CIYABBY genes (CIYABBY4, CIYABBY6) showed the wound-responsive WUN motif. CIYABBY2, CIYABBY5, and CIYABBY6 showed TC-rich repeats that demonstrate responses in defence and stress, respectively. The MBS element was present in 2 CIYABBY genes (CIYABBY2, CIYABBY4) and was connected to drought-inducibility. The auxin responsive TGA element was present in 2 CIYABBY genes (CIYABBY6, CIYABBY8). Ethylene response element ERE is present in 3 CIYABBY genes (CIYABBY1, CIYABBY5, and CIYABBY8). The MYB, which was involved in plant development, is present in 6 CIYABBY genes (CIYABBY1, CIYABBY4, CIYABBY5, CIYABBY6, CIYABBY7, and CIYABBY8), but only one CIYABBY8 gene possesses the gibberellin responsive GARE-motif, and another gibberellin responsive element p-box contains only 1 CIYABBY2 gene. The STRE, which was engaged in stress-related processes, was demonstrated by 5 CIYABBY genes (CIYABBY1, CIYABBY2, CIYABBY3, CIYABBY7, and CIYABBY8). Only 1 CIYABBY2 gene includes the ABRE4 which mediates ABA-dependent stress responses, as well as ABRE3a, a positive regulator of abiotic stress and ABA signalling. The MYC, which stimulates the expression of proliferative genes, was found in 2 CIYABBY genes (CIYABBY2, CIYABBY8). No CIYABBY genes, however, have been discovered to be deficient in the previously stated cis regulatory regions.
The eight watermelon YABBY genes' identified CREs are displayed in (Fig 4) and Table S4 along with their functional annotations.

Fig 4: Cis-regulatory elements in putative CIYAB promoters which are associated with different plant developmental process.

Transcriptomic analysis of watermelon YABBY genes: With the aid of high throughput sequencing, the differentiating expression patterns of all watermelon YABBY genes at distinct developmental stages were also examined (Song, Joshi, DiPiazza, et al., 2020) (Song, Joshi, & Joshi, 2020). There were two different kinds of studies on the effects of salt and drought stress. Watermelon seedlings were subjected to short-term salt stress in order to study and discover genes and pathways associated with response to salt stress treatments, such as differences in photosystem II photosynthetic efficiency and free amino acids. A heat map displaying the expression of the watermelon CIYABBY genes (Fig 5).

Out of the eight genes that were subjected to salt stress, CIYABBY2, CIYAABY3, CIYABBY4, CIYABBY6, and CIYABBY8 showed expression in five of them. CIYABBY1 nevertheless displayed expression, but in the control group. The ethylene biosynthesis signaling pathways were activated during salt stress treatments, according to the hypothesis, and photosystem efficiency was decreased under salt stress. Therefore, the genes that were expressed play a role in the ethylene biosynthesis signaling pathway, which explains why they were highly expressed. However, CIYABBY1, CIYABBY7, and CIYABBY5 play a role in pathways that increase photosynthesis by enhancing the efficiency of photosystem II. This is why they were not expressed in watermelon seedlings under salt stress (Song, Joshi, & Joshi, 2020)

Fig 5: Heat map shows the expression profile of the CIYAB genes in leaves of watermelon cultivar under different levels of Drought stress.

The goal of the drought stress experiment was to explore the accumulation of drought-induced citrulline in watermelon leaves by monitoring stress treatments using physiological measurements. Watermelon seedlings were stressed in a semi-controlled environment and open field. A heat map represents the expression profiles of the
watermelon ClYABBY genes (Fig 6) Because only 5 hits from the available RNA seq data were identified, the expression of 5 of the 8 ClYABBY genes, namely ClYABBY1, ClYABBY2, ClYABBY3, ClYABBY4, and ClYABBY8, was detected. Based on the assumption, these genes play a part in the biosynthesis of citrulline, which accumulated in the leaves of watermelon seedlings under drought stress conditions, while ClYABBY5, ClYABBY6, and ClYABBY7 are thought to play a minor role in the biosynthesis of citrulline or may not be directly associated to its creation in young watermelon plants, which may explain why they did not exhibit expression (Song, Joshi, DiPiazza, et al., 2020).

![Fig 6: Heat map shows the expression profile of the ClYAB genes in seedlings of Watermelon cultivar under different levels of salt stress.](image)

**Putative miRNA targets in watermelon:** The results showed that nine miRNAs in total were discovered, and four of the eight ClYABBY genes were their targets. None of these miRNAs targeted the remaining four ClYABBY genes (ClYABBY2, ClYABBY3, ClYABBY4, and ClYABBY5) (Table S6). These miRNAs were between 20 and 22 amino acids long. Per the ClYABBY gene, there were between 156 and 825 miRNAs that targeted these genes. Cla-miR156c, Cla-miR157b, Cla-miR159, Cla-miR820, Cla-miR157a, and Cla-miR159b were six of the nine mature miRNAs that target ClYABBY6, while ClYABBY7, ClYABBY8, and ClYABBY1 were each targeted by just one mature miRNA. Cla-miR162, Cla-miR825, and Cla-miR812 all target ClYABBY7, ClYABBY8, and ClYABBY1, respectively (Table S6). Therefore, ClYABBY6 was the sole gene that had the greatest number of miRNAs targeting it. The most targeted group was Group INO, which had six mature miRNAs target it. In contrast, just one miRNA targeted Group CRC, Group YAB5, and Group YAB3, whereas no miRNAs targeted Group YAB2.

**Discussion**

YABBY PSTrFs were divided into 5 families (Group AtINO, AtCRC, AtYAB5, AtAFO/AtYAB3, AtYAB2), following the phylogenetic and domain analysis patterns of A. thaliana (D. Lijavetzky, P. Carbonero, & J. Vicente-Carbajosa, 2003), citrus (Wu, Fu, & Yi, 2016), and eggplant (Wei et al., 2018). The Cucurbit Genomics Database's recently released data was utilized to identify ClYABBY genes at the genome level (CuGenDB) (http://cucurbitgenomics.org/) (Table 1). Using phylogenetic analysis, the watermelon's 8 YABBY genes were divided into five groups (Group AtINO, AtCRC, AtYAB5, AtAFO/AtYAB3, AtYAB2) (Fig. 1, Table 1). Watermelon had less YABBY genes than rice (30 OsYABBY) (X. Yang & Tuskan, 2006), Arabidopsis (36 AtYABBY) (X. Yang & Tuskan, 2006), tomato (34 SiYABBY) (Cai et al., 2013), banana (74 MaYABBY) (Dong, Hu, & Xie, 2016), and Chinese cabbage (76 BrATYABBY) (Ma, Li, Wang, Tang, & Xiong, 2015).
Exons and introns were present in all the genes, albeit in varying amounts (Table 1). Genes with the same number of exons and introns were classified into the same clade following analysis. The watermelon CIYABBY genes found in the same family (Groups) generally had the same exon-intron architecture, however variants were found in other families. Similar intron-exon configurations have also been found in Arabidopsis, rice, and soybean (Diego Lijavetzky, Pilar Carbonero, & Jesús Vicente-Carbayosa, 2003) (Gu et al., 2013) suggesting that these structures were evolutionarily preserved.

The examination of conserved motifs added more support to the classification of CIYABBY genes. To find the conserved motifs, the MEME analysis tool was used to upload all of the CIYABBY protein sequences. There were consequently a total of fifteen conserved motifs found. The length of the motifs of CIYABBY proteins discovered by MEME ranged from 6 to 50 amino acids. All of the watermelon YABBY proteins shared Motif-1, or YABBY (Fig. 2). Motif-1 and Motif-2 are present in all Group INO proteins as well as all Group YAB2, YAB3, CRC, and YAB5 proteins with the exception of CIYABBY8 which lack Motif-2. Extra particular motifs found in several YABBY proteins may be important for various functions. The most intricate motif pattern, Motif-5, was found in the YABBY proteins from Group YAB5, and Motif-10 is unique to this subgroup. In contrast to Group YAB5, the members of the YAB2 group have a simpler motif pattern, although some of the group members also feature group-specific patterns, like Motif-4. Additionally, the YAB3 group included group-specific motifs like Motif-3 and Motif-6. In order to comprehend the probable functions of the Group CRC-specific motifs, the GO annotations of the Group CRC genes in Arabidopsis were first examined. We discovered something intriguing: In contrast to the Arabidopsis YABBY genes (Table 2) in other groups, AtCRC genes in Group CRC have responsibilities in "Polarity establishment in carpel and nectary development," which is comparable to GO annotations of watermelon CIYABBY7 gene in Group CRC which have roles in "nectary development," "definition of floral organ identity," "Stamen development," "style development," and "polarity specification of adaxial/a (Table 2). Rice (Yamaguchi et al., 2004), Arabidopsis (Alvarez & Smyth, 1999; Bowman & Smyth, 1999; Lee et al., 2005) and cucumber (Yin et al., 2022) have all shown similar functions for genes belonging to the CRC clade. This suggests that these genes function similarly in Arabidopsis, rice, cucumber, and watermelon.

The distribution of motifs in the watermelon YABBY proteins (Fig. 2) shows how they evolved, as inferred from the phylogenetic tree (Gupta et al., 2015) (Malviya et al., 2015). The alignment of the watermelon CIYABBY protein sequences and the motif data analysis by Motif finder and domain analysis by NCBI CDD all pointed to a highly conserved YABBY domain. Evolutionarily, YABBY PSTFs have been preserved in a variety of plants. In addition to the YABBY domain, 14 newly discovered unique motifs were identified and distributed differently among the CIYABBY genes (Fig. 2). Consequently, one or two motifs were revealed to be preserved in the same group of distinct comparative crop species; however, some differences were observed between motifs of other groups, suggesting that members of the same group functioned similarly. Furthermore, the CIYABBY genes’ structural arrangement was retained throughout all five separated groups, including other species such as Arabidopsis, Cucumber, and Musk melon (Dong et al., 2016; Diego Lijavetzky et al., 2003; Nasim, Malviya, Kumar, & Yadav, 2016; X. Yang & Tuskan, 2006). In addition, an analysis of the subcellular localization of CIYABBY proteins using the online tool WoLF PSORT (https://wolfpsort.hgc.jp/) revealed that all CIYABBY proteins had nuclear localization, with only minor variances among the cytoplasm, peroxysomes, and chloroplasts, all of which were typically found in the nucleus. (Table 2, Fig S3). A gene's position on a chromosome can be used to evaluate gene duplication. It is known that segmental duplication occurs when two or more genes from the same species are located on distinct chromosomes of the same species. Tandem duplication occurs when two or more genes are present on the same chromosome (Panchy, Lehti-Shiu, & Shiu, 2016). Despite the fact that the high number of watermelon YABBY genes on chromosome 5 (Fig. S4) indicates the occurrence of tandem duplication, some of the CIYABBY genes were also displaying segmental duplication (Fig. S4). In the YABBY family of genes, segmental duplication predominated in chickpea (Nasim et al., 2016) and pigeon pea (Malviya et al., 2015). While domain duplication may have contributed to the larger number of YABBY genes throughout the evolution of eukaryotic plants, gene duplication is the primary method involved in the proliferation of gene families (Moore & Purugganan, 2005; Taylor & Raes, 2004).

Understanding amino acid substitution requires an understanding of the Ka/Ks ratio (Fig. 3). When Ka/Ks is less than 1, purifying selection takes place, and when Ka/Ks is greater than 1, positive selection takes place (Hurst, 2002; Ziheng Yang & Bielawski, 2000). Variations in selection pressure typically result in changes to a protein’s unique amino acid pattern, which were crucial for understanding how various proteins operate (Morgan, Loughran, Walsh, Harrison, & O’Connell, 2010). There was just a small amount of variation in Ka/Ks ratios between CIYABBY genes. Despite the changes, all anticipated values of Ka/Ks ranged from 0.17 to 0.36, which is less than 1, indicating that the sequences of YABBY present in all groups underwent strong purifying selection pressure and that only a small number of sites may have been impacted by
positive selection only during the process of evolution (Fig.3).

Using the available RNA seq data, the expression of all \textit{CIYABBY} genes in several watermelon experiments was examined. Five \textit{CIYABBY} genes were found to express in the leaf tissue of seedlings during a drought stress experiment. As stated in the results, the cis-regulatory element MBS in \textit{CIYABBY2} and \textit{CIYABBY4} has a role in drought inducibility and expressed itself in this experiment. Furthermore, Go annotations of these genes in comparison to Arabidopsis showed how these gene expressions occur in leaf and stem tissue. The functions of \textit{CIYABBY2}, "abaxial cell fate specification" and "inflorescence meristem growth," which are consistent with those of Arabidopsis (Siegfried \textit{et al.}, 1999), confirm their function in the leaves of early watermelon seedlings under drought stress. \textit{CIYABBY2} is a member of Group YAB3, which also includes the axial regulators \textit{AtYAB3} and \textit{AtAFO} Arabidopsis genes. When watermelon seedlings are subjected to drought stress, above mentioned genes are engaged in pathways that create citrulline, which was used in biosynthesis (Song, Joshi, DiPiazza, \textit{et al.}, 2020).

However, it was discovered in additional studies on salt stress that mechanisms for the manufacture of ethylene are activated under salt stress. Therefore, 5 \textit{CIYABBY} genes were expressed in watermelon seedlings, indicating that they may be involved in ethylene responsiveness and stress-mediated response. \textit{CIYABBY6} has CREs TC-rich repeats, which have functions associated in defence and stress responsiveness, and \textit{CIYABBY8} has CRE ERE, which has functions involved in ethylene responsiveness, as described in the results. Since photosystem II efficiency decreases under salt stress, it follows that the genes that did not exhibit any expression should have a role in light sensitivity and photosynthetic activity. This supports the cause for the expression of these genes. \textit{CIYABBY1} has the light responsive CREs G-box and Gt1 motif, while \textit{CIYABBY7} has the light responsive CREs I-box, TCCC-motif, and TCT-motif (Song, Joshi, 

From the developmental stage to pathogen defence and maintaining healthy internal conditions, microRNAs play a critical role in controlling plant growth (Carbone \textit{et al.}, 2019; Samad, 2017; Spanudakis, 2014; Terzi, 2008). Regardless of the type of species they were found in, miRNAs are present in the majority of plant species in a manner that is preserved, which explains their specific function. The majority of the \textit{CIYABBY} genes have transcriptional activity linked with them, which suggests that they inhibit miRNA activity. Due to this, members of the Cla-miR156 and Cla-miR825 families only targeted one out of the eight \textit{CIYABBY} genes (Table. S6). Only one \textit{CIYABBY6} gene was targeted by each of the two Cla-miR157 (Cla-miR157a, Cla-miR157b) and Cla-miR159 (Cla-miR159a, Cla-miR159b), whereas Cla-miR162 was targeted by \textit{CIYABBY7}. Because these two genes are on the same chromosome 5, it is likely that the majority of their origin and activity is on chromosome 5. Numerous plant species include Cla-miR156c, which inhibits the growth of adventitious roots (Xu, Chen, Ying, 

The surprising variations in \textit{YABBY} gene counts among different plant species are an intriguing part of our research. While certain plants, such as tomato, pepper, and potato, have almost comparable quantities of \textit{YABBY} genes, others, such as bottle gourd, melon, wax gourd, and watermelon, have far fewer. The hypothesis suggests that the differences in gene numbers were a result of gene duplication and loss events that occurred during plant evolution. Furthermore, it was discovered that segmental duplications were more important in the expansion of the \textit{YABBY} gene family in watermelon than tandem duplications. This finding highlights the mechanisms responsible for the diversification of this gene family within the watermelon genome. The research also investigated tissue-specific expression patterns, revealing that \textit{CIYABBY} genes exhibited high expression levels in roots and seedlings. This information sheds light on the critical roles that these genes play in early

Conclusion

The \textit{CIYABBY} PSTrFs gene family inside the watermelon genome was explored in this comprehensive study. The research focused on the structural and functional properties of the eight \textit{CIYABBY} genes, classifying them into five separate groups based on Arabidopsis classification. These genes were essential not only for normal watermelon growth and development, but also for alleviating salt and drought-induced stress. Furthermore, the study examined miRNAs that target \textit{CIYABBY} genes, revealing their role in defense mechanisms and stress responses, notably under salt and drought stress. This study's computational results offer potential for applications in molecular cloning, gene expression profiling, and exploring relationships with various transcription factors (TFs).

Table 3: Help us understand the roles played by miRNAs, more investigation is needed.
plant growth and root function. Finally, distinct CREs linked to responses to light, wound, drought, auxin, stress, salicylic acid, and abscisic acid (ABA) in seedlings and roots were identified through cis-regulatory element (CRE) studies. This study enhances the understanding of plant adaptation and stress response systems, with direct implications for agriculture and plant science. In conclusion, the research offers valuable insights into the activities and regulatory networks of CIYaYABBY genes, serving as a valuable resource for future plant biology research.

Authors’ Contribution

References


neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC Evolutionary Biology,* **10**(1), 39.


Publisher’s note: JOARPS remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.