African Journal of Biological Sciences

https://doi.org/ 10.33472/AFJBS.6.5.2024. 7976-7994

Evolutionary and Regulatory Analysis of B3 DNA-Binding Superfamily Genes in Sponge Gourds

Singh Vasundhra [1], Gupta Divya [1,2], Vaish Swati [1], Singh Nootan [1], Tiwari Neeraj[1] and Sharma Sapna [1]*

[1]Faculty of Biosciences, Institute of Biosciences and Technology,
Shri Ramswaroop Memorial University,
Lucknow-Deva Road, Barabanki 225003, Uttar Pradesh, India
[2] Division of Endocrinology and Metabolism, Department of Medicine
University of Pittsburg, Pittsburgh, USA

[1*]Corresponding author:
Dr. Sapna Sharma
Assistant Professor
Faculty of Biosciences Institute of Bioscience and Technology,
Shri Ramswaroop Memorial University, Lucknow-Deva Road, Barabanki, 225003, Uttar Pradesh, India
sapnasharma.fobs@srmu.ac.in, sapsharma1903@gmail.com

Article History
Volume 6, Issue 5, 2024
Received: 22 May 2024
Accepted: 29 May 2024
doi: *10.33472/AFJBS.6.5.2024. 7976-7994*

Abstract

This study focuses on the REM, RAV, ARF, and LAV families of B3 DNA-binding superfamily genes in sponge gourds. Through in silico genome-wide analyses, the study identifies these families and highlights the absence of ABI3 and HSI families, indicating distinct evolutionary adaptations. Different forms of the REM, ARF, and LAV families suggest that their functions change during different stages of development or in response to environmental factors. The study identifies a total of 52 B3 genes and categorizes them into four classes: ARF, REM, LAV, and RAV. Physicochemical investigation reveals that the protein is hydrophilic and located in the nucleus. Phylogenetic analysis shows distinct clades specific to each class based on sequence similarity. Gene duplication patterns across chromosomes reveal evolutionary dynamics, while analysis of gene architecture and conserved motifs emphasizes diverse functional roles. Structural and phosphorylation analyses provide insights into protein function, and expression profiling reveals tissue-specific roles. Overall, this study provides a molecular framework for understanding the functional analysis of B3 genes in sponge gourds, particularly during fruit development. It also contributes valuable information about protein properties and tissue-specific functions, improving our understanding of the dynamics of the B3 gene family in sponge gourds.

Keywords

DNA binding proteins, Sponge gourd, evolutionary analysis, multiple alignment

1. Introduction

Sponge gourd (2n = 2x = 26), a member of the Cucurbitaceae family and the Luffa genus, is an annual climbing herb indigenous to subtropical and tropical regions of Asia (Tyagi et al., 2020). The consumable fruits of sponge gourd boast a nutritional profile encompassing carbohydrates, protein, vitamins, crude fibre, and minerals (Rodríguez-Moreno et al., 2011). Additionally, derivatives and individual components derived from these fruits' manifest diverse pharmacological effects, including immunomodulation, antioxidant attributes, anticancer properties, and anti-inflammatory responses, contributing beneficially to human health (Shendge & Belemkar, 2018).

The B3 domain was initially discovered in the VIVIPAROUS (VP1) gene of maize, characterized by three distinct regions labelled B1, B2, and B3 (Suzuki et al., 1997). An analogous orthologue, ABI3 (ABSCISIC ACID-INSENSITIVE3), was subsequently identified in Arabidopsis thaliana (Giraudat et al., 1992). These genes, which share the B3 domain, fall into five major classes distinguished by their structural similarities, including proteins from the ABI3/VP1 [1], RAV (Related to ABI3/VP1) (Kagaya et al., 1999), REM (Reproductive Meristem) (Franco-Zorrilla et al., 2002), ARF (Auxin Response Factor) (Ulmasov et al., 1997), and HSI (High-level expression of sugar-inducible gene) (Suzuki et al., 2007) families. The DNA binding specificity of the B3 domain has been explored in families for the mentioned genes. The ABI3 family's B3 domain recognizes the Sph/RY element in the CATGCA sequence (Monke et al., 2004). RAV family proteins have an N-terminal DNA binding AP2/EREBP domain recognising the CAACA sequence and a C-terminal B3 domain recognising the CACCTG sequence (Kagaya et al., 1999). The ARF family is distinct feature by an N-terminal B3 domain recognizing the TGTCTC sequence (auxin response elements - AuxREs), a divergent centred domain serving as a transcriptional activation or repression domain (Ulmasov et al., 1999), and a C-terminal dimerization domain containing motifs III and IV similar to motifs found in Aux/IAA proteins (Ulmasov et al., 1997). Notably, research has shown that B3 domains from various families display binding preferences for distinct DNA sites. Despite this

diversity, these peptides/proteins share the structural similarities in framework for DNA recognition. A Nuclear Magnetic Resonance spectroscopy examination of the B3 domain in the At1g16640 protein from Arabidopsis, classified within the REM family, unveiled a unique fold reminiscent of RAV1. This structural fold features an open barrel of arranged seven-stranded β-sheet, complemented by two small α-helices (Waltner et al., 2005). Importantly, the amino acid sequence of the gene At1g16640 deviates significantly from other superfamily members, prompting inquiries into its DNA-binding capability. Nevertheless, despite the dissimilarity in sequence, studies have established that VRN1 (VERNALIZATION1), another REM family member, displays DNA-binding activity *in vitro* in a non-sequence-specific manner (Levy et al., 2002). This suggests the potential for a reduction in specific DNA binding within the At1g16640 domain, while still retaining a general affinity for DNA.

Proteins harbouring the B3 domain play vital roles in diverse plant processes. Specifically, three transcriptional activators—FUSCA3 (FUS3), LEAFY COTYLEDON2 (LEC2), and ABSCISIC ACID INSENSITIVE3 (ABI3)—along with three repressors—HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE 2 (HSI), HSI L1, and HSIL2 or VP1/ABI3-LIKE (VAL)—from the ABI3 and HSI/VAL families have been recognized as contributors to seed development and maturation (Stone et al., 2002). While RAV genes lack comprehensive characterization, some have been associated with the regulation of growth, development, and flowering time (Castillejo & Pelaz, 2008).The most extensively studied B3 family is the ARF family, renowned for its regulatory role in various responses to auxin, encompassing additional regulatory systems (Guilfoyle & Hagen, 2007). In contrast, the vital REM family has limited functional details available as of now (Franco-Zorrilla et al., 2002), except for VRN1 (VERNALIZATION 1), which facilitates flowering (Sung & Amasino, 2004). Functionally characterized B3 proteins from the ABI3, HSI, RaV, and ARF families have been demonstrated to primarily engage in hormone signalling pathways, including those associated with auxin, gibberellin, brassinosteroid, and abscisic acid.

2. Methods

2.1. B3 gene identification through the Cucurbita Genome database

To ascertain members of the B3 gene family in Sponge gourd, an in-silico genome-wide pBLAST search was executed on the Cucurbita genome database (http://cucurbitgenomics.org) using an E-value of 0.00000. The query set included 244 amino acid sequences of B3 gene family members obtained from Arabidopsis, rice, and citrus. Arabidopsis thaliana B3 protein sequences were sourced from The Arabidopsis Information Resource (TAIR), Oryza sativa amino acid sequences from the UniPort database, and Citrus sinensis amino acid sequences from the "Citrus pan-genome to breeding database" (Lie et al., 2020). Multiple full-length B3 genes were identified from the Sponge gourd genome database. Amino acid, genomic DNA sequences, and coding DNA sequences (CDS) of the crops were acquired from their respective genome databases. Initially designated as *Momordica charantia* (Mc), the identified genes were further annotated based on their class and chromosomal localization. Following this, the downloaded sequences underwent additional physical and chemical characterization. The validated amino acid sequences from Sponge gourd underwent analysis through the NCBI Batch CD-search with default parameters to authenticate characteristic domain organization and ascertain their corresponding classes (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi).

2.2. Physical and Biochemical Characterization and subcellular localization of Sponge gourd B3 genes

The protein sequences derived from Sponge gourd were further subjected to physicochemical characterization, encompassing an examination of amino acid composition, molecular weight, isoelectric point (pI), total counts of negative and positive charged residues, extinction coefficient, instability index, aliphatic index, and Grand Average of Hydropathy (GRAVY). This comprehensive analysis was performed utilizing the ExPASy ProtParam tool (http://web.expasy.org/protparam/) with standard parameters. The Sponge gourd protein sequences are subcellularly localization and autonomously predicted using three different tools: DeepLoc (http://www.cbs.dtu.dk/services/DeepLoc/), Plant-mSubP (http://bioinfo.usu.edu/Plant-mSubP/), and WoLF PSORT (www.genscript.com/wolfpsort.html).

2.3. Conserved motifs and Genomic organization analysis of Sponge gourd B3 genes

The Sponge gourd proteins with conserved motifs were identified employing the Multiple Em for Motif Elicitation (MEME) program (http://meme-suite.org/). The analysis involved the search for 15 motifs with a motif width ranging from 6 to 50. The remaining parameters were configured to default settings. The assessment of intron-exon numbers and their organization in Sponge gourd was executed utilizing the Gene Structure Display Server 2.0 (GSDS; http://gsds.cbi.pku.edu.cn/). The genomic DNA sequences and coding DNA sequences (CDS) were formatted in FASTA and systematically arranged in distinct text files, ensuring meticulous attention to sequence consistency.

2.4. Chromosomal localization, phylogenetic and analysis of Gene duplication of the B3 gene of Sponge gourd

To conduct phylogenetic research, the B3 sequences of Sponge gourd were aligned with the sequences of previously characterized B3 gene families from *O. sativa, C. sinensis,* and *A. thaliana* by using Clustal Omega. The evolutionary tree was created using the iTOL program, using the tree data acquired from Clustal Omega. An investigation of gene duplication occurrences was carried out by doing a pBLAST search of Sponge gourd B3 genes with each other using the NCBI pBLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). Homologous gene pairs placed within a 100-kb range on a single chromosome are defined as tandem duplicated (TD), whereas those positioned beyond the 100-kb zone or on different chromosomal sites are categorized as segmental duplicated (SD) genes. The evaluation of non-synonymous rate (dN), synonymous rate (dS), and evolutionary constraint (dN/dS) among the duplicate Sponge gourd B3 gene pairs was conducted using the PAL2NAL online tool (https://bio.tools/pal2nal).

2.5. Alignment of Protein sequence of Sponge gourd

To highlight the unique sequences of each class in Sponge gourd, the pertinent catalytic residue amino acid sequences from the crops were aligned with the protein sequences of *O. sativa, A. thaliana,* and *C. sinensis* using Clustal Omega. Following the alignment, the aligned protein was visualized through ESPript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

2.6. Protein secondary structure analysis of Sponge gourd

The secondary structure of Sponge gourd B3 proteins was forecasted using the Self-Optimized Prediction Method with Alignment (SOPMA) (https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html).

2.7. Protein phosphorylation site analysis of Sponge gourd

To predict potential phosphorylation sites, the local prediction software GPS6.0 (http://gps.biocuckoo.cn/online.php) was employed, utilizing a high threshold value.

2.8. Cis-acting regulatory elements analysis of Sponge gourd

The promoter region, spanning 2kb-pairs upstream of the transcription start site on the Sponge gourd B3 genomic DNA sequence, was obtained from the Cucurbita genome database. In order to analyze these (cis-acting) regulatory elements, the locus ID was used to extract this area. We used the PlantCARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to analyze the extracted regulatory sequences and find several components that respond to hormones, plant growth, and stress. The resulting schematic representation of the promoters was generated using TBtools software.

2.9. Expression profiling of B3 gene using RNA-seq data of Sponge gourd

To examine the foundational gene expression levels of the B3 genes under diverse tissue development conditions and abiotic stress, RNA-Seq data for Sponge gourd was acquired from the Cucurbit Genomics Database (http://cucurbitgenomics.org/v2/expression).

3.Results

3.1. Identification of B3 DNA Binding superfamily members in Sponge gourd

The B3 DNA binding superfamily includes five families in plants namely ARF, ABI3, HSI, RAV and REM. In sponge gourds predominantly as seen by analysis of B3 DNA binding proteins by in silico genome-wide pBLAST showed the presence of 52 B3 genes. These genes further grouped into four groups REM, RAV, ARF and LAV as illustrated in Fig. 1 on the basis's presences of B3 domain and additional domain. However, ABI3 and HSI have not been observed. There are multiple isoforms of REM which have been observed followed by ARF and LAV. A total of 20

copies of REM have been observed in sponge gourd as per this analysis. The B3 domain is seen in families with REM showing only the B3 domain distinctly. LAV family has the B3 and zfCW domains. Many of the ARFs have B3, Auxin IAA superfamily present in them. RAV has an AP2 domain in addition to B3 (Kandeel et al., 2023).

3.2. In silico physicochemical characterization of LcB3 DNA binding proteins and their subcellular localization prediction

Various physicochemical characterization analyses manifest that amino acid length ranged from 213 (LcREM4) to 1111 (LcARF12) with their complementary molecular weight 23.90 kDa to 123.58kDa. The isoelectric point ranged from 5.21 (LcLAV5) to 9.88 (LcREM3), with an average value of 7.39. Among the 52 LcB3 proteins, 29 had an acidic character, whereas 23 showed a basic nature. The overall mean hydropathy values of the majority of B3 members across all classes are mostly negative, suggesting that all B3 members exhibit a hydrophilic characteristic (TableS1). B3 DNA binding superfamily being a transcription factor have predominant nuclear localization. A similar observation is seen even in sponge gourds with the bulk of them having nuclear and very few having chloroplast and cytoplasmic localization.

3.3. Analysis of phylogenetic relationship, chromosome location, and gene duplication analysis of identified B3 gene

In order to determine the evolutionary connection among the 52 LcB3 proteins, a multiple sequence alignment was performed, and phylogenetic tree was created (Fig.2). The LcB3 proteins were categorized into six distinct subgroups, labelled as ARF, RAV, LAV, REM, ABI3 and HSI. The REM subgroup displayed the maximum number of members while the HSI and ABI3 subgroup comprised just 3 individuals. It is important to note that the 4 subfamilies of RAV, REM, LAV, and ARF, which are categorized based on gene domains, do not exhibit noticeable grouping in the phylogenetic tree. The chromosomal location of various gene families has shown Fig 3. REM family genes have been found to localize on chromosomes 4,7,10,11 and 12 with maximum localization on chromosome 7. Similarly, ARF genes showed localization on chromosomes 1,2,3,4,8,12 and 13. RAV family genes were found to localise on chromosomes 2,3,4,6,8, 11 and

13.LAV genes were relatively less in number with localization on 4,6,9 and 10 as presented. In this study we get two pair of tandem duplication and one pair of segmental duplication (TableS2). The dN/dS ratio was less than 1, indicating that all three gene pairs were subject to significant purifying selection. The replication period of these gene pairs was also computed and found to be between 5.83 and 67.66 million years ago (Fig.4).

3.4. Conserved motifs and Gene architecture of identified B3 genes

The investigation focused on the identification of LcB3 genes by selecting 15 conserved motifs. The findings indicated that individuals within the same subgroup had comparable motifs and gene architectures, suggesting possible functional commonalities (Fig.5) (TableS3). The LcARF subgroup had the greatest number of motifs, totalling 11, while LcRAV and LcLAV subgroups had the fewest, with just 3 and 2 respectively. The main motifs in LcARF, LcRAV, and LcLAV are motif 2 and motif 5. Conversely, the primary motifs found in LcREM subgroups consist of motif 5, motif 12, motif 14. It is crucial to observe that motif 9 is only found in LcREM5 and motif 15 is only found in LcREM13 and LcREM14 and is absent in the other subgroups. The number of exons within each subgroup exhibited significant heterogeneity. LAV and RAV show a similar pattern of presence of conserved domains. In the case of LAV, the multiple gene copies show the presence of only exons without introns and upstream or downstream elements. In the case of RAV introns are of varying length and the cis-acting elements were extremely short. However, in the case of ARF, the introns were present in all genes albeit there was variation in its length which could be seen in Fig. 5 provided below. All the genes had upstream or downstream elements. In the case of REM, there are extensive differences among genes. LcREM14, LcREM11 and LcREM12 had introns were very short. In the case of othe**r** REM either upstream or downstream or both were present with distinct introns (Liu et al., 2020).

3.5. Multiple sequence alignment of LcB3 proteins

Multiple sequence alignments were performed using the Clustal omega program in order to illustrate the conserved residues. Through the process of aligning the various B3 protein sequences of sponge gourd along with Arabidopsis, rice, and citrus the discovery of class specific conserved

residues was accomplished. It was discovered that there are three highly conserved sequences in ARF class (RGQPK/RR), (Fig 6A) RAV class, (WN/RSSQS) (Fig 6B) and the LAV family, (WPNNKSR) (Fig 6C). There was a significant amount of sequence variation in the REM class (Fig. 6), which indicates that they have functional diversity.

3.6. Cis-acting regulatory element analysis of identified LcB3 gene

The expression pattern of genes is mostly determined by a promoter sequence. Additionally, the cis-regulatory regions located within the promoter sequence often include specific locations where transcription factors and other proteins may bind. This binding can subsequently result in the activation or suppression of gene expression. To determine the significant cis-regulatory elements located on the B3 genes promoters, we evaluated the start site 2 kb upstream sequences of each B3 gene using the care database. The examination of promoter regions of all B3 genes using computational methods indicated the existence of many distinct cis-elements in the upstream region of these genes. Cis-acting elements showed wide differences among the identified family members. The present study identified 36 cis-elements, classified according to their significance in physiology of the plant, such as cellular development-related element, light-responsive element, stress-responsive element, hormone-responsive element, and other elements, in addition to the four core elements (AT ~ TATA box, CAAT box, TATA box, TATA). The light-responsive elements exhibited the largest quantity, followed by hormone- and stress-responsive elements (Fig. 7). LcARF9 had the greatest count of 61 cis-elements, although LcREM8 had just 15 cis-elements in its promoter region. The numbers 193 and 205 corresponded to the maximum occurrences of MYB and MYC cis elements, respectively. Subsequently, the number of abscisic acid-responsive elements was the largest, totalling 153. The abundance in the promoter region of stress-responsive elements (MYB and MYC) may be linked to their function in defending sponge gourd plants against various stressors by increasing the overall enzyme activity and expression of LcB3 transcripts (TableS4).

3.7. Protein sequence and secondary structure analysis of identified B3 genes of Sponge gourd

B3 gene family proteins in Sponge gourd show predominantly coiled structure followed by beta-strand, alpha helix and beta-turn as seen in the Fig 8A (Table S5) This observation is highly correlated with existing B3 Protein structures from other plant species (Romanel et al., 2009).

3.8. Analysis of Protein phosphorylation of identified B3 genes of Sponge gourd

As can be seen from Fig. 8B, B3 superfamily proteins show phosphorylation. The phosphorylation is varied with some families showing up to 50% phosphorylation and others showing only 27% phosphorylation (Table S6).

3.9. Expression profiling and subcellular localisation of identified B3 genes

The expression pattern of all 52 LcB3 genes was analysed using RNA sequencing information obtained from different sponge gourd sections including leaf, root, flower, fruit, stem, tendril and shoot apex (Fig. 4). According to the study, the LcARF family showed increased levels of expression in many parts of the plant, including the flower, fruit, stem, tendril, and shoot apex. LcARF16 has no expression in male flowers and tendrils. In addition, LcARF3 and LcARF17 had almost comparable expression levels across all stages of development. Within the RAV family, LcRAV1 and LcRAV7 exhibited a downregulation in fruit development. LcRAV8 exhibited elevated expression levels across all stages of development. The remaining members displayed a modest level of expression. LcLAV1 and LcLAV4 exhibit enhanced expression in all plant organs developments, but LcLAV2 and LcLAV3 have suppressed expression in many plant structures, except for the tendril where no expression is seen. LcLAV5 exhibited downregulation only in the root and shoot apex, with no expression seen throughout other stages of growth. However, several members of the REM family showed increased expression in different plant sections, except for LcREM6, LcREM7, LcREM8, and LcREM11-14. as observed in Fig 9.

4. Discussion

Plant development is regulated by many transcription factors (TFs) that collaborate in a combinatorial manner. Multiple families of transcription factors have been widely discovered, including the plant-specific B3 transcription factor family. This family plays a crucial role in controlling many pathways associated with plant growth and stress responses.

Additionally, according to Swaminathan et al. (2008), B3 transcription factors (TFs) play an important role in several hormone-related signalling processes. The current research used the existing whole genome sequence of sponge gourd (Zhao et al. 2020) to identify all 52 B3 domain-containing proteins. This paper represents the first investigation into the identification and molecular characterization of B3 proteins in sponge gourd. The 52 B3 proteins were categorized into four subfamilies: ARF, LAV, RAV, and REM, according to the classification of Swaminathan et al. in 2008. Nevertheless, some publications have classified them into five separate subfamilies, with LAV being further subcategorized as ABI3/VP1 and HIS (Romanel et al., 2009). Significantly, the absence of ABI3 and HSI families in sponge gourds suggests unique evolutionary adaptations. The existence of multiple isoforms within the REM, ARF, and LAV families indicates functional diversity, potentially linked to specific developmental stages or responses to environmental stimuli (Kandeel et al., 2023). The LcREM class had the largest gene count, totalling 21, which was comparable to Pepper (70) (Wang et al. 2024) Camelina sativa (Kandeel et al. 2023), and citrus (38) (Liu et al. 2020) as determined by using various computational tools. Domain analysis is a crucial component of structural genomics that offers insights into the protein's evolution, function, and structure. Every B3 protein has a minimum of one B3 domain that exhibited a high degree of conservation across all B3 proteins. Significantly, individuals within the same group in the evolutionary tree had comparable quantities and varieties of domains, alongside the B3 domain. This implies the presence of structural preservation among members of the same group. The predominantly globular nature of these proteins, as indicated by GRAVY scores, suggests their involvement in various cellular processes rather than membrane association. This information is crucial for predicting protein function and interactions. Comparative examination of gene families

from various crop species has shown that individuals belonging to the same clade have identical motif structure, domain organization, and intron-exon structure. This indicates that they originated from a common ancestor and possess a conserved function. (Bhattacharjee et al., 2015; Li et al., 2017; Singh et al., 2014; Tang et al., 2016). The current investigation included the clustering of 52 LcB3 proteins into six distinct clades. Each clade was well supported by evidence such as comparable gene structure and the existence of conserved motifs. Past studies have shown that genes belonging to the REM subfamily tend to cluster together in the genome. This clustering phenomena has been seen in several species, such as citrus, grapes, rice, tobacco, Arabidopsis, and others. A similar situation was seen in sponge gourd, where the largest gene cluster was located on chromosome 7, that includes 7 REM genes (Fig. 3). In addition, it was found that many LcB3 genes were clustered together in different regions of the remaining chromosomes.

Gene duplication plays a crucial role as a primary driver in the evolutionary process of genes and gene families. Gene duplication events, including tandem and segmental duplications, play a significant role in the enlargement of a gene family (Cannon et al. 2004). The recorded instances of gene duplications across chromosomes underscore the evolutionary dynamics inherent in the B3 gene family. Regarding the duplication of LcB3 genes, it is evident that these genes mostly developed by tandem duplication, since their frequency was higher than that of segmental duplications (Fig. 4). In the field of genetics, the dN/dS substitution ratio is used to ascertain the kind of selection (positive, diversifying, or purifying) that is influencing a group of homologous protein-coding genes. A dN/dS ratio of 1 signifies positive selection, whereas dN/dS >1 suggests positive (diversifying) selection, and dN/dS <1 denotes negative selection (Zhou et al. 2010). The current investigation used the dN/dS ratio to ascertain the rate of evolution in homologous protein-coding B3 genes. It was noted that the B3 genes undergo purifying selection, as seen by all gene pairs having a dN/dS ratio of less than 1 (Online Resource 8). The process of purifying selection ensures long-term stability by removing harmful mutations (Sironi et al., 2015). Thus, the significance of the B3 gene family in the growth of sponge gourd plants has made it necessary to

preserve and protect its members. Examination of the intron-exon structure of the B3 genes revealed notable discrepancies in the quantity and size of introns, indicating that these genes may have undergone intron loss or intron gain during the process of evolution. Three B3 genes without introns were detected in sponge gourd (Fig. 5). These genes may have undergone intron loss in order to develop efficient and rapid transcription regulatory systems (Deutsch and Long 1999). A protein motif refers to a brief sequence that is essential for the correct folding of proteins (Bailey et al., 2015). Significant levels of conservation were detected among the identified motifs within the B3 proteins of sponge gourd. Notably, individuals belonging to the same clade in the phylogenetic tree were discovered to possess comparable motifs and also had similar motif structure, hence providing further confirmation for the phylogenetic study. In addition, this kind of similarity in motif patterns indicates the presence of structural and functional redundancy among members of the same group. The multiple sequence alignments revealed the presence of conserved amino acid residues in the LcB3 family. (Fig.6). However, the REM subgroup exhibited a significant degree of variation. Specifically, we observed that the amino acid sequences unchanged within each subfamily (with the exception of the REM subfamily). As an example, conserved amino acid residues in the RAV subfamily were WNSSQ. In contrast, the comparable residues in the LAV, and ARF subfamilies were WPNNN, and RGQPR, respectively. This study confirms and expands upon other studies in Arabidopsis (Swaminathan et al., 2008), and amino acid residues castor bean were shown to be largely conserved within each subfamily (Wang et al., 2021). Results from the study highlight the predominance of coiled structures in B3 gene family proteins aligned with known structures in other plant species. The variable phosphorylation levels among B3 gene families indicate potential differences in post-translational regulation. The significance of phosphorylation in modulating protein activity and stability underscores the complexity of regulatory networks involving B3 proteins. The diverse expression patterns across plant parts suggest specific roles of B3 gene family members in different tissues. The nuclear localization of most B3 proteins aligns with their function as transcription factors. Variations in expression levels provide clues about the regulatory networks these genes participate in during plant development.

5. Conclusion

In the current study reveals the intricate landscape of B3 DNA-Binding superfamily genes in sponge gourds, revealing evolutionary patterns, gene architecture, and functional diversity. The absence of specific families indicates unique adaptations, while detailed analyses of structure and phosphorylation deepen our understanding of protein characteristics. Tissue-specific expression profiles suggest specific roles in plant development. Future research should explore the regulatory networks governing B3 genes, investigate potential interactions with other pathways, and uncover the ecological implications of their evolutionary dynamics. These insights hold promise for advancing agricultural practices and harnessing the pharmacological benefits of sponge gourds.

6. Conflicts of Interest

No conflict of interest.

7. Author Contributions

VS conducted the investigation, collected data, and wrote the manuscript. DG and SS helped to develop the topic, design, supervise, correct, and approve the text. SV, NT and NS assist in the analysis of the collected data.

8. Acknowledgement

Authors are thankful to the Shri Ramswaroop University, Lucknow-Deva Road, Barabanki, Uttar Pradesh for the providing the opportunity to accomplish this study.

9. References

1. Castillejo C, and Pelaz S,. The balance between Constant and Tempranillo activities determines FT expression to trigger flowering. Curr. Biol., 2008; 18 (17), 1338-1343.

2. Franco-Zorrilla, J.M.; Cubas, P.; Jarillo, J.A.; Fernandez-Calvin, B. and Salinas, J. AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. Plant Physiol. 2002; 128: 418–427.

3. Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, et al. Isolation of the Arabidopsis-Abi3 Gene by Positional Cloning. Plant Cell 1992; 4: 1251–1261.

4. Guilfoyle TJ, Hagen G. Auxin response factors. Curr Opin Plant Biol, 2007; 10 (5), 453-460.

5. Kabir N, Lin H, Kong X, Liu L, Qanmber G, Wang Y. Zhao, N. Identification, evolutionary analysis and functional diversification of RAV gene family in cotton (*G. hirsutum* L.). Planta, 2022; 255, 1-16.

6. Kagaya Y, Ohmiya K, Hattori T. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucl Acids Res 1999; 27: 470–478.

7. Kandeel M, Morsy MA, Abd El-Lateef H M, Marzok M, El-Beltagi HS, Al Khodair KM. and Venugopala KN. Genome-wide identification of B3 DNA-binding superfamily members (ABI, HIS, ARF, RVL, REM) and their involvement in stress responses and development in *Camelina sativa*. Agronomy, 2023; 13(3), 648.

8. Kwon S H, Lee BH, Kim EY, Seo YS, Lee S, Kim WT, and Kim JH. Overexpression of a *Brassica rapa* NGATHA gene in *Arabidopsis thaliana* negatively affects cell proliferation during lateral organ and root growth. Plant Cell Physiol. 2009; 50 (12), 2162-2173.

9. Lee BH, Kwon SH, Lee SJ, Park SK, Song JT, Lee S. and Kim JH. The Arabidopsis thaliana NGATHA transcription factors negatively regulate cell proliferation of lateral organs. Plant Mol Biol; 2015; 89, 529-538.

10. Levy YY, Mesnage S, Mylne JS, Gendall AR, and Dean C. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. Science; 2002; 297(5579), 243-246.

11. Liu Z, Ge XX, Wu XM, Xu Q, Atkinson RG, and Guo WW. Genome-wide analysis of the citrus B3 superfamily and their association with somatic embryogenesis. BMC Genomics; 2020; 21(1), 1-15.

12. Lu Q, Zhao L, Li D, Hao D, Zhan Y, and Li WA. GmRAV ortholog is involved in photoperiod and sucrose control of flowering time in soybean. PLoS One, 2014; 9 (2): e89145.

13. Monke G, Altschmied L, Tewes A, Reidt W, Mock HP, et al. Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta. 2004; 219: 158–166.

14. Parakkunnel R, Bhojaraja Naik K, Susmita C, Girimalla V, Bhaskar KU, Sripathy KV and Bhat KV. Evolution and co-evolution: Insights into the divergence of plant heat shock factor genes. Physiology and Molecular Biology of Plants, 2022; 28(5): 1029-1047.

15. Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, and Garcia-Mas J. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC genomics, 2011; 12(1): 1-14.

16. Romanel EA, Schrago CG, Couñago RM, Russo CA, and Alves-Ferreira M. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. PLOS one, 2009; *4*(6); e5791.

17. Shendge PN, Belemkar S. Therapeutic potential of *Luffa acutangula*: A review on its traditional uses, phytochemistry, pharmacology and toxicological aspects. Front Pharmacol. 2018; 9: 1177.

18. Shin HY, Nam KH. RAV1 negatively regulates seed development by directly repressing MINI3 and IKU2 in Arabidopsis. Mol Cells, 2018; 41(12): 1072.

19. Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, and Harada JJ. Leafy cotyledon2 encodes a B3 domain transcription factor that induces embryo development. PNAS; 98(20): 2001; 11806-11811.

20. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. Trends Plant Sci. 2008; 13: 647–55

21. Sung S. Amasino RM. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. Nature; 2004; 427(6970): 159-164.

22. Suzuki M, Kao CY, McCarty DR. The Conserved B3 Domain of VIVIPAROUS1 Has a Cooperative DNA Binding Activity. Plant Cell 1997; 9: 799–807.

23. Suzuki M, Wang HHY, McCarty DR. Repression of the LEAFY COTYLEDON 1/B3 Regulatory Network in Plant Embryo Development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 Genes. Plant Physiol 2007; 143: 902–911.

24. Tyagi R, Sharma V, Sureja AK, Das Munshi A, Arya L, Saha D, and Verma M. Genetic diversity and population structure detection in sponge gourd (*Luffa cylindrica*) using ISSR, SCoT and morphological markers. Physiol Mol Biol Plants, 2020; 26: 119-131.

25. Ulmasov T, Hagen G, Guilfoyle TJ. Activation and repression of transcription by auxin-response factors. PNAS. 1999; 96: 5844–5849.

26. Ulmasov T, Murfett J, Hagen G, Guilfoyle TJ. Aux/IAA Proteins Repress Expression of Reporter Genes Containing Natural and Highly Active Synthetic Auxin Response Elements. Plant Cell. 1997; 9: 1963–1971.

27. Wang WB, Ao T, Zhang YY, Wu D, Xu W, Han B, Liu AZ (2021) Genome-wide analysis of the B3 transcription factors reveals that RcABI3/VP1 subfamily plays important roles in seed development and oil storage in castor bean (*Ricinus communis*). Plant Divers 2021; 44: 201-212.

28. Wang T, Long C, Chang M.et al. Genome-wide identification of the *B3* transcription factor family in pepper (*Capsicum annuum*) and expression patterns during fruit ripening. Sci Rep. 2024; 14: 2226.

29. Waltner JK, Peterson FC, Lytle BL, Volkman BF. Structure of the B3 domain from Arabidopsis thaliana protein At1g16640. Protein Sci, 2005; 14(9): 2478-2483.

30. Wu H, Zhao G, Gong H. A high-quality sponge gourd (*Luffa cylindrica*) genome. *Hortic Res* 2020; **7**: 128.

10. Figures Captions

Figure 1 NCNI-CD search of B3 gene superfamily. Different domain represents by different colours.

Figure. 2: A phylogenetic tree of the LcB3 gene family. The four subfamilies are visually depicted using distinct coloured lines: RAV is denoted by the green, REM by the peach, LAV by the dark pink, and ARF by the sky-blue. The six subgroups are denoted by distinct colour schemes.

Figure 3. Chromosomal Localization of identified B3 gene on 13 chromosomes.

Figure. 4. Gene duplications in sponge gourd. The dark pink lines represent duplicated genes of LcB3 genes. The black lines represent the LcB3 genes as the background. Different chromosomes are represented by the blocks in yellow colour.

Figure 5: Conserved motif for LcB3 proteins, and gene structures for LcB3 genes (A) UTRs are prominently highlighted in pink, exons denoted by blue, and black lines represent introns of gene structures. (B)The 15 domains under investigation are visually distinguished by employing a range of colours, each representing a unique domain.

Figure.6: Multiple sequence alignment of LcB3 proteins with the B3 proteins of Arabidopsis and citrus. (6A) The consensus sequence for LcARF is RGQPK/RR and (6B) LcRAV class

consensus sequence is WN/RSSQS. (6C) The LAV class had WPNNKSR as the consensus sequence. (6D) Multiple sequence of REM class.

Figure 7. Cis-acting regulatory elements of B3 gene superfamilty

Figure 8 A) Secondary Structure of B3 gene superfamily B) Protein phosphorylation of identified LcB3 genes.

Figure 9 Expression profiling of B3 gene superfamily in different tissues developments.