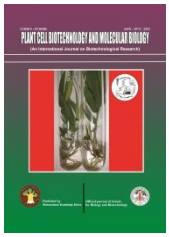


Plant Cell Biotechnology and Molecular Biology

Volume 25, Issue 11-12, Page 84-106, 2024; Article no.PCBMB.12450
ISSN: 0972-2025

In silico Analysis and Characterization of the F5H Gene Family in *Sorghum bicolor* and Its Role in Lignin Production

Anumandla Vinod Kumar [a], Prashanth B [a], Vikas Chandra [b] and Prashant Singam [a*]

[a] Department of Genetics, Osmania University, Hyderabad – 500 007, Telangana, India.
[b] Department of Biotechnology, School of Interdisciplinary Education and Research, Guru Ghasidas Vishwavidyalaya, Bilaspur - 495 009, Chattisgarh, India.

Authors' contributions

This work was carried out in collaboration between all authors. Author AVK designed the study, performed the bioinformatics analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors PS and PB managed the analyses and review of the study. Author VC managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.56557/pcbmb/2024/v25i11-128885

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://prh.ikprress.org/review-history/12450

Original Research Article

Received: 01/08/2024
Accepted: 03/10/2024
Published: 08/10/2024

ABSTRACT

Ferulate 5-hydroxylase (*F5H*), a cytochrome P450-dependent monooxygenase, catalyzes the hydroxylation of coniferaldehyde, a crucial step in the formation of syringyl lignin monomer (S). However, evolutionary divergence, expression patterns under abiotic stress conditions (ABA, PEG and NaOH) and lignin content-related features of the *F5H* gene family in *Sorghum bicolor* have not

**Corresponding author: E-mail: prashantsingam@gmail.com;*

Cite as: *Kumar, Anumandla Vinod, Prashanth B, Vikas Chandra, and Prashant Singam. 2024. "In Silico Analysis and Characterization of the F5H Gene Family in Sorghum Bicolor and Its Role in Lignin Production". PLANT CELL BIOTECHNOLOGY AND MOLECULAR BIOLOGY 25 (11-12):84-106. https://doi.org/10.56557/pcbmb/2024/v25i11-128885.*

been explored. This study envisaged mining of Sorghum genomic data leading to the identification of 61 *SbF5H* genes. Bioinformatics analysis revealed the phylogenetic evolutionary relations, gene structures, conserved motifs, physicochemical properties, and promoter cis-acting elements related to these genes and their encoded proteins. Based on the gene structural and phylogenetic features, these 61 *SbF5Hs* were grouped into 4 subclasses. The *in silico* expression analysis revealed higher accumulation of *SbF5H*1 transcripts in embryo and in root under stress conditions. Similarly, Other *SbF5H* genes have shown expression in stem and root, thus indicating *SbF5H* genes involvement in Sorghum lignin biosynthesis. By exploring into the functional aspects of the *F5H* gene, our study sought to shed light on its significance in influencing not only the chemical makeup of lignin but also the resultant plant phenotypes. This insight into the molecular mechanisms governing lignin biosynthesis can have implications for bioenergy production and crop improvement.

Keywords: Ferulate 5-hydroxylase; sorghum bicolor; phylogeny; biomass; lignin biosynthesis.

1. INTRODUCTION

Lignin biosynthesis is a complex process in plants and is essential for structural integrity, to withstand mechanical stress, and to defend against pathogens. Lignin biosynthesis primarily occurs through the phenylpropanoid pathway, starting with the conversion of phenylalanine to cinnamic acid via phenylalanine ammonia-lyase (PAL) [1,2,3,4]. In recent times, there has been significant research focus on manipulating lignin structures to enhance the extraction and transformation of lignin into various aromatic products [5,6,7]. The Gramineae family, comprising monocotyledonous grasses, encompasses vital food crops like rice, maize, wheat, and sorghum. As these crops generate biomass as agricultural by-products, there is the potential to establish new avenues for biomass-based energy and material production [8]. Though lignin bioengineering has been extensively studied in dicotyledonous model species like *Arabidopsis thaliana* and *Poplar*, there is limited information on lignin engineering in monocotyledonous grass species. In this regard, our research group has been specifically exploring lignin engineering in Sorghum (*Sorghum bicolor*), which serves as a notable model grass and holds commercial significance as a crop (Paterson, et al. 2009). In grasses, the composition of lignin's typically revolves around two primary units, guaiacyl (G) and syringyl (S). In contrast to biopolymers like nucleic acids and proteins, which undergo template-directed synthesis within the cell, the formation of the lignin polymer occurs through chemical reaction where different monomer units combine in a random manner to form a polymer. These reactions link monolignols to the developing polymer in the apoplast [9]. These units are formed through the oxidative coupling of two specific monolignols coniferyl alcohol for guaiacyl and sinapyl alcohol for syringyl. Furthermore, there is a lesser presence of p-hydroxyphenyl (H) units, and these are derived from p-coumaryl alcohol. This chemical arrangement and the varying proportions of these units contribute to the unique structure and properties of lignin in grasses, impacting their potential applications and behavior in different contexts [1,2,4]. Ferulate 5-hydroxylase (F5H) is a cytochrome P450-dependent monooxygenase, that catalyzes the hydroxylation of coniferaldehyde, a crucial step in the formation of one of the three main monolignols i.e., sinapyl alcohol. The current investigation of the sorghum *F5H* gene family provides valuable data for use in gene function analysis towards improving *Sorghum* biomass quality and reducing *Sorghum* cultivation costs. This investigation also helps on the functional significance of the *F5H* gene in determining lignin composition and properties in *Sorghum*. As future prospects researchers can examine how genetic modification in *F5H* expression or activity impacts lignin content, and lignin monomer composition (e.g., S/G (Syringyl and Guaiacyl) ratio), subsequently affecting the plant's cell wall structure, properties to improve biomass utilization, forage quality, or biofuel production efficiency. Overall, our study helps in the understanding of the SbF5H gene family in Sorghum bicolor by setting the base for future research intended for lignin biosynthesis, sustainable agriculture and bioenergy.

2. MATERIALS AND METHODS

Identification of *F5H* genes in *Sorghum bicolor*: With the help of protein sequences of model plant *A. thaliana* as the reference sequence, homologous *F5H* sequences within the *Sorghum bicolor* genome (*S. bicolor* v3.0) were detected based on results of BLASTP sequence alignments (using an e value cut off of

10^{-10}) against the Phytozome database (https://phytozome-next.jgi.doe.gov/). The presence of the *F5H* conserved domain was confirmed using the NCBI-CDD search (http://www.ncbi.nlm.nih.gov/Structure/cdd/ - assessed on 30 January 2024) online tool.

Analysis of sequence and characterization of *SbF5H* genes (Multiple sequence alignment, Conserved Domain, Gene structure and Phylogenetic Analysis): The Circos imagining method was employed to depict the sequence similarity between the *SbF5H* gene family and its orthologs in *A. thaliana*, *O. sativa*, and *Z. Mays* (https://bat.infspire.org/circoletto/_- assessed on 22 February 2024). Then, *F5H* protein sequences were first aligned using ClustalW, a plugin that is provided with MEGA 7.0. Next, a phylogenetic tree was built using MEGA 7.0 with the neighbour-joining (NJ) method and with 1000 bootstrap test. The exon/intron configurations of *Sorghum**F5H* genes were analysed to understand evolutionary divergency using the TBtools (https://github.com/CJ-Chen/TBtools__- assessed on 29 February 2024) and GTF annotation files [10]. The conserved motifs of *SbF5H* proteins were analysed using the online MEME website (http://meme-suite.org/tools/meme) (Malavasi, U. C., Davis, A. S., &Malavasi, M. D. M. [11], Maury, S., Geoffroy, P., & Legrand, M. [12], with the parameter defining the maximum number of identified motifs set to 10. The results were visualized using the TB tool.

Analysis of ProtParam, Sub-Cellular localization and PPI Network of *F5H* Proteins of *Sorghum bicolor:* The online program ExPAsy (https://web.expasy.org/protparam) was used to predict the basic characteristics of *SbF5H* proteins from their sequences [13]. Wolfpsort (https://wolfpsort.hgc.jp/results/pAA52df14345167a446b2a4275373bf2842.html) was used to predict subcellular localization sites of *F5H* proteins [14]. Meanwhile, the Protein–protein interaction (PPI) network of *Sorghum bicolorF5H* proteins were constructed and then used to predict the PPI network of *SBF5H* proteins using the STRING database (https://cn.string-db.org/_- assessed on 15 March 2024).

***SbF5H* proteins 3D Structures prediction and Ramachandran plot analysis:** To analyse the three-dimensional structures of *SbF5H* homologs, we utilized two computational tools: the SOPMA server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html_- assessed on 30 December 2023) and the SWISS-MODEL server. Following the structure modelling, Ramachandran plot was generated using the PROCHECK server. Additionally, the stability of these 3D structures was verified through the online PSVS server. Furthermore, we delved into the Protein-Protein interactions of *SbF5H* using the STRING database. This comprehensive approach allowed us to assess both the structural characteristics and potential interactions of *SbF5H* homologs, providing valuable insights into their functional roles and relationships in biological processes.

Chromosome location / physical mapping, Synteny, and Ka/Ks analysis of *SbF5H* homologs: The anticipated homologs of *SbF5H* were positioned on the sorghum genome utilizing TBtools and GTF annotation files. By using MCScanX software (TB tools - assessed on 20 June 2024) collinearity of *F5H*gene was analysed between *Sorghum and Arabidopsis, Sorghum and Zea mays*, as well as Sorghum and *Oryza sativa*. The Ka/Ks Calculator, employing the GLWL model, was applied to calculate Ka (nonsynonymous substitutions), Ks (synonymous substitutions), and the Ka/Ks ratio. Notably, a p-value of 0.05 was considered significant and retained in the analysis.

Analysis of cis-acting elements in the promotor regions of *SbF5H* genes: Cis-acting elements of *SbF5H* gene promoters were analysed using TBtools [15]; then, the 1.5 kb upstream region of the *SbF5H* gene's protein-coding sequence (CDS) was extracted from the sorghum genome using the TBtools and saved as a FASTA file. The Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/_____- assessed on 28 June 2024) web server was used to predict cis-acting elements [16], and then the results were visualized using the TBtools.

Prediction of transcription factors and miRNA of *SbF5H* genes: In our research, we predicted the transcription factors that interact with *SbF5H* genes by utilizing the Plant Transcription Factor Database (Plant TFD - assessed on 19 February 2024). Additionally, we examined miRNA binding sites associated with these genes through the Plant RNA prediction tool. To comprehensively understand the regulatory networks involved, we visualized the interactions between transcription factors and miRNAs using the Cytoscape tool. This approach allows us to explore and interpret

the complex relationships and regulatory mechanisms involving *SbF5H* genes, transcription factors, and miRNAs within the biological context.

Gene ontology study: In our study, we subjected the homologs of *SbF5H* to a comprehensive Gene Ontology enrichment analysis, employing the Shiny GO online tool. This analysis allowed us to gain insights into the functional annotations and biological roles associated with these homologous genes. Additionally, we anticipated the co-expression patterns of *SbF5H* homologs by leveraging the Sorghum Functional Genomics Database (assessed on 28 July 2024). Co-expression analysis provides valuable information about genes that are likely to be functionally related or involved in common biological processes. To visualize and interpret the complex relationships inferred from the co-expression analysis, we constructed a co-expression network using the Cytoscape tool. This network representation offers a graphical depiction of the interactions and associations among *SbF5H* homologs, facilitating a more comprehensive understanding of their functional relationships within the context of sorghum functional genomics.

Analysis of expression patterns of *SbF5H* genes: The *Sorghum* transcriptome data FPKM (Fragments Per Kilobase of transcript per Million mapped reads) was obtained from the *Sorghum* functional genomic data base (http://structuralbiology.cau.edu.cn/sorghum/gene_detail.php?gene=Sobic.001G196300) and Gramene database (https://www.gramene.org/__- assessed on 05 June 2024). This dataset encompasses the baseline expression of *SbF5H*s across diverse organs of *Sorghum*, including both vascular and non-vascular tissues, stem internodes of *Sorghum*, and the expression patterns in shoot and root tissues under various stress conditions. The visualization of these expression patterns was accomplished through the creation of a heatmap using TB tools.

3. RESULTS

Identification and fundamental data of *Sorghum bicolor F5H* genes: The Phytozome database's sequences and the reference *F5H* protein sequences of A. thaliana were compared using BLASTP, and the results showed that the sorghum genome encodes homologous *F5H* protein sequences. Then, the conserved domain of *SbF5Hs* was confirmed by the CDD database (Table 1). Ultimately, 61 *SbF5Hs*genes were identified based on E value (E < 1e-5) (Fang et al. 2022). All *SbF5Hs* contained the conserved P450 domain (except *SbF5H 10, SbF5H31 and SbF5H 53*), thus indicating thatthe*SbF5Hs* belong to the P450 superfamily. Based on previously reported findings and genome assembly scaffold data, the *SbF*5H genes were named in sequential order (*SbF5H1* to *SbF5H61*).

The ExPASy website online tools were used to analyse the physicochemical properties of predicted SbF5H proteins (Table 2). The *SbF5H* sequences ranged from 461 (*SbF5H31*) to 579 (*SbF5H15*) amino acids (aa), with an average length of 524 aa. The molecular weights (MWs) ranged from 50.907KDa (*SbF5H31*) to 64.013KDa (*SbF5H15*), with an average MW of 58.371KDa. The theoretical pI values ranged from5.95 (*SbF5H14*) to 9.43 (*SbF5H44*), with an average pI of 7.68. Overall, the *SbF5H*family proteins had pI values that were <9.5, thus indicating they are neutral or slightly acidic proteins. Moreover, grand average of hydropathy (GRAVY) values were also examined (Table 2). An analysis of the 61*SbF5Hs* using the online WoLFPSORT tool, which is used to predict subcellular localization of proteins, predicted that the sixty-one *SbF5Hs*genes show chloroplast localization, extracellular, vacuolar, endoplasmic reticulum, mitochondrial and plasma membrane (PM) localization. However, most of them are localized on Chloroplast.

Analysis of sequence and characterization of *SbF5H* genes (Phylogenetic Analysis, Multiple sequence alignment, Conserved Domain and Gene structure): Multiple sequence alignment of *SbF5H* homologs were performed using ClustalW. The sequence similarity between *SbF5H* gene family and orthologs in *A. thaliana, O. sativa* and *Z. mays* were visualised by Circos (https://bat.infspire.org/circoletto/).The results of the multiple sequence alignment analysis were visualised by Circostool, showed that the protein homology of the *SbF5H* proteins ranged from 50 to 99% (Fig. 4). *SorghumF5H*genes displayed 99% similarity with *O. sativa, Z. mays F5H* genes and 50 to 75% similarity with Arabidopsis *F5H* genes. Fig. 7 also shows that the *Sorghum bicolorF5H* amino acid sequences share high degrees of identity with the *F5H* sequences of other plants. For example, *SbF5Hs* 1–61 share identity with *OsF5H1, ZMF5H1 and ATF5H1.*

The evolutionary tree was constructed using the complete protein sequences of 61 *SbF5Hs,* top 5 *AtF5Hs*, top 5 *OsF5Hs*, and top 5 *ZmF5Hs* (Fig. 1). All these F5Hs grouped into four clades (I, II, III and IV) based on evolutionary relationship analysis. Our analysis revealed that *F5Hs* from these four species can be categorized into four distinct groups: I, II, III, and IV. Group I comprise only one *Oryza sativa* gene (*OsF5H4*), which is shared with 26*Sorghum bicolorF5Hs*. Group II consists of 8*Sorghum bicolorF5Hs* genes (*SbF5H18, SbF5H53, SbF5H43, SbF5H51, SbF5H44, SbF5H48, SbF5H19, and SbF5H21*) exclusively, without sharing with other species. Group III includes 5*Zea maysF5Hs* (ZmF5H 1 - 5), 5 *Arabidopsis thalianaF5Hs* (*AtF5H* 1 - 5), 3 *Oryza sativaF5Hs* (*OsF5H1, OsF5H3, and OsF5H5*), along with various *Sorghum bicolorF5H* genes. Group IV harbours 1 *Oryza sativaF5H* (*OsF5H2*) and 1*Sorghum bicolorF5H* (*SbF5H26*) exclusively. Upon examining this distribution, it becomes apparent that group I contains only one Oryza sativa gene (*OsF5H4*) and 26 sorghum bicolor SbF5H genes while Group II exclusively containsSbF5H genes. Group IV is characterized by the exclusive presence of 1 *Oryza sativa* (*F5H2*) and 1 *Sorghum bicolor* (*F5H26*) gene. *SbF5H*, *ZmF5H, AtF5H*and *OsF5H* genes are dispersed throughout group III. In summary, the phylogenetic tree illustrates the evolutionary consistency of *F5Hs* across different species.

SbF5Hs and other plant *F5Hs* (ZmF5H, AtF5H and OsF5H) are highly homologous and contain the cytochrome p450domain (PFGSGRRSCPG), wherein a conserved cysteine acts as a heme-binding ligand, thus providing further evidence that *SbF5Hs* belong to the cytochrome P450 super family (Fig. 3c). The complex range of differences in gene structures is crucial for the development of various gene families. This enables the formation of specific groups of genes, each with its own functions and traits. To analyse the *SbF5Hs* gene structures, we analysed their DNA sequences to determine their intron and exon compositions and overall length (Fig. 2). The results of this analysis revealed significant differences in the architecture of the *Sorghum F5H* gene family members. Most *SbF5H* genes typically consist of two exons (Fig. 2). However, some specific genes such as *SbF5H* genes 11, 12, 15, 26, 35, 37, 38, and 41 have only one exon. Conversely, *SbF5H* genes 16, 17, 19, 21, 24, 33, 50, 52, 59, and 60 each comprise three exons. Notably, *SbF5H* gene 13 has the highest number of exons (four in total). Among all the *SbF5H* genes (from *SbF5H1 to SbF5H61*), *SbF5H* gene 20 is the longest, while *SbF5H* genes 35 and 37 are the shortest and lack UTR sequences.

Thereafter, in our analysis, we employed the MEME online tool (http://meme-suite.org/) to predict protein motifs and delve into the structural characteristics of *SbF5H* proteins. Through this conserved motif analysis, a maximum of 10 conserved motifs within the *SbF5H* proteins was identified (Fig. 3a). Each motif, labelled from 1 to 10, was found to be distinct. Clusters of proteins within the same group displayed similar motif compositions.

Notably, all the predicted *SbF5H* proteins were found to possess motifs 1-10, with the exception of *SbF5H5, SbF5H6, SbF5H12, SbF5H19, SbF5H27, SbF5H49 and SbF5H50*. Also, the *SbF5H29*has only motif 3 and 5. This comprehensive motif analysis provides insights into the conserved structural features of *SbF5H* proteins and highlights distinct motif patterns within different groups of these proteins.

Homology modelling, prediction of protein 2D, 3D structures, PPI analysis, and Ramachandran Plot of *SbF5H* Homologs: The predicted 2D structure of *SbF5H1* homolog consists of more amount of alpha helix (46 %) followed by Random coil (36%) than extended strand and beta turn (Fig. S1 and S2). The structural characteristics observed in the *SbF5H1* protein may contribute to its ability to perform a range of functions and adapt to various structural configurations.

Producing the three-dimensional structure of a protein is crucial for bridging the gap between its sequence and its actual configuration. Constructing 3D models provides valuable insights into the structure, function, localization, and interaction network of the protein. The models with high confidence and identity percentage were selected. All the predicted 3D structures of *SbF5H* (*SbF5H1 – SbF5H 61*) showed significant similarity with previous studied proteins (as shown in Fig. S3).

A key analysis conducted was the Ramachandran plot, The results showed that over 90% of the protein's residues fell within the favourable region of the Ramachandran plot, indicating high structural integrity (Fig. 6). This suggests that the backbone dihedral angles are mostly within acceptable ranges, affirming the accuracy of the generated models. These

findings instil confidence in the reliability of the predicted 3D model, crucial for advancing research in understanding biological mechanisms of *F5H* gene.

The PPI analysis of *SbF5H* homologs were analysed (Fig. 5a). Among the all *SbF5H* homologs, *SbF5H1* exhibited interactions with Cinnamyl alcohol dehydrogenase (Fig. 5b), one of the major enzymes involved in lignin biosynthesis of sorghum. Thus, the string analysis of *sorghum F5H* protein exhibited the interaction with various lignin biosynthetic pathway and secondary metabolites (Fig. 5 a and b).

Chromosome localization/physical mapping, synteny, and ka/ks analysis of *SbF5H* homologs: The chromosomal location of the *SbF5Hs* genes were explored to analyse how they are located within the genome. A total of 61 genes were mapped in accordance with their position in the chromosome. The chromosomal positions of *SbF5H* genes across the ten chromosomes of the sorghum genome were depicted using TB tools software and GTF annotation files (Fig. 7). All 61 *SbF5H* genes were found distributed across chromosomes 1 to 10. The majority of *SbF5H* genes (15 genes) were located on chromosome 1, with 9 genes on chromosome 2, while the remaining *F5H* genes were dispersed across the other chromosomes. Our analysis reveals an uneven distribution of *SbF5H* genes across sorghum chromosomes, with approximately 75% clustered near the chromosome centres. Notably, chromosomes 1 and 2 emerged as hotspots for *SbF5H* homologs that is identified with chromosomal location mapping analysis.

Additionally, we explored the collinearity of *SbF5H* genes. Comparative analysis was conducted between *SbF5H* proteins and homologs from three other related plants (*Arabidopsis thaliana* (Fig. 8a), *Oryza sativa* (Fig. 8b), and *Zea mays* (Fig. 8c) to explore deeper into the convergence of the *F5H* gene family across different species. The findings revealed that *SbF5H* proteins exhibited homology with both *Oryza sativa* and *Zea mays* species, while no homologs were identified in *Arabidopsis thaliana*. This suggests that the *F5H* gene family in sorghum exhibits stronger collinearity with monocots (*Oryza sativa and Zea mays*) but lacks convergence with Eudicots (*Arabidopsis*), indicating evolutionary divergence within monocots.

The Ka/Ks ratio, indicating the substitution rate ratio, was determined for the 61 *SbF5H* genes. In each of the 17 interspecific duplicated gene pairs, all *SbF5H* genes displayed a Ka/Ks ratio of less than one (Table 3). This indicates that all segmentally duplicated *SbF5H* genes have undergone purifying selection [17].

Exploration of Cis-acting elements in the promoter region of *SbF5H* genes: We employed the Plant Care website to delve into the cis-acting elements within the sorghum *SbF5H* gene family and their potential regulatory functions. This analysis enabled us to anticipate and scrutinize the cis-regulatory elements present in the 1500 bp nucleotide sequences upstream of the transcriptional start site of the *SbF5H* genes (Fig. 3b). These cis-acting elements, located in the promoter region of *SbF5H* genes, serve vital roles in responding to diverse developmental, biotic, and abiotic signals, thereby modulating the expression of downstream genes. Our study identified four primary classes of cis-acting elements in the promoter region of the *SbF5H* gene family: those associated with development and growth, light responsiveness, stress responsiveness, and hormone responsiveness. This suggests the existence of a sophisticated regulatory network governing the expression of *SbF5H* genes in response to a broad spectrum of environmental and developmental cues. The Methyl jasmine responsive elements (MeJRE) and Light responsive elements (LRE) were detected in most of the *SbF5H*s genes. It indicates that these genes may involve in plant defence responses. We observed the presence of gibberellic acid responsive elements (GARE) and auxin responsive elements (ARE), indicating their roles in growth and development. Additionally, the identification of Abscisic acid responsive elements (ABA RE), crucial for drought stress response, suggests the involvement of these genes in drought stress tolerance. Moreover, our analysis unveiled several other cis-regulatory elements within the promoter regions of *SbF5H* genes, such as MYB, SARE (Sialic acid responsive elements), LTR (Low temperature responsive elements), and WRE (Wound responsive elements), all of which are associated with lignin biosynthesis. These comprehensive findings highlight the potential regulation of *SbF5H* gene expression by a variety of stressors or external stimuli.

Prediction of transcription factors and miRNAs targeting *SbF5H* genes: The initiation

of transcription marks a pivotal phase in gene expression, with a critical event being the interaction between RNA polymerase and the promoter region. The composition and configuration of the promoter significantly influence the binding affinity of RNA polymerase, consequently impacting the extent of gene expression [13].

Table 1. Characteristics of *SbF5H* homologs

Transcript ID.	Gene name	Chromosome	Location start	Location End	Strand	CDS (bp)
Sobic.001G196300.1. v3.2	*SbF5H-1*	1	17685325	17690280	forward	1590
Sobic.010G106800.1. v3.2	*SbF5H-2*	10	10322497	10324579	forward	1560
Sobic.010G204200.1. v3.2	*SbF5H-3*	10	54643881	54645731	forward	1542
Sobic.010G203600.1. v3.2	*SbF5H-4*	10	54604663	54607246	forward	1539
Sobic.010G203500.1. v3.2	*SbF5H-5*	10	54600169	54602528	forward	1563
Sobic.003G009900.1. v3.2	*SbF5H-6*	3	862851	865801	reverse	1674
Sobic.003G010200.1. v3.2	*SbF5H-7*	3	879743	881711	reverse	1659
Sobic.003G010300.1. v3.2	*SbF5H-8*	3	884470	886126	reverse	1554
Sobic.003G010100.1. v3.2	*SbF5H-9*	3	876989	879450	reverse	1569
Sobic.003G009700.1. v3.2	*SbF5H-10*	3	854887	856782	reverse	1632
Sobic.009G064400.1. v3.2	*SbF5H-11*	9	6867135	6869496	reverse	1554
Sobic.009G142900.1. v3.2	*SbF5H-12*	9	50013542	50015736	reverse	1602
Sobic.005G228400.1. v3.2	*SbF5H-13*	5	71574181	71581342	reverse	1554
Sobic.005G064900.1. v3.2	*SbF5H-14*	5	7284316	7286285	reverse	1626
Sobic.005G158000.1. v3.2	*SbF5H-15*	5	63113875	63116031	reverse	1740
Sobic.005G127000.1. v3.2	*SbF5H-16*	5	55231045	55238054	reverse	1629
Sobic.005G206100.1. v3.2	*SbF5H-17*	5	69210090	69212342	forward	1632
Sobic.005G217500.1. v3.2	*SbF5H-18*	5	70333616	70335727	reverse	1578
Sobic.008G039300.1. v3.2	*SbF5H-19*	8	3699932	3702303	reverse	1533
Sobic.008G106200.1. v3.2	*SbF5H-20*	8	49976165	49984597	reverse	1551
Sobic.008G039501.1. v3.2	*SbF5H-21*	8	3818742	3823993	forward	1542
Sobic.008G105800.1. v3.2	*SbF5H-22*	8	49839492	49844969	reverse	1548
Sobic.008G107100.1. v3.2	*SbF5H-23*	8	50355638	50360525	forward	1554
Sobic.008G058500.1. v3.2	*SbF5H-24*	8	6145463	6147947	reverse	1671
Sobic.008G107200.1. v3.2	*SbF5H-25*	8	50404741	50407012	forward	1539
Sobic.004G007500.1. v3.2	*SbF5H-26*	4	665104	667038	forward	1512
Sobic.004G108800.1. v3.2	*SbF5H-27*	4	10493841	10495822	reverse	1623
Sobic.004G139300.1. v3.2	*SbF5H-28*	4	39976163	39978076	forward	1587
Sobic.004G068700.1. v3.2	*SbF5H-29*	4	5601967	5604203	forward	1533
Sobic.004G068600.1. v3.2	*SbF5H-30*	4	5584654	5587070	reverse	1563
Sobic.004G069700.1. v3.2	*SbF5H-31*	4	5655328	5657079	forward	1386
Sobic.004G068900.1. v3.2	*SbF5H-32*	4	5618765	5620768	forward	1605
Sobic.004G068800.1. v3.2	*SbF5H-33*	4	5612341	5614859	forward	1494
Sobic.007G003400.1. v3.2	*SbF5H-34*	7	315338	317447	forward	1659
Sobic.007G149000.1. v3.2	*SbF5H-35*	7	57980821	57982444	reverse	1623
Sobic.001G176300.1. v3.2	*SbF5H-36*	1	14789118	14791894	reverse	1539
Sobic.001G018600.1. v3.2	*SbF5H-37*	1	1559058	1560675	reverse	1617
Sobic.001G176000.1. v3.2	*SbF5H-38*	1	14754668	14756267	reverse	1599
Sobic.001G235500.1. v3.2	*SbF5H-39*	1	23744184	23746267	reverse	1524
Sobic.001G176600.1. v3.2	*SbF5H-40*	1	14850308	14852961	reverse	1551
Sobic.001G018300.1. v3.2	*SbF5H-41*	1	1538412	1540518	forward	1587
Sobic.001G362900.1. v3.2	*SbF5H-42*	1	65234965	65237598	forward	1593
Sobic.001G338900.1. v3.2	*SbF5H-43*	1	62661827	62664269	forward	1563
Sobic.001G128900.1. v3.2	*SbF5H-44*	1	10139545	10141822	forward	1548
Sobic.001G229500.1. v3.2	*SbF5H-45*	1	22159567	22163252	reverse	1533
Sobic.001G173100.1. v3.2	*SbF5H-46*	1	14518859	14521341	reverse	1560
Sobic.001G012000.1. v3.2	*SbF5H-47*	1	1031519	1034071	reverse	1698
Sobic.001G326400.1. v3.2	*SbF5H-48*	1	61353023	61355357	forward	1602
Sobic.001G176400.1. v3.2	*SbF5H-49*	1	14825098	14829135	reverse	1539
Sobic.006G044100.1. v3.2	*SbF5H-50*	6	30770149	30772258	reverse	1620
Sobic.006G010200.1. v3.2	*SbF5H-51*	6	1501662	1503498	reverse	1599
Sobic.006G043800.1. v3.2	*SbF5H-52*	6	30398856	30401232	reverse	1584
Sobic.002G090900.1. v3.2	*SbF5H-53*	2	9495657	9497800	reverse	1584
Sobic.002G273600.1. v3.2	*SbF5H-54*	2	65661886	65664257	forward	1521
Sobic.002G273800.1. v3.2	*SbF5H-55*	2	65668673	65671325	forward	1521
Sobic.002G190300.1. v3.2	*SbF5H-56*	2	57663874	57665873	forward	1527
Sobic.002G065800.1. v3.2	*SbF5H-57*	2	6495171	6497230	forward	1536
Sobic.002G065700.1. v3.2	*SbF5H-58*	2	6482062	6484042	forward	1542
Sobic.002G110100.1. v3.2	*SbF5H-59*	2	13386991	13389287	forward	1683
Sobic.002G110200.1. v3.2	*SbF5H-60*	2	13389688	13393034	reverse	1626
Sobic.002G273700.1. v3.2	*SbF5H-61*	2	65666390	65668673	forward	1518

Characteristics of Sorghum F5H genes; CDS-coding sequence(bp)

Table 2. Protparam analysis of *SbF5H* homologs

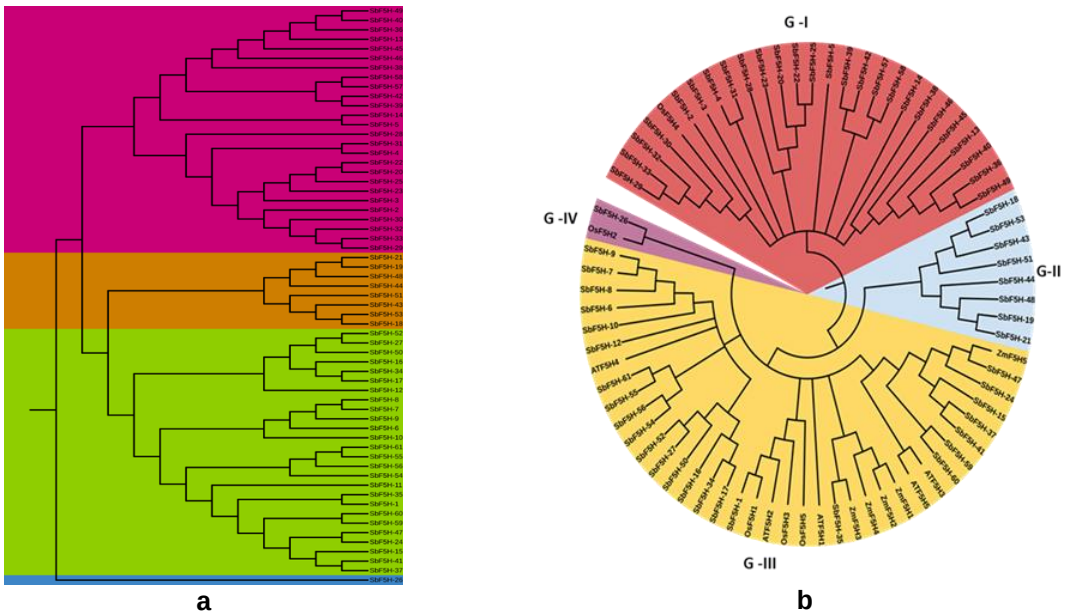
Gene name	Protein length (A.A)	MW (KDa)	pI	GRAVY Grand average of hydropathicity	No. of Exons / introns	Sub cellular localization
SbF5H-1	530	58.387	7.16	-0.055	2:1	Chloroplast
SbF5H-2	519	58.234	9.03	0.009	2:1	Extracellular
SbF5H-3	513	56.191	9.33	-0.053	2:1	Chloroplast
SbF5H-4	512	57.037	8.88	-0.039	2:1	Chloroplast
SbF5H-5	520	57.664	6.70	-0.088	2:1	Chloroplast
SbF5H-6	557	60.114	7.82	-0.031	2:1	Chloroplast
SbF5H-7	552	61.417	6.67	-0.126	2:1	Chloroplast
SbF5H-8	517	57.270	7.75	-0.044	2:1	Chloroplast
SbF5H-9	522	57.951	6.94	-0.109	2:1	Chloroplast
SbF5H-10	543	58.601	6.76	0.026	2:1	Chloroplast
SbF5H-11	517	57.573	6.42	-0.104	1:0	Chloroplast
SbF5H-12	533	58.356	6.20	-0.155	1:0	Chloroplast
SbF5H-13	517	58.727	8.82	-0.073	4:3	Chloroplast
SbF5H-14	541	58.575	5.95	0.040	2:1	Vacuolar
SbF5H-15	579	64.013	7.23	-0.158	1:0	Endoplasmic reticulum
SbF5H-16	542	60.945	7.08	-0.208	3:2	Chloroplast
SbF5H-17	543	61.131	7.71	-0.067	3:2	Chloroplast
SbF5H-18	525	59.154	8.42	-0.119	2:1	Chloroplast
SbF5H-19	510	56.615	7.71	-0.016	3:2	Chloroplast
SbF5H-20	516	57.542	6.91	0.038	2:1	Chloroplast
SbF5H-21	513	57.051	8.50	-0.050	3:2	Chloroplast
SbF5H-22	515	58.525	8.66	0.019	2:1	Chloroplast
SbF5H-23	517	58.028	7.22	-0.004	2:1	Chloroplast
SbF5H-24	556	62.535	8.86	-0.129	3:2	Chloroplast
SbF5H-25	512	57.156	8.82	-0.039	2:1	Chloroplast
SbF5H-26	503	55.438	6.04	-0.026	1:0	Chloroplast
SbF5H-27	540	59.214	6.51	0.0040	2:1	Chloroplast
SbF5H-28	528	59.275	6.67	-0.126	2:1	Chloroplast
SbF5H-29	510	56.951	7.19	-0.013	2:1	Chloroplast
SbF5H-30	520	58.101	8.76	-0.068	2:1	Chloroplast
SbF5H-31	461	50.907	6.35	-0.069	2:1	Mitochondrial
SbF5H-32	534	59.148	9.15	-0.037	2:1	Chloroplast
SbF5H-33	497	55.791	6.59	-0.067	3:2	Chloroplast
SbF5H-34	552	61.586	6.67	-0.226	2:1	Chloroplast
SbF5H-35	540	60.771	7.02	-0.112	1:0	Chloroplast
SbF5H-36	512	58.316	9.05	-0.100	2:1	Chloroplast
SbF5H-37	538	59.267	6.56	-0.070	1:0	Chloroplast
SbF5H-38	532	58.721	8.89	-0.112	1:0	Chloroplast
SbF5H-39	507	56.337	7.59	-0.003	2:1	Chloroplast
SbF5H-40	516	59.006	8.55	-0.115	2:1	Chloroplast
SbF5H-41	528	58.669	7.28	-0.105	1:0	Vacuolar
SbF5H-42	530	58.416	8.90	-0.051	2:1	Chloroplast
SbF5H-43	520	57.580	8.26	-0.149	2:1	Chloroplast
SbF5H-44	515	57.695	9.43	-0.132	2:1	Chloroplast
SbF5H-45	510	58.128	8.72	-0.081	2:1	Chloroplast
SbF5H-46	519	58.695	8.86	-0.07	2:1	Chloroplast
SbF5H-47	565	62.546	9.22	-0.243	2:1	Chloroplast
SbF5H-48	533	59.155	6.73	-0.107	2:1	Chloroplast
SbF5H-49	512	58.394	8.60	-0.058	2:1	Chloroplast
SbF5H-50	539	59.604	6.37	0.060	3:2	Chloroplast
SbF5H-51	532	58.556	7.11	-0.091	2:1	Chloroplast
SbF5H-52	527	57.957	6.08	0.070	3:1	Chloroplast
SbF5H-53	527	57.968	6.52	0.013	2:1	Chloroplast
SbF5H-54	506	55.274	7.77	-0.049	2:1	Chloroplast
SbF5H-55	506	55.272	6.59	-0.014	2:1	Chloroplast
SbF5H-56	508	55.788	6.66	-0.129	2:1	Chloroplast
SbF5H-57	511	56.501	8.86	0.029	2:1	Chloroplast
SbF5H-58	513	57.582	8.74	-0.103	2:1	Chloroplast
SbF5H-59	560	63.122	9.16	-0.056	3:2	Plasma membrane
SbF5H-60	541	61.014	9.23	-0.161	3:2	Chloroplast
SbF5H-61	505	55.122	6.30	0.029	2:1	Endoplasmic reticulum

Analysis of protein parameters of Sorghum F5H genes; protein length (amino acids), pI- isoelectric point, Sub-cellular localization and GRAVY- grand average of hydropathicity

Table 3. Duplicated *SbF5H* genes in Sorghum and history of their duplication

S. No.	Paralogous pairs of genes		Ka	Ks	Ka/Ks	Time How many million years ago (MYA) they duplicated
	Seq_1	Seq_2				
1	Sobic.001G176400 (SbF5H-49)	Sobic.001G176600 (SbF5H-40)	0.076	0.217	0.352	5.883
2	Sobic.002G065700 (SbF5H-58)	Sobic.002G065800 (SbF5H-57)	0.167	0.325	0.512	12.816
3	Sobic.001G362900 (SbF5H-42)	Sobic.001G235500 (SbF5H-39)	0.143	0.438	0.326	10.975
4	Sobic.005G064900 (SbF5H-14)	Sobic.010G203500 (SbF5H-05)	0.437	0.750	0.582	33.578
5	Sobic.004G069700 (SbF5H-31)	Sobic.010G203600 (SbF5H-04)	0.296	0.720	0.410	22.739
6	Sobic.008G105800 (SbF5H-22)	Sobic.008G106200 (SbF5H-20)	0.074	0.283	0.262	5.707
7	Sobic.004G068800 (SbF5H-33)	Sobic.004G068700 (SbF5H-29)	0.036	0.075	0.479	2.760
8	Sobic.008G039501 (SbF5H-21)	Sobic.008G039300 (SbF5H-19)	0.008	0.068	0.118	0.617
9	Sobic.002G090900 (SbF5H-53)	Sobic.005G217500 (SbF5H-18)	0.170	0.658	0.259	13.087
10	Sobic.006G043800 (SbF5H-52)	Sobic.004G108800 (SbF5H-27)	0.050	0.171	0.289	3.811
11	Sobic.007G003400 (SbF5H-34)	Sobic.005G206100 (SbF5H-17)	0.232	0.676	0.343	17.819
12	Sobic.003G010300 (SbF5H-08)	Sobic.003G010200 (SbF5H-07)	0.125	0.289	0.432	9.595
13	Sobic.002G273700 (SbF5H-61)	Sobic.002G273800 (SbF5H-55)	0.051	0.193	0.265	3.936
14	Sobic.007G149000 (SbF5H-35)	Sobic.001G196300 (SbF5H-01)	0.590	0.867	0.681	45.358
15	Sobic.002G110200 (SbF5H-60)	Sobic.002G110100 (SbF5H-59)	0.463	1.793	0.258	35.580
16	Sobic.001G012000 (SbF5H-47)	Sobic.008G058500 (SbF5H-24)	0.246	0.465	0.529	18.904
17	Sobic.001G018300 (SbF5H-41)	Sobic.001G018600 (SbF5H-37)	0.046	0.082	0.553	3.509

Ka- nonsynonymous; Ks- synonymous.



a **b**

Fig. 1a. Rectangular Phylogeny of *SbF5H*, 1b. Circular phylogeny of *SbF5H* with other species

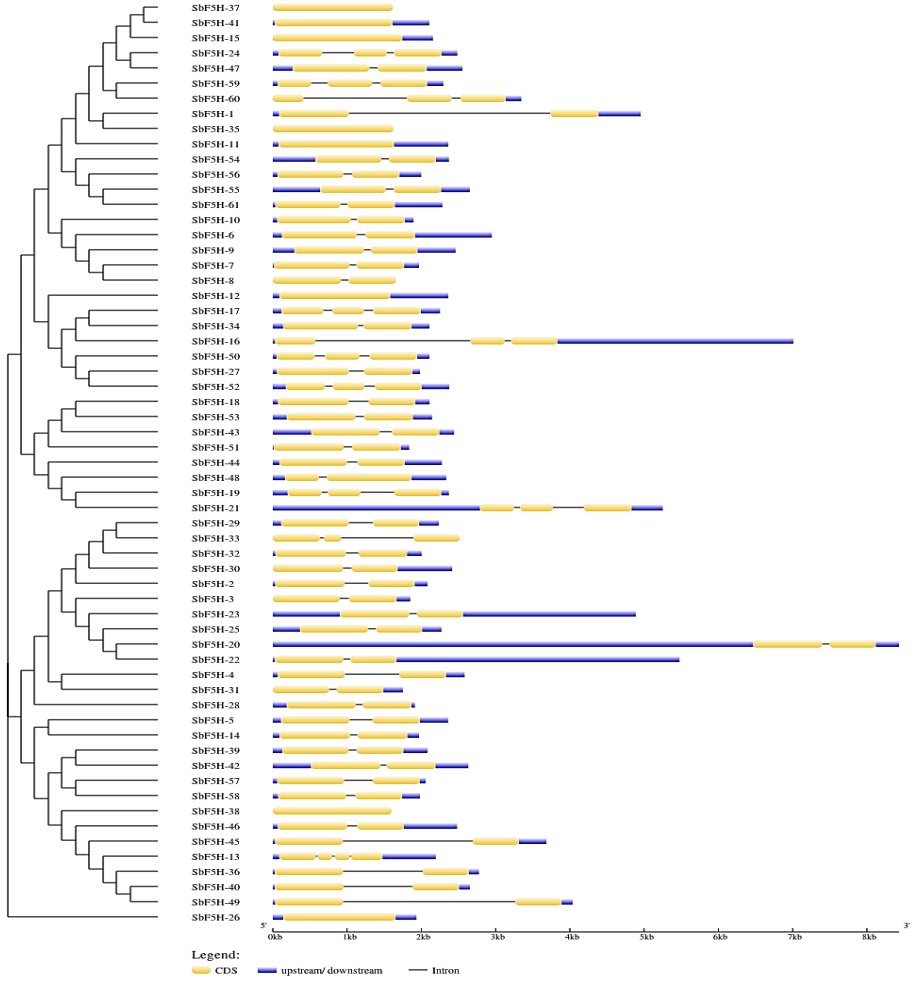


Fig. 2. Gene structure analysis of *of SbF5H genes*

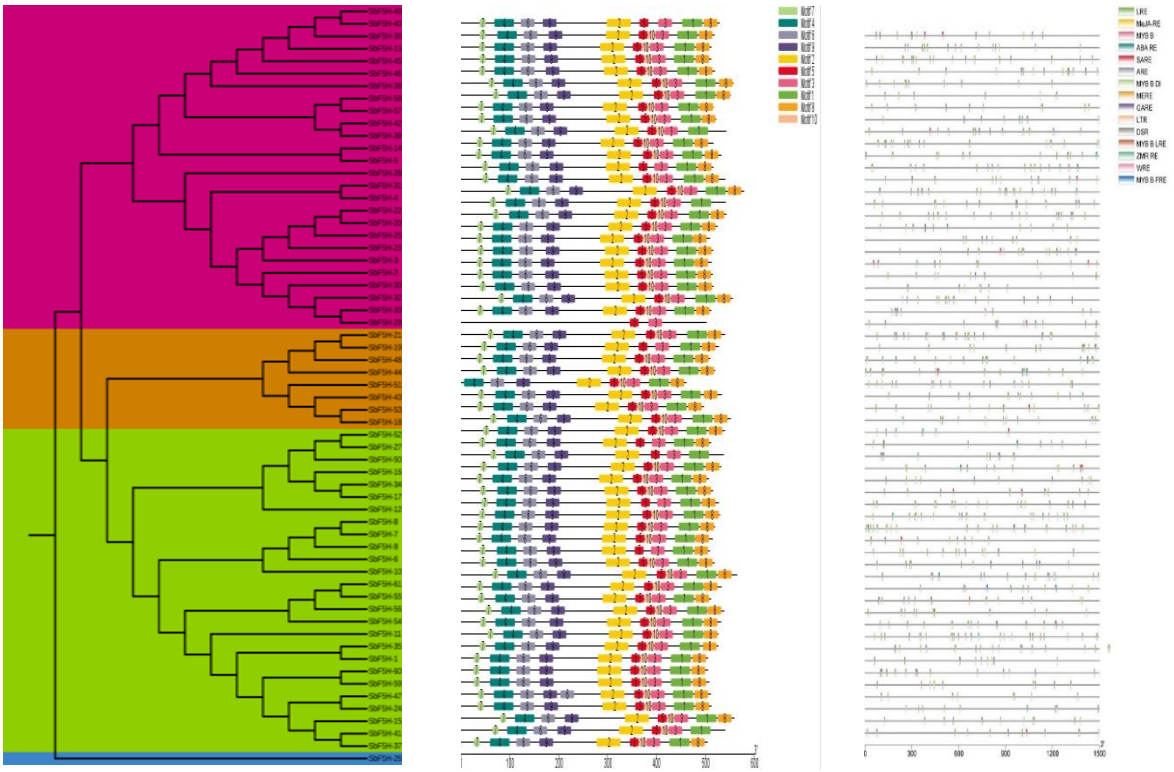


Fig. 3a. Phylogeny with Motifs Fig. 3b. Cis regulatory analysis

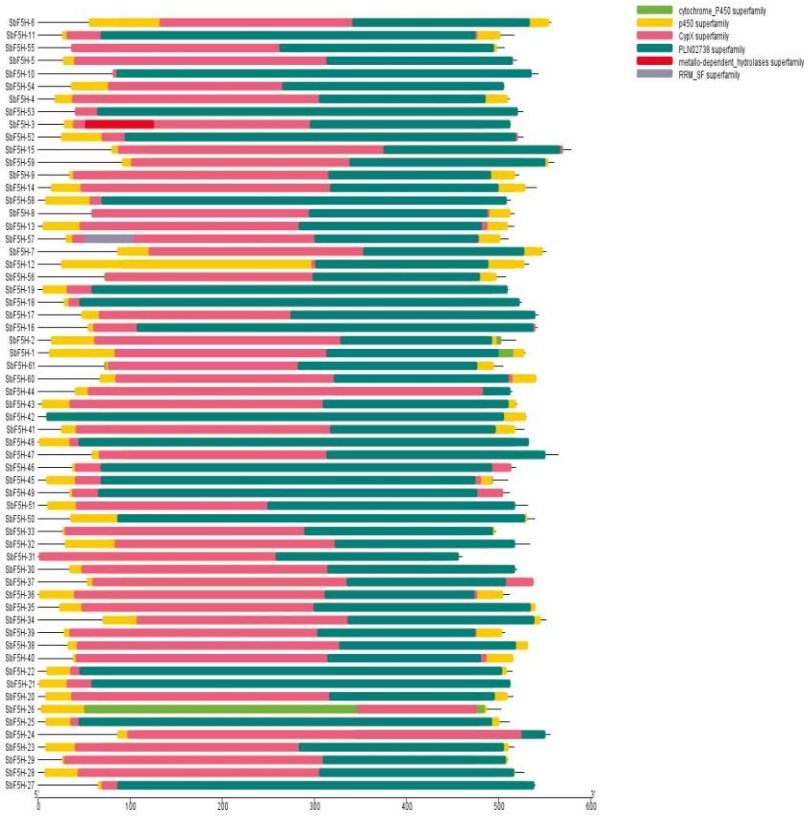


Fig. 3c. Conserved domain analysis of *SbF5H*

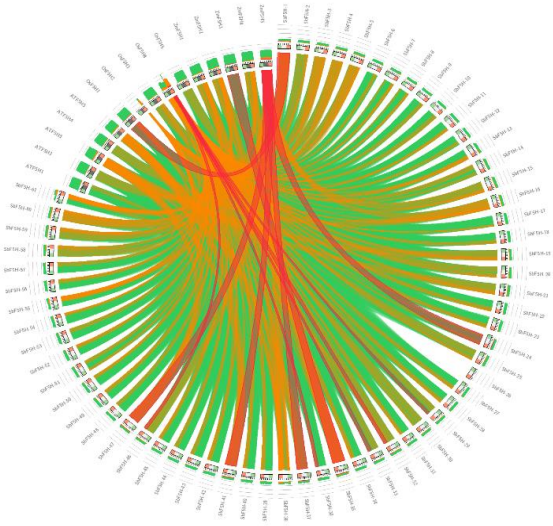


Fig. 4. Sequence similarity analysis *of SbF5H* genes by circoletto
Note: Coloring with green<=75, orange<=99.9999, red>99.9999

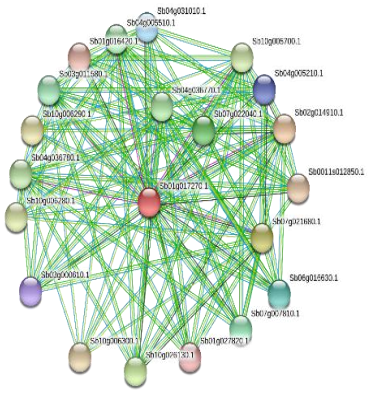


Fig. 5a. PPI (Protein interaction) network analysis *of SbF5H*

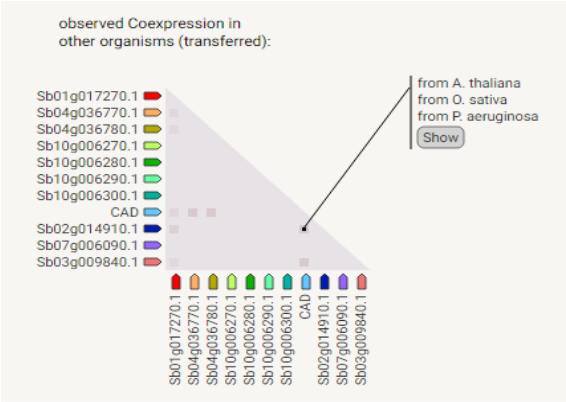


Fig. 5b. Gene co expression in other organisms *of SbF5H*

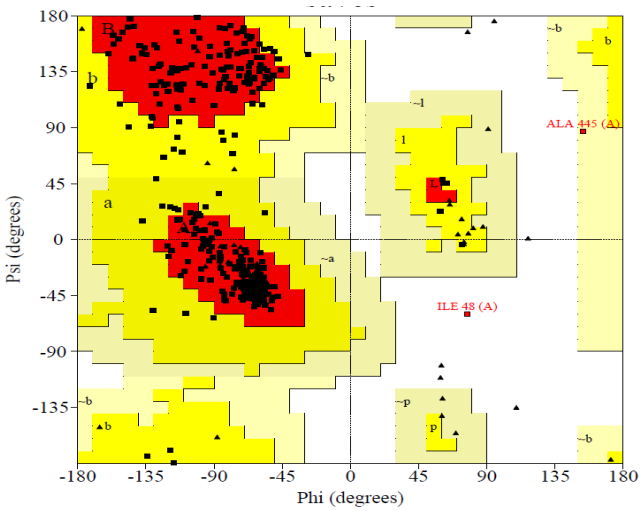


Fig. 6. Analysis of *SbF5H* protein by Ramachandran plot

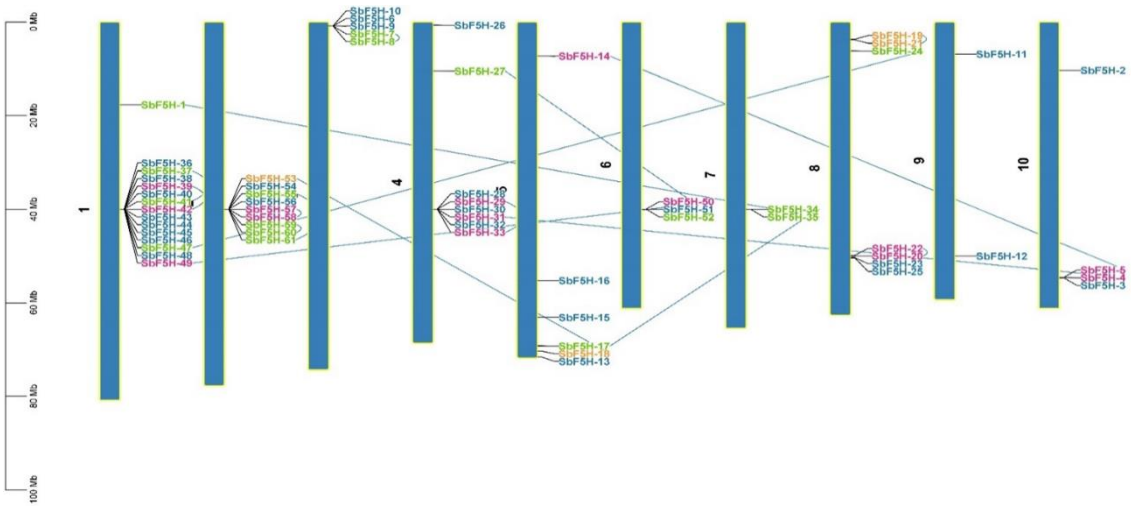


Fig. 7. Chromosomal location mapping *of SbF5H genes*

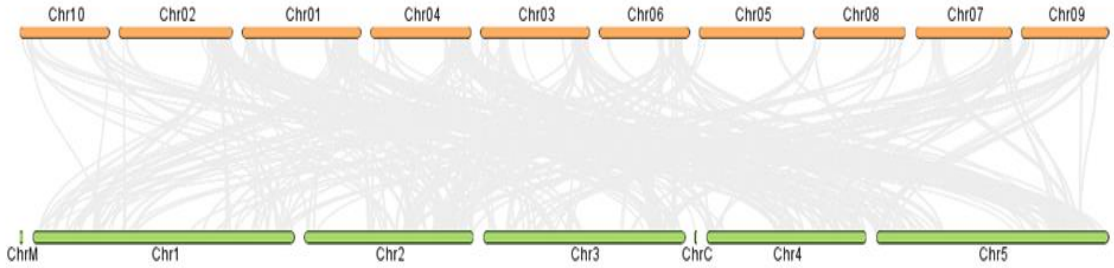


Fig. 8a. Synteny analysis of *Sorghum bicolor* and *Arabidopsis thaliana*

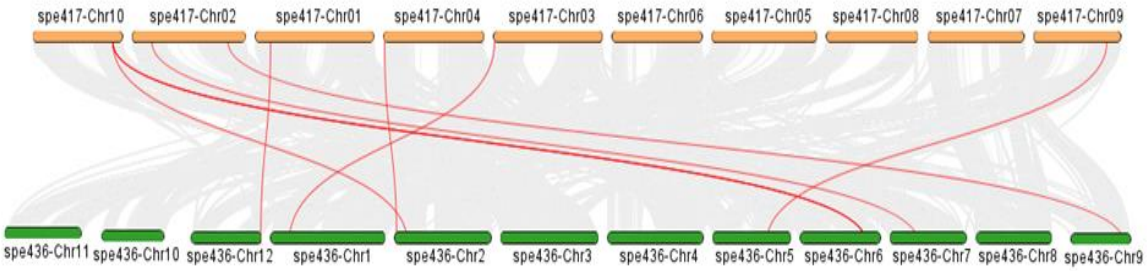


Fig. 8b. Synteny analysis of *Sorghum bicolor* and *Oryza sativa*

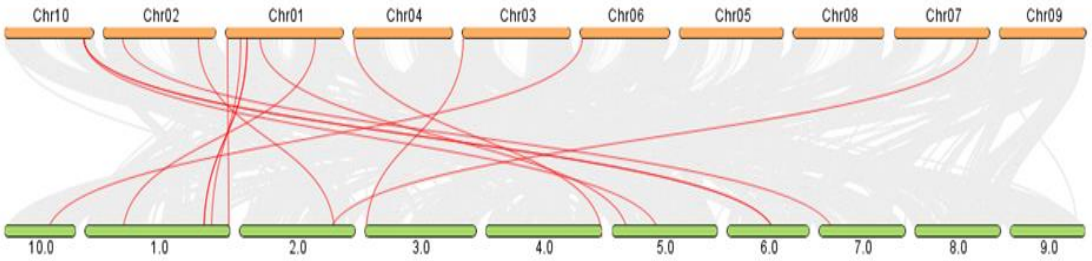


Fig. 8c. Synteny analysis of *Sorghum bicolor* and *Zea mays*

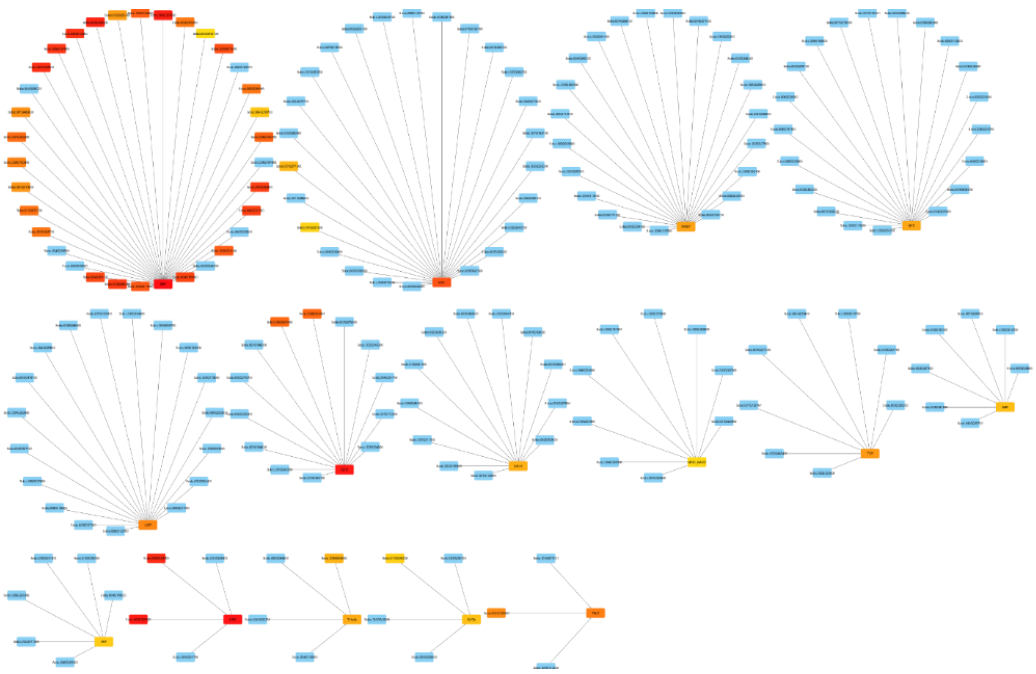


Fig. 9. Transcription factor analysis *Sorghum F5H* genes

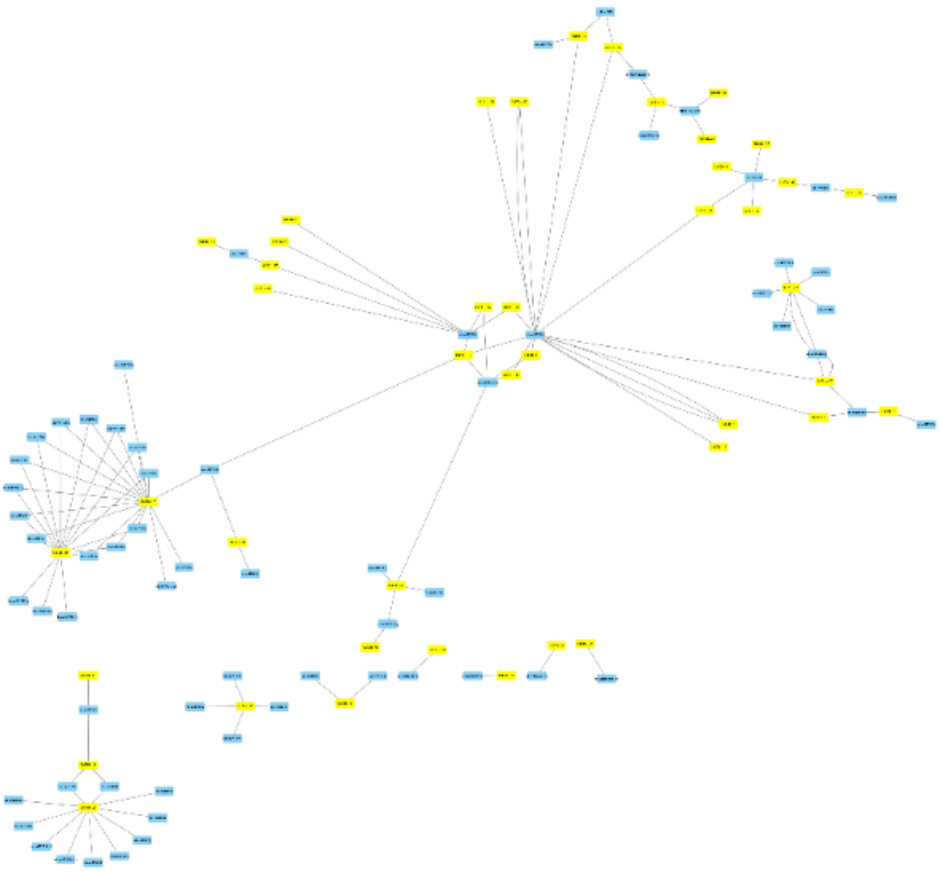


Fig. 10. miRNA prediction analysis *Sorghum F5H* genes

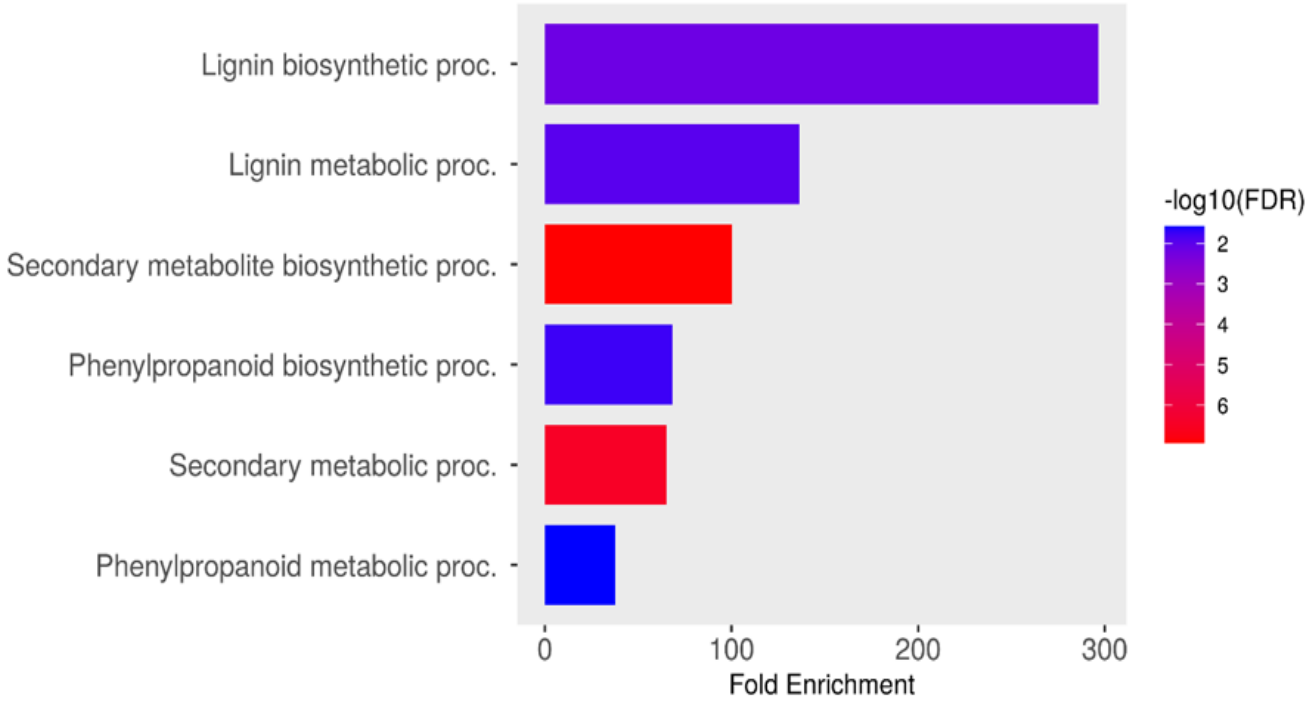


Fig. 11. Gene ontology enrichment study of *Sorghum F5H* genes

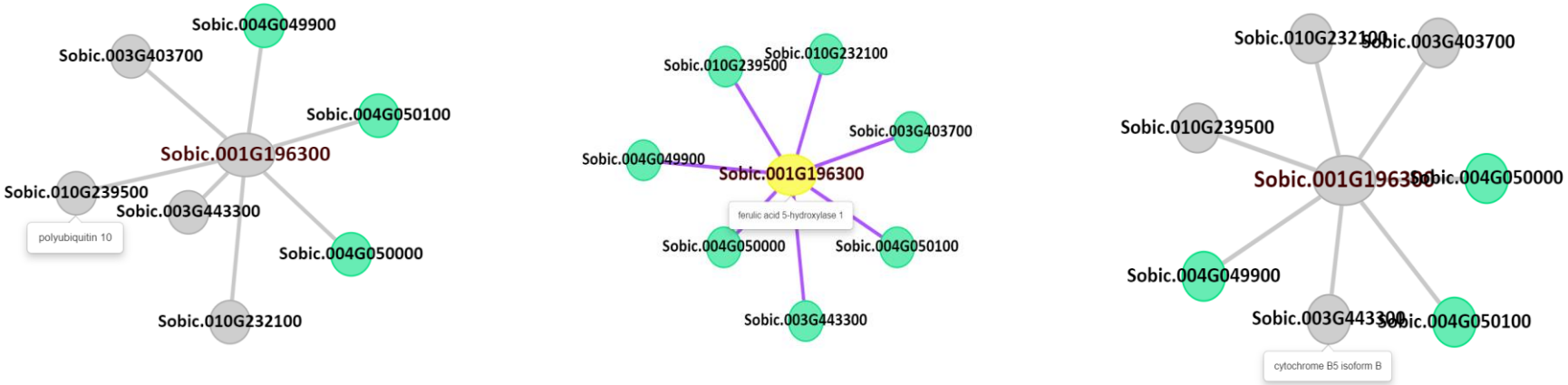


Fig. 12. Gene co-expression analysis of *Sorghum F5H* genes

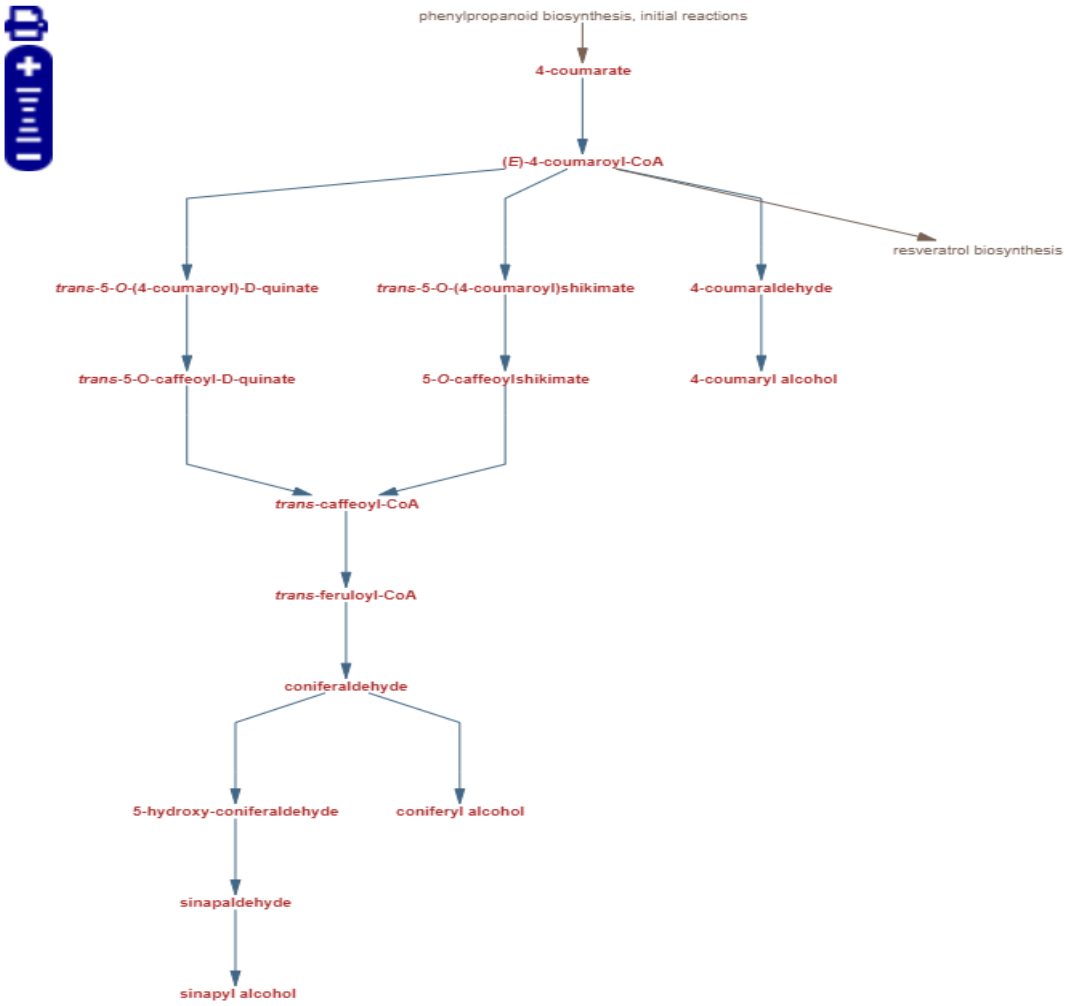


Fig. 13a. Phenylpropanoid biosynthesis pathway – KEGG analysis

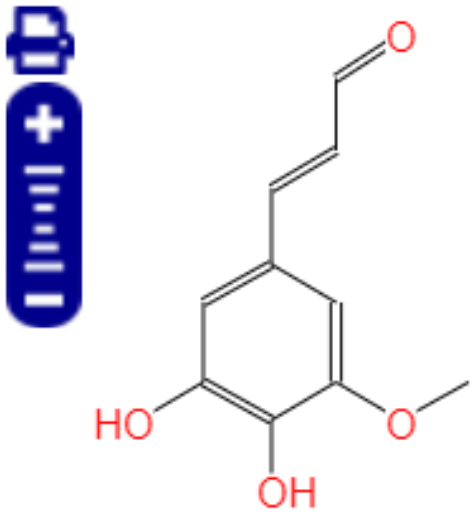


Fig. 13b. Structure of *Ferulate 5 hydroxylase*

Expression analysis:

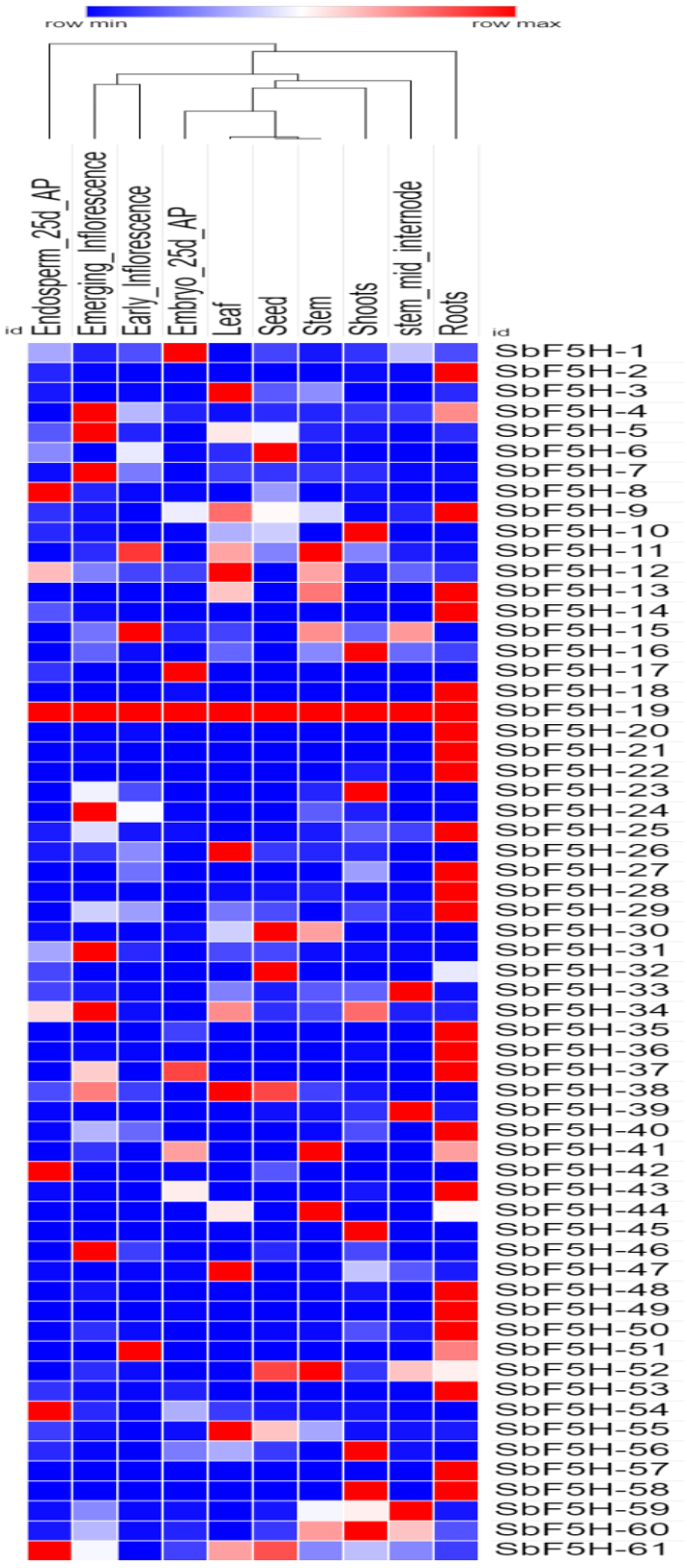


Fig. 14. Base line expression analysis of *SbF5H* genes in various tissues/organs of Sorghum bicolor

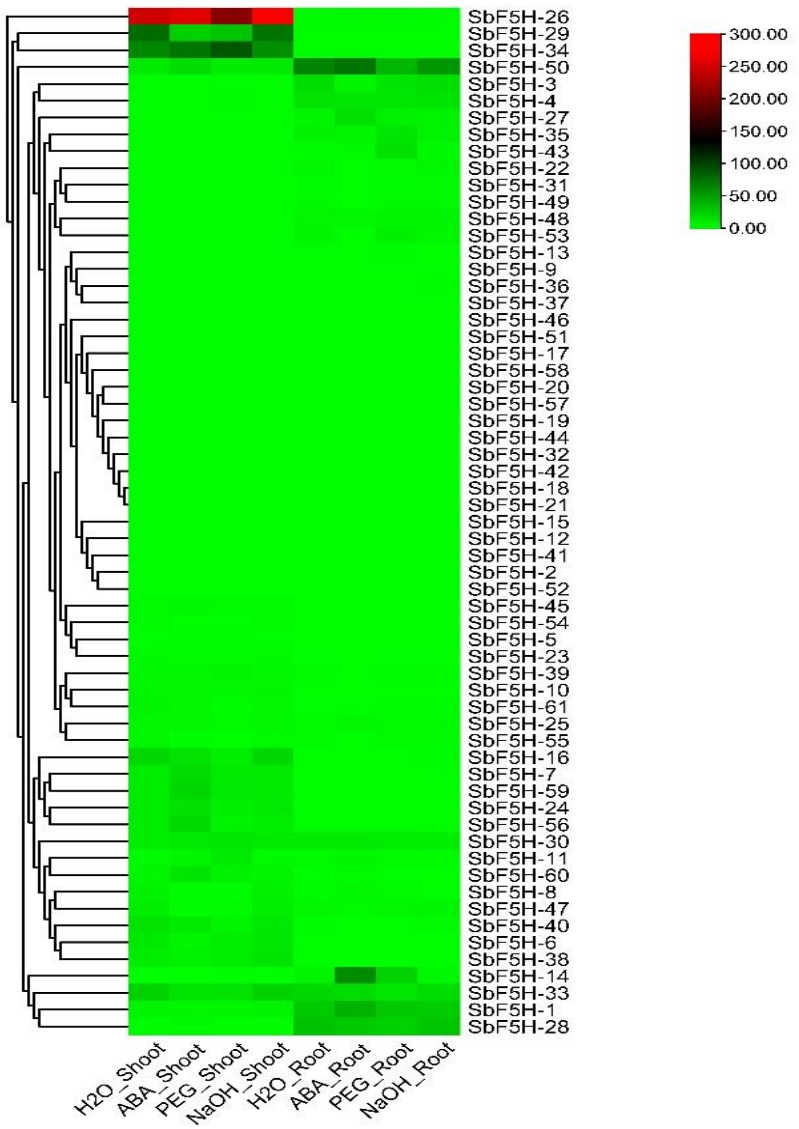


Fig. 15a. Expression analysis of *SbF5H* genes in different stress conditions

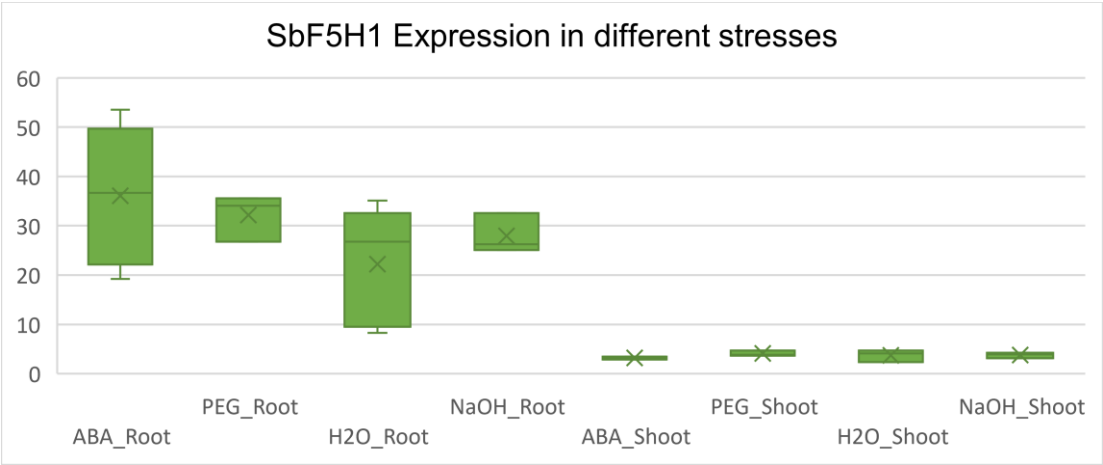


Fig. 15b. Graphical representation of expression of *SbF5H1* gene in different stress conditions

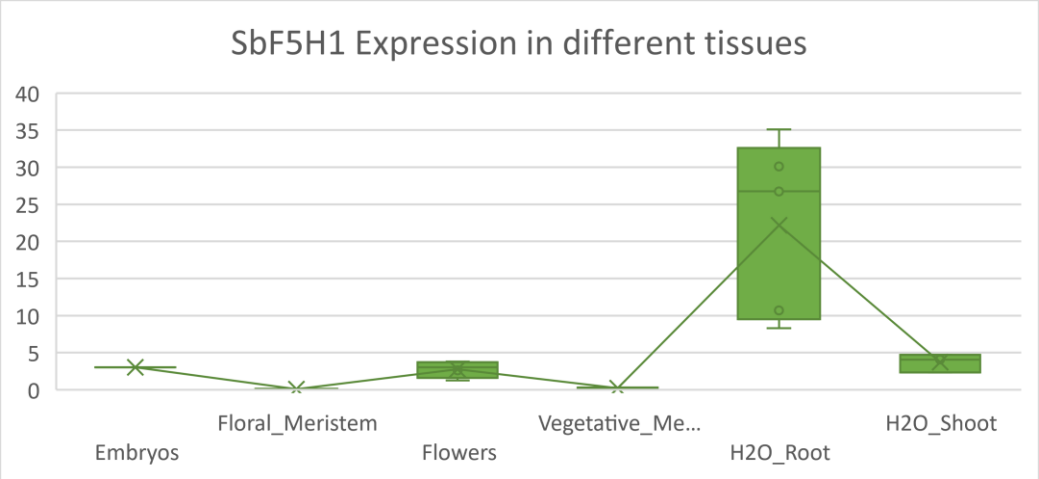


Fig. 15c. Graphical representation of expression of SbF5H1 gene in different tissues/organs of Sorghum bicolor

We observed the presence of numerous cis-elements in sorghum *F5H*s (Fig. 3b), which potentially serve as binding sites for transcription factors crucial in lignin production. To delve deeper, we utilized Plant TFDB to identify transcription factors binding to the regulatory elements of *SbF5H*s. Our analysis unveiled interactions with a variety of transcription factors, notably ERF, NAC, WRKY, MYB, bZIP, C2H2, bHLH, MIKC, MADS, TCP, SBP, ARF, LBD, Trihelix, GATA, and TALE. These transcription factors were found to interact with all *SbF5H*s homologs, indicating their involvement in lignin biosynthesis and response to abiotic stress conditions (ABA, PEG and NaOH). (Fig. 9) Additionally, we identified miRNAs potentially targeting the 61 *SbF5H*s to regulate their expression (Fig. 10). According to the analysis results, distinct miRNAs targeted various *SbF5H* genes, with Sbi-MiR5386 specifically affecting the function of the *SbF5H1* target. This comprehensive transcription factor analysis and miRNA prediction provide valuable insights into the regulation of *SbF5H* expression, offering potential avenues for engineering biomass accumulation and lignin biosynthesis in *Sorghum bicolor*.

Gene ontology: According to the findings from the GO enrichment study, the *SbF5H* genes exhibited significant enrichment in terms related to Hydroxylase activity, particularly GO:0046424 (Ferulate 5-hydroxylase activity) (Table S6; Fig. 11). This indicates their involvement in catalysing the hydroxylation of ferulate, a key step in lignin biosynthesis. Moreover, the enrichment in the Lignin biosynthesis process further underscores their crucial role in this fundamental biological pathway (accessible at http://structuralbiology.cau.edu.cn/sorghum/gene_detail.php?gene= Sobic.001G196300).

Moreover, the co-expression analysis of *SbF5H* homologs revealed intriguing associations with other genes. Notably, several genes, including polyubiquitin, cytochrome B5 isoform, and various transcription factors, were found to be co-expressed with *SbF5H* genes. This suggests potential regulatory interactions and functional relationships between *SbF5H* genes and these co-expressed genes, which could have implications for understanding the broader regulatory networks governing lignin biosynthesis and plant development (Fig. 12).

Additionally, our KEGG pathway analysis unveiled the diverse involvement of *SbF5H* genes in various processes critical to plant metabolism. Specifically, they were found to participate in the Lignin biosynthetic process and Lignin metabolic process, crucial for the formation and maintenance of plant cell walls. Furthermore, their association with the phenylpropanoid pathway highlights their role in the biosynthesis of phenylpropanoids, which are precursors to lignin, as well as their participation in the biosynthesis of various secondary metabolites essential for plant growth and defence mechanisms (Figs. 13 a & b).

***In silico* study of expression patterns of *SbF5H* genes:** The expression patterns of *SbF5H*s have been explored using RNA-sequence data from various tissues of sorghum

as well as under ABA, PEG, and NaOH stress conditions. All of discovered *SbF5H*genes have been found to be expressed in various tissues, while some of the genes exhibiting tissue-specific expression (Figs. 14 and 15; Table S4 and S5).

4. DISCUSSION

The study began by comparing Phytozome database sequences with reference *F5H* protein sequences from *Arabidopsis thaliana*, revealing homologous *F5H* protein sequences in the sorghum genome. Subsequent confirmation of conserved P450 domains indicated membership in the cytochrome P450 superfamily for 61 identified *SbF5H* genes [18]. Physicochemical properties of predicted *SbF5H* proteins were analysed, showing variations in sequence length, molecular weight, and isoelectric point. Subcellular localization predictions indicated diverse placements, with chloroplasts being predominant [17]. Multiple sequence alignment revealed protein homology rates ranging from 50% to 99% among *SbF5H* proteins. An evolutionary tree constructed from *SbF5H* and homologous proteins categorized them into four clades, demonstrating evolutionary consistency across species. The identification of conserved motifs through motif analysis provided insights into structural features and group distinctions within *SbF5H* proteins [19].

Structural analyses including 3D structure prediction and homology modelling contributed to understanding *SbF5H* protein configurations. Evaluation of SbF5H 3D models confirmed high structural integrity, crucial for subsequent studies. Protein-protein interaction analysis revealed associations with lignin biosynthesis enzymes, emphasizing functional roles within metabolic pathways.

Chromosomal mapping demonstrated uneven distribution across sorghum chromosomes, with hotspots on chromosomes 1 and 2. Syntany analysis suggested stronger relationships with monocots than Eudicots, indicating evolutionary divergence. Purifying selection was evident in Ka/Ks ratio calculations for segmentally duplicated *SbF5H* genes [17].

Cis-acting element analysis uncovered regulatory elements responsive to various stimuli, hinting at a complex regulatory network governing *SbF5H* gene expression. Transcription factor binding and miRNA predictions offered further insights into regulatory mechanisms, potentially impacting biomass accumulation and lignin biosynthesis [20].

Gene ontology enrichment analysis highlighted involvement in hydroxylase activity and lignin biosynthesis, corroborated by KEGG pathway analysis implicating roles in phenylpropanoid metabolism and secondary metabolite biosynthesis. Co-expression analysis unveiled potential regulatory interactions with other genes, enriching understanding of broader regulatory networks [21].

Overall, the study provides a comprehensive analysis of *SbF5H* genes, shedding light on their structural, evolutionary, and regulatory characteristics, with implications for lignin biosynthesis and plant development. Also, this study will expedite the enhancement of understanding of information stored in the database. It will have promising uses for experimental biologists who intend to improve plant crop performance under abiotic stress conditions environments.

5. CONCLUSIONS

This study on *SbF5H* genes in Sorghum bicolor unveils their structural, evolutionary, and regulatory characteristics. Bioinformatic analysis identified conserved P450 domains and diverse protein properties. Evolutionary analysis showed consistency across species and revealed potential divergence between monocots and eudicots. Further, analysis into regulatory networks involving cis-acting elements, transcription factors, and miRNAs were uncovered, impacting biomass and lignin biosynthesis. Gene ontology study highlighted roles in hydroxylase activity and phenylpropanoid metabolism. Protein interactions and co-expression patterns provided insights into lignin biosynthesis and stress response. Overall, our study enhances understanding of the *SbF5H* gene family in Sorghum bicolor, laying the groundwork for future research to improve biomass, lignin biosynthesis, and crop traits for sustainable agriculture and bioenergy.

SUPPLEMENTARY MATERIALS

Supplementary materials available is this link:https://ikprress.org/media/2024_PCBMB_12450.pdf

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

ACKNOWLEDGEMENTS

Anumandla Vinodkumar is thankful to the Department of Genetics, Osmania University, Hyderabad, for providing support for research.

Prashant Singam acknowledges financial support from DBT-BUILDER Program with grant number BT/INF/22/SP41415/2021.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annual Review of Plant Biology. 2003;54(1):519-546.
2. Bonawitz ND, Chapple C. The genetics of lignin biosynthesis: Connecting genotype to phenotype. Annual Review of Genetics. 2010;44(1):337-363.
3. Umezawa T, Goto D, Aoki S, Ishijima K, Patra PK, Sugawara S, Nakazawa T. Variations of tropospheric methane over Japan during 1988–2010. Tellus B: Chemical and Physical Meteorology. 2014;66(1):23837.
4. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. Plant physiology. 2010;153(3):895-905.
5. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Wyman CE. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344(6185):1246843.
6. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. Opportunities and challenges in biological lignin valorization. Current Opinion in Biotechnology. 2016;42:40-53.
7. Rinaldi R, Jastrzebski R, Clough MT, Ralph J, Kennema M, Bruijnincx PC. Paving the way for lignin valorisation: recent advances in bioengineering, biorefining and catalysis. Angewandte Chemie International Edition. 2016; 55(29):8164-8215.
8. Takeda Y, Koshiba T, Tobimatsu Y, Suzuki S, Murakami S, Yamamura M, Umezawa T. Regulation of Coniferaldehyde 5-Hydroxylase expression to modulate cell wall lignin structure in rice. Planta. 2017;246:337-349.
9. Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF. Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. Phytochemistry Reviews. 2004;3:29-60.
10. Liu X, Zhao C, Gong Q, Wang Y, Cao J, Li X, Grierson D, Sun C. Characterization of a caffeoyl-CoA O-methyltransferase-like enzyme involved in biosynthesis of polyethoxylated flavones in Citrus reticulata. J. Exp. Bot; 2020.
11. Malavasi UC, Davis AS, Malavasi MDM. Lignin in woody plants under water stress: a review. Floresta e Ambiente. 2016;23:589-597.
12. Maury S, Geoffroy P, Legrand M. Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. Plant Physiology. 1999; 121(1):215-224.
13. Liu SJ, Huang YH, Chang-Jiu HE, Fang C, Zhang YW. Cloning, bioinformatics and transcriptional analysis of caffeoyl-coenzyme A 3-O-methyltransferase in switchgrass under abiotic stress. J. Integr. Agric. 2016;15:636–649.
14. Alam O, Khan A, Khan WU, Khan WA, Ahmad M, Khan LU. Improving heat tolerance in betel palm (*Areca catechu*) by characterisation and expression analyses of heat shock proteins under thermal stress. Crop and Pasture Science. 2024;75(9).
15. Nakamura N, Katsumoto Y, Brugliera F, Demelis L, Nakajima D. Flower color modification in Rosa hybrida by expressing the S-adenosylmethionine: Anthocyanin 3′, 5′-O-methyltransferase gene from Torenia hybrida. Plant Biotechnology. 2016; 32:109–117.
16. Nakashima K, Ito Y, Yamaguchi-Shinozaki K. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis

and grasses. Plant Physiology. 2009; 149(1):88-95.
17. Alam O, Khan LU, Khan A, Salmen SH, Ansari MJ, Mehwish F, Wang HF. Functional characterisation of Dof gene family and expression analysis under abiotic stresses and melatonin-mediated tolerance in pitaya (*Selenicereus undatus*). Functional Plant Biology. 2024;51(4).
18. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Rokhsar DS. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research. 2012;40(D1):D1178-D1186.
19. Puri H. Sorghum-Sugarcane Aphid Interactions: Multi-Omic Approaches to Elucidate Plant Defence Against Sap-Feeding Insects (Doctoral dissertation, The University of Nebraska-Lincoln); 2023.
20. Li M, Zhao X, Li Y, Zhao X, Mai W, Li, Chen X. MePHD1. 2 affects the synthesis of cyanogenic glycosides by regulating the transcription of MeCYP79D2 in cassava. Biorxiv. 2024;2024-04.
21. Sarath G, Mitchell RB, Sattler SE, Funnell D, Pedersen JF, Graybosch RA. Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. Journal of Industrial Microbiology & Biotechnology. 2008;35(5): 343–54. Available:https://doi.org/10.1007/s10295-007-0296-3 PMID: 18205019

Peer-review history:
The peer review history for this paper can be accessed here:
https://prh.ikprress.org/review-history/12450