Genome-wide identification of *WRKY* gene family in Rice Bean

Dissertation submitted in fulfilment of the requirement of

MASTERS OF SCIENCE

IN

BIOTECHNOLOGY

BY

AKSHITA CHAUDHARY (225111007)

Under the supervision of

DR. SHIKHA MITTAL

2024

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN

DECLARATION

We hereby declare that the project work entitled "**Genome-wide identification of *WRKY* gene family in Ricebean (*Vigna umbellata*)"** has been solely submitted to the Department of Biotechnology and Bioinformatics, **Jaypee University of Information Technology, Waknaghat (Solan)** is a record of an original work done by us under the supervision of **Dr. Shikha Mittal.**

Signature:

Akshita Chaudhary (225111007)
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat, Solan.

Date:

SUPERVISOR CERTIFICATE

This is to certify that the work reported in the M.Sc. Dissertation report **"Genome-Wide identification of *WRKY* gene family in ricebean (*Vigna umbellata*)**" submitted by Akshita Chaudhary at Jaypee University of Information Technology, Waknaghat, India, is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Dr. Shikha Mittal
Assistant Professor
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology
Waknaghat, Distt-Solan, H.P. - 173234

Date:

ACKNOWLEDGEMENT

I would like to express my profound gratitude to my guide Dr. Shikha Mittal for her guidance, support and constant encouragement throughout the course of this project work. She has been more than just my project guide; at times a mentor to rescue me out of my doubts. She has always helped me to work hard and also taught me how to implement different ideas to deal with the problem. Moreover, she taught me to not give up and many other valuable lessons.

Furthermore, I would like to acknowledge Vice-Chancellor Prof. (Dr.) Rajendra Kumar Sharma, Prof. (Dr.) Ashok Kumar Gupta, Dean of academics & research for providing me with an opportunity to be a part of the institute and to complete my Master's Degree.

I also want to mention the HOD of Biotechnology and Bioinformatics Prof. (Dr.) Sudhir Kumar has been a source of immense motivation and inspiration both for my academic and personal life. He was never, and I know will never be, more than just a phone call away. He has helped me in almost every aspect I have asked him for.

In addition, I would like to thank all the faculty members of the BT/BI Department of JUIT, who have helped me whenever I needed and also would like to thank all the lab engineers and specially Ms. Somlata Sharma for providing me with a workplace and for always motivating me.

I would also like to appreciate the part that my classmates (Kritika, Anuja, Shan and Priyanka) have played in shaping this project work. They have been my constant support and cheered me up at hard times. They helped me whenever I had any doubts. Thanks a lot!

I would like to thank the almighty God for his grace throughout my life. Last but not the least I would like to thank my Mother and Father who have always supported me through thick and thin and have been a constant source of encouragement and support; also, who has never given up on me and always motivated me.

[Thanks to JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY]

Akshita Chaudhary
M.Sc. Biotechnology
JUIT, Solan

TABLE OF CONTENT

LIST OF FIGURES

ABSTRACT

The study of *WRKY* gene family had been steadily advancing in recent years, primarily in the areas of hormone response and environmental stress. But in order to fully understand the *WRKY* gene family, is also necessary to concentrate on the identification and whole-genome identification of the family. Previous research has extensively examined the entire *WRKY* gene members in *Arabidopsis thaliana, Vigna species, Glycine max,* legumes, and other plants. The whole-genome and RNA-seq data of *Vigna umbellata* were used for comprehensive analysis in this study, which included *WRKY* gene family identification, multiple sequence alignment (MSA), phylogenetic tree construction, analysis of conserved motifs along with their domains, identification of all cis-acting elements, intron-exon distribution analysis, synteny analysis along with expression analysis of *WRKY* genes. However, there is some information about the *WRKY*-family in *Vigna umbellata*. In this study, 84 *WRKY*-gene members were identified in *Vigna umbellate* (ricebean). So, after removal of non *WRKY* domains and their sequences, total 79 *WRKY* genes of ricebean were obtained. Based on the phylogenetic analysis, 79 *WRKY* genes was divided into 3 major groups based on their conserved motifs: Group1 which consisted of 48 *WRKY* genes and has conserved motif1 and motif 2, Group2 consisted of 14*WRKY* genes and has conserved motif 1, motif 2 and motif 4 and Group 3 consist of 17 *WRKY* genes which has conserved motif1, motif 2, motif 3 as well as motif 5. In this study, total 66 cis-acting regulatory elements were analyzed which includedA-box, AAGAA-motif, Box4, CAAT-box, CAT-box, CCAAT-box, CGTCA motif, CTAG-motif etc. The RNA-seq data which is used for expression analysis is represented by heatmap which shows that all 79 *WRKY* genes of ricebean exhibited discrete expression during development stages. Synteny analysis is done across genomes of *Vigna umbellata* with *Arabdiopsis thaliana, Vigna radiata, Vigna angularis, Vigna unguiculata, Glycine max and Cicer arietinum* to identify conserved syntenic regions and orthologous gene pairs. The output of these combinations indicates the possibility of studying the evolutionary and functional aspects of *WRKY*-gene family, they can be exploited as essential parameters for production as well as research of *Vigna umbellata*.

CHAPTER 1-

INTRODUCTION

INTRODUCTION

The comprehensive examination of *WRKY* genes, at the level involves an investigation into the *WRKY* gene family across an organisms entire set of genetic material. This examination encompasses aspects, including the identification of *WRKY* genes their evolutionary history, how they express themselves their response to types of stress and their structural characteristics. Researchers employ methods to predict and analyze *WRKY* genes. They also study where these genes are located within chromosomes identify shared domains among them.

Perform analysis to understand their evolutionary relationships. Additionally, this examination entails exploring the elements found in the sequences preceding *WRKY* genes and investigating how these genes express themselves in tissues under diverse conditions. Numerous studies had been conducted on genome analysis of *WRKY* genes in numerous plant species (rice, buckwheat, tomato, *Acer truncatum* and *Aquilaria sinensis*). These studies have provided insights, into the diversity and functions of *WRKY*-gene family in plants. Genomewide analysis of *WRKY* genes include:

- Functional dynamics
- Evolutionary dynamics
- Structure analysis
- Network analysis
- Epigenetic regulation
- Responses to biotic/abiotic stresses

There are numerous approaches that can be used for genome wide analysis of *WRKY* genes in crops and plants. The most common approach is to use homology search to identify *WRKY* genes based on their sequence similarity to known *WRKY* genes from other plants [1].

THE *WRKY* GENE FAMILY

Figure 1.1 *WRKY* domain

There are specific sequences in DNA to which only transcription factors can bind. The transcription factors can also be involved in activation or repression of transcription of the downstream target genes. Proteins called transcription factors constitute an integral part of the gene expression, involving turning on or off the information (DNA) used to produce RNA, and eventually proteins.

Here are key features and functions of transcription factors:

DNA Binding

Transcription factors have specific DNA binding domains(figure1.1) which recognize and binds to regions that are specific DNA sequences, usually on regulatory region of target gene. The interactions of transcription factors with the different genes is very specific leading to binding of these binding events.

Activation or Repression

They might serve by activating/inhibiting the transcription of particular targeted genes. The activation of transcriptional factors help in attracting the transcriptional machinery and hence expression of genes is promoted. On the other hand, repressor transcription factors suppress gene expression by limiting the affinity between transcriptional machinery and RNA polymerase.

Modulation of RNA Polymerase

Transcription factors affect the action of RNA polymerase which produces RNA from DNA. These factors act either positively or negatively to allow RNA polymerase to initiate transcription.

Cellular Response to Signals

There are several transcription factors that respond to signals from the external environment of the cell. Such as in response to hormones, growth factors and/or cellular stress. It enables cells to change the way of their gene expression according to changing circumstances.

Role in Development

Cell development and differentiation are dependent on particular transcription factors. These gene products regulate the specific set of genes that define a particular cell type or stage of development, determining fate of cells [2].

In plants, *WRKY*-gene family members has a large group of transcriptional factors that engages in a numerous of physiological processes. They are essential to the growth, expansion, development, and stress-reactions of plants species. It's basically a protein that activates another gene and that helps to regulate their expression.

Figure 1.2 depicted a circular format diagram, with five interconnected text boxes describing key aspects of *WRKY* genes. These features highlight the main characteristics and functional roles of *WRKY* genes in biological processes.

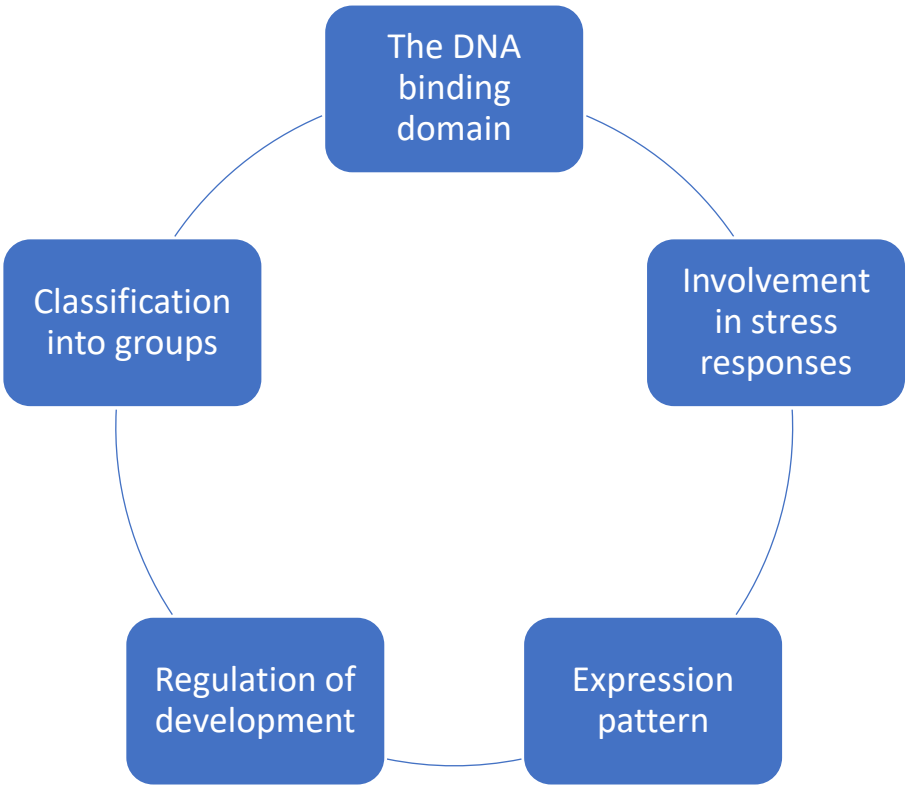


Figure 1. 2 Key features of *WRKY* genes

Structural characteristics:

WRKY gene family carry at a minimum, one extremely conserved domain and consist of about sixty (aa) amino acids residues signature domain with a very conserved heptapeptide sequence ' *WRKY*GQK' pursued by Zn finger motif (C2HC type or C2H2) in the C terminal end portion and DNA- binding domain in N-terminal end region of the sequence [3]. The transcriptional factor (TF) is attached to W-box (C/T) TGAC (C/T) in upstream of the inducible gene. The process of transcriptional autoregulation and cross regulation by TF via w-Box takes place by this sequence specific binding element.

Figure 1.3 shows *WRKY*'s transcriptional factor (TF) in dark-blue box adheres to a W box cis regulatory sequence in a light yellow-box(C/T) TGAC(C/T) for administration and execution of transcriptional processes and other environmental processes.

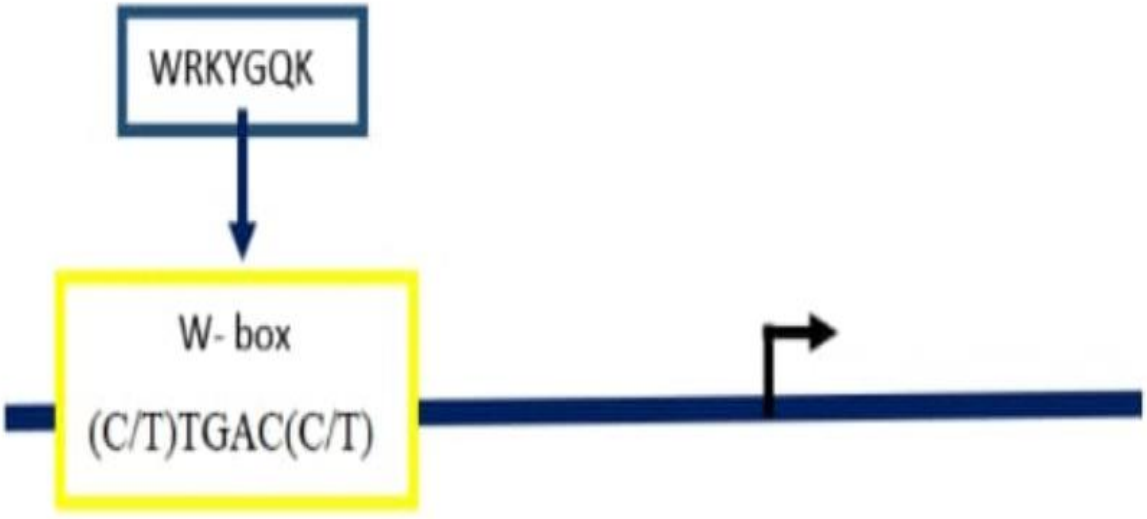


Figure 1.3 *WRKY* domain binding to W-box sequence

Certain *WRKY* proteins interact with one another to promote the *WRKY* gene's promoter region, and transcriptional processes that cross-and auto-regulate *WRKY* genes also stimulate other genes through a signal cascade [4].

A single *WRKY* TF may control several pathways. Through a hydrophobic interaction with the Gly residue in the middle of the N-terminus, four ß sheets in the *WRKY* domain with the zinc binding pocket—which contain Cys and His zinc finger residues will provide structural stability. The *WRKY* gene grouping, an important team of information providers in crops, has undergone organized categorization. Its significant involvement in a variety of biotic along with abiotic stress responses is revealed by genome-wide analyses. Based on the number of *WRKY* domains along with zinc finger motif pattern, members of the *WRKY* family can be rigorously categorized into groups. The comprehensive study of *WRKY*-genes transcriptional factors extends to the investigation of shared synteny among different plant species, enhancing our understanding of their evolutionary relationships.

WRKY gene families are usually classified into three major categories based on expression patterns and synonymous substitution rates, among other considerations. This analytical approach facilitates a nuanced understanding of the functional diversity within the entire *WRKY* gene family. Genome-wide characterizations emphasize the significance of the *WRKY*

domain(s) harbored by family members, highlighting their role as transcriptional regulators in higher plants [5].

WRKY-genes are generally classified into 3 major groups based on the sum total of *WRKY*-domain(s) and the type of zinc finger-like motifs contain, The protein is fragmented into 3 groups depending on the type of domain and zinc finger motif it has together with the number of domain available, and these include: group I has title 2-*WRKY* domains with C2H2-zinc finger motifs, group II with only one *WRKY* domain along with C2H2-zinc finger motifs, and group III with one *WRKY* domain along C2H2 zinc finger motif. It has a region called DNA binding domain which is positioned in N-terminal and another motif called zinc finger in the C-terminal. The theme of both of these two motifs is critical to achieve the tight combination of *WRKY* transcription factors.

RICEBEAN

Till now, no analysis has been done on *WRKY* gene family in rice bean. The ricebean genome is small (414mbp) as compared to the genomes of other crops, such as maize and soybean. However, it is still larger than the genomes of many other plants, such as *Arabdiopsis thaliana* (135mb). A warm-seasonal, periodical legume vine that bears yellow colour flowers and has small comestible and edible ricebean (*Vigna umbellata*). The plant widely grows in tropical and sub-tropical areas, such as Asia and Africa.

Here are some key facts about rice bean:

Scientific name: *Vigna umbellata*

Family: Fabaceae

Common names: Black gram/ rice gram (vaibha), dewa-baani / cow pea ricebean (vaibha-shaalam), ur-dal (ur-dahi), (vaibha-shaalam), cholai (chaal).

Origin: Southeast Asia

Habitat: Warm season climates and well drained soils.

Growth habit: Vine

Height: Up to 2 meters

Flowers: Yellow

Fruits: Pods containing 4-6 small seeds

Seeds: Fit for consumption, fit for food, feed, or for cooking oil.

Nutritional value: The protein, fiber, vitamin and mineral content of the meal.

Uses: Food, feed, forage, soil improvement

As ricebean is included in the popular crop in the world, particularly in South as well as South East Asia, large number of rice beans are cultivated each year. To appeased this demand, enhancing yield is very essential. Research on the planting microenvironment has demonstrated that biotic and abiotic stressors play a significant role in limiting ricebean yield.

OBJECTIVES

1. To identify *WRKY* family genes in *Vigna umbellata* using orthologs of *Vigna* Species.

2. Expression analysis of identified *WRKY* family genes using RNA-Seq data.

CHAPTER 2- REVIEW OF LITERATURE

Overview of *WRKY* transcription factor

Within plant species, the *WRKY*-gene family is one of the largest and the biggest members of transcriptional factors. These protein family play a very significant and integral roles in increase of growth, elongation, expansion and response to stress conditions and environmental conditions. A conserved *WRKY* domain, which helps these genes to adhere particular DNA-sequences and controls the gene-expression of downstream target genes, is one of their distinguishing features. The W-box (TTGAC/T) in the target gene's promoter region is where the *WRKY* domain binds. Its unique amino acid(aa) sequence (*WRKYGOK*) and zinc (Zn) finger-like conserved motif, which binds zinc ions to preserve structural integrity, are its defining features. *WRKY* genes are particularly acclaimed for their function in plant defence systems against pathogens and in responding to environmental stressors such temperature changes, salt, and drought [6].

Significance in plant biology

WRKY genes are essential for a variety of plant functions, such as:
Stress Response: They control the reactions to biotic and abiotic stress.
Plant Immunity: They are very important components of plant defence systems that ward off invaders.
Secondary Metabolism: The manufacture of secondary metabolites is regulated by *WRKY* proteins.
Development: Growth regulation and developmental processes are impacted by *WRKY* variables.
Senescence and dormancy: The control of dormancy and senescence, which affects crop quality and yield, is linked to *WRKY* proteins.

Mechanism of action

It is well established that the W-boxes in promoters region of the target genes are recognition and adhering site of *WRKY*-transcription factors. According to these, these proteins either activate or repress the gene. Meanwhile DNA-binding activity can be redox-state-dependent, phosphorylation, and activity can be modulated by interacting with other proteins [6].

Regulation of *WRKY* transcription factors

Regulatory modes: Additional internal and external signals target *WRKY* genes for transcription. Furthermore, several *WRKY* proteins undergo post-translational modifications which are phosphorylation in *WRKY*; these can have a direct effect on subcellular localization, stability and protein-protein interactions to influence the protein's activity. Studies show that *WRKY* gene factors may occur as to homo- or hetero-dimers. They interact with other proteins groups as well and therefore modulate their activity and show their expression [7].

Role in stress responses

The *WRKY*-gene family is important for plant's capacity reaction to different kinds of stress. These transcription factors aid plants in adapting to harsh environmental conditions by frequently serving as the central component of stress response networks. Key elements of *WRKY* gene family's execution of plant's stress responses are listed below:

Stress-Responsive Pathway Activation: promoting the expression of genes that are responsive to stress signalling conditions.

Cross-Talk Among Stress Signalling Pathways: Additionally, *WRKY* TFs integrate multiple signals and pathways possible and warrant that the plant could respond to multiple complex or simultaneous stress factors. Further, they generally execute an important role in fine-tuning stressful responses in both positive as well as negative regulatory loops.

- **Abiotic-stress tolerance**: *WRKY*-transcription has implicated in acknowledgement to drought, salinity and extreme temperatures, including in the execution of water-stress-related genes, ion transporters and genes involved in osmolyte biosynthesis [8].

- *WRKY* TFs contributes to the protection of cellular integrity through oxmo-regulated stress responses, in order to maintain cell turgidity under circumstances that can result in dehydration or oversaturation of the cells [9].

Developmental regulation under stress:

Impact on growth and development. *WRKY* proteins also influence developmental processes in stress environments, including seed germination, root elongation, and senescence. Stress-induced developmental changes. *WRKY* transcription factors mediate developmental adjustments induced by certain stresses, which are called stress-induced flowering or precocious senescence.

Modulation of Plant immune responses:

They are known for their role in pathogen defence because they control genes participating in the salicylic acidic dependent as well as jasmonic acid pathways/ethylene dependent pathways.

In a (SAR) systemic acquired resistance model of "whole-plant" resistance that provides continuous immunity towards an array of pathogens, some *WRKY* transcription factors are involved [10].

Evolutionary adaptations:

Conserved and Novel Functions: While certain *WRKY* protein functions are shared by all plant species, other *WRKY* protein functions might be unique responses to particular environmental stressors.

Gene Family Expansion and Diversification: Plants ability to withstand a wide range of stress conditions is probably attributed to the evolutionary adaptations that have guided to the enlargement, expansion and diversification of the *WRKY*-gene family members [11].

Technology and biotechnological applications:

Genetic Engineering: Plant stress tolerance may be improved by genetically modifying *WRKY* genes using tools like CRISPR/Cas9.

Molecular Breeding: Molecular breeding programs can benefit from our growing understanding of the role and function of *WRKY* genes in stress responses to develop crop varieties that are resistant to stress.

Use in crop improvement

Molecular Breeding: Marker-assisted selection may be aided by the discovery of *WRKY* genes linked to desired features. Genetic engineering: Certain *WRKY* genes can be overexpressed or silenced to produce crops with higher yields or stress tolerance.

Genome wide identification strategies:
Several essential approaches are involved in genome-wide identification:

Whole Genome Sequencing: *WRKY* gene families can be thoroughly identified thanks to the availability of full plant genome sequencing.
Computational Annotation: Based on sequence homology and domain features, bioinformatics methods aid in the prediction and annotation of *WRKY* genes.
Phylogenetic Analysis: The evolutionary links and functional divergence of *WRKY* genes are revealed by phylogenetic trees built from various sequence alignments.
Expression Profiling: Methods like RNA sequencing (RNA-seq) and microarrays are used to help determine how *WRKY* genes express themselves in different environments.

Research highlights in different plants

The study analyzes that potatoes (*Solanum tuberosum*) contain 79 *WRKY* genes. *WRKY* was categorized using phylogeny and multi-sequence alignment into three groups. Group II comprised of over fifty-two *StWRKY*s, while Group III comprised fourteen *StWRKY*s and the First Group consisted of thirteen *StWRKY*s.

In the phylogenetic tree, the superfamily II finally had five subgroups. The *StWRKY* genes combined with *StWRKY*79 which were located in eight homologous gene pairs and seven corresponding gene pair of the *StWRKY* family gene were also mapped on the potato chromosomes. The microarray assay was performed on 22 *StWRKY*s quantified their expression level and overall, it was found that their expressions are variable. With respect to

the promoter regions of *StWRKY*s, the cis-element predictive analysis found multiple motifs that are putatively related to heat, drought, and salicylic acid which are the world of xenobiotic detoxification. Analysis of the motif revealed 20 genes within the genome sequences which can earlier provided deep insight into the diversity among genetics genes. the C terminal *WRKY* domain has been created with the motifs 1-4 as key elements. Motifs 1-4 appear in genes of almost all cases, meanwhile, the W-box-binding domain is mainly formed by motifs 6, 7 and 13. *WRKYGQK or WRKYGQK*-like domains made up motif 1 and motif 6, which were widely found in all *StWRKY* genes. Using the GSDS2.0 website for analysis of exon/intron structure, it was possible to determine that nearly all *StWRKY* genes had at least one intron. Approximately 50% of *StWRKY*s had two introns. Additionally, 26 members had more than three introns, while only six members had one and having one to five introns on average. Two members, *StWRKY*23 and *StWRKY*24, did not, however, possess introns.

Five *StWRKY* genes with varying responses to heat, drought, salt, and SA treatments were identified through expression profiling. These might be the subject of more studies as potential genes for abiotic stress signaling [12], Similarly, in case of *Arabdiopsis thaliana,* 72 *WRKY* genes were divided into 3 groups based on their sequence analogously. *AtWRKY*39 involved in tolerance to heat. *WRKY* gene were found to be involved in the regulation of leaf senescence. Another study identified a *WRKY* gene that is involved in the defense response against fungal pathogens [13].

Another study showed, 95 *DcWRKY* genes were found using the carrot genome and transcriptome, and these genes were then splitted into 3 main groups. Phylogenetic analysis of the carrot and *Arabidopsis WRKY* protein separated them into seven subgroups. The exon/intron structure showed that majority of *DcWRKY*s, forty-two, had two introns. These were pursued by 18 *DcWRKY*s with only one intron and 12 *DcWRKY*s with 3 introns. Exon–intron structures of *DcWRKY* genes that belonged to the same group appeared to be similar. The MEME program predicted the conserved motifs in order to investigate the diversity within each group. A *WRKYGQK* sequence was present in motifs1, 3, and 5.

In order to shed light on the source and dispersal of *WRKY*-gene family members, we have provided a comparative study along with schematic phylogeny of 22 species (plants, protozoa, and slime mould). To identify the 9 homologous factor groups in lower and upper plants from various taxonomic groups, a thorough investigation was conducted. Using the carrot genome,

the 38 *DcWRKY* proteins that serve as orthology-related interaction partners between carrot and Arabidopsis were determined. As shown by the yeast two-hybrid experiment, *DcWRKY*20 interacts with *DcMAPK1* and *DcMAPK4*.

These *DcWRKY* genes are engaged in root/shoot elongation, development and responded to biotic and abiotic stresses, according to transcriptome data and qRT-PCR. Comprehensive examination can be used to examine the biological functions and the evolutionary history of *WRKY* genes. *DcWRKY*33, was shown to be upregulated in carrot plants under drought stress. Another *WRKY* gene, *DcWRKY*6 was shown to be engaged in the carrot responded to salt stress. *DcWRKY*11, was shown to be engaged in the carrot defense responded against fungal pathogens and viruses and bacteria [14].

According to an analysis of the cucumber *CsWRKY* gene expression, under normal growth and developmental conditions, Altogether, 48 *WRKY* genes displayed either higher transcript abundance or modulated expression patterns. Out of the *WRKY* related genes, 23 genes expressed differently to minimum one abiotic stressses (nutritional deficiency, drought, extreme temperatures, salinity or cold). Despite the fact, the associated expression profiles of the putative *Arabidopsis WRKY's* gene (*AtWRKY*) orthologs and stress inducible *CsWRKY* genes except for the ones from group 3 *WRKY*-genes were significantly at relatively higher levels, yet the suppressive effects of the treatment could still be observed. Moreover, one must note that during this evolution, positive selection pressure has been applied on more than one duplicated group 3 (*AtWRKY*) genes [15].

Additionally, 61 *WRKY* family genes from cucumber were found via annotation of latest assembled genome (V3. 0). By analyzing the homologous genes in the related species, the phylogenetic and syntenic analysis were implemented to explore the development and progression of the (Cucumis sativus) cucumber *WRKY*-family genes members. It was seen that the three nuclei of 61 *CsWRKY*s were recognized, and the gene structures and motif composition of these groups were consistent.

According to the tissue gene-expression profiles of the *WRKY* genes, some *WRKY* genes showed expression specific to particular organs and tissues, and twentyfour *CsWRKY* genes showed intrinsic and essential expression as fragment per kilobase per million mapped fragments (FPKM) less than 1 in nearly every samples, indicating that all these *WRKY*s may

be very essential for the growth, expansion, elongation, organogenesis and tissue development of cucumber plants. Examining the *CsWRKY* gene expression level patterns is important because it reveals that 3*CsWRKY*- genes simultaneously reacted to every treatment tested, twelve genes were found to be expressed in response to downy along with powdery mildew infections which causes death, and 5*CsWRKY* genes strongly responded to heat and salt stresses.

Multiple *Cucumis sativus WRKY* genes were detected to be activated or even restrain edover dissimilar and alternative period following abiotic as well as biotic stress treatment, suggesting that cucumber *(Cucumis sativus*) *WRKY*-genes may have distinct functions during various stress reactions and their expression level patterns may change in reaction to all the stressors and other environmental conditions [16].

In case of *Melastoma dodecandrum* 126 *WRKY* members were found. Phylogenetic study classified them into 3 major primary groupings, with group II further fragmented into 5 groups. Group I (26) following group 2 (80), and group 3 (20). Group II comprises the following five major subgroups: IId (10), IIc (28), IIe (13), IIa (11) and IIb (18). On the phylogenetic tree, groups IIa, IIb, and IId and IIe were firmly distinguished, but group II-c's distribution was more similar to group I's, suggesting a closer relationship between the two.

Distribution of the *MedWRKY* genes proved to be out of balance along the 12 chromosomes. On the whole, all *MedWRKY* members comprise of three and eight motifs. Majority of *MedWRKY*, Motifs 1, 2, and 4 were identified in these groups. Certain motifs were only found in particular groups; Motif 9 was unique to one, Motif five was unique to IIa and IIb, and Motif ten was likely unique to IIb. The 126 *MedWRKY*s had varying distribution of exons along with introns having 2–13 exons following 0–12 introns in each gene and *MedWRKY*6 and Med *WRKY*36 had zero introns. *MedWRKY* 32 was one with highest exons and introns. According to the analysis, genes with comparable structures were generally accumulating together in the same class [17].

Furthermore, there was functional diversity indicated by the similarity and difference in sequence composition and gene structure within and between groups. The *WRKY* gene family has 12 cis-acting regulatory elements including bZIP, CATA, AT-HOOK, bHLH, and C2H2 were examined and shown. Therefore, for the total group, there are 2805 others, 1718 bZIPs

and 2030 bHLHs, which account for about 58% of the group. GATA (201), TBP (246), and Dehydrin (266a) had fewest numbers, accounting for only 6% of sum total of genes. The entire *MedWRKY* gene was either ZF-HD or bZIP-dominant in every case. for example, these areas were linked to plant responses, secondary metabolism synthesis, and plant growth and development and had TATC-box, which is a part of gibberellin responsive elements, MBS (which is a abbreviation for dehydration-responsive elements with abscisic acid and abscisic acid) and CRF stress responsive elements.

Physicochemical properties analysis shows that *MedWRKY*70 had the highest value, with 816 amino acids (aa) and *MedWRKY* 112 has the minimum with 110 amino acids(aa) and has molecular weights(mw) of 12.14 kDa along with 88.78 kDa. Aliphatic index (AI) is between 36.46 (*MedWRKY*64) to 79.95 (*MedWRKY*43), *MedWRKY* isoelectric points (PI) had an alkaline mean of 7.38 and varies from 4.91 (*MedWRKY*4) to 9.99 (*MedWRKY*60. Of the whole *MedWRKY* family of the proteins, the nucleus was their major preferred place for their functions. Therefore, the nucleus is the place they probably served best. The hydrophilic characters for these 126 MedWKRY proteins were described as the negative grand average of hydropathicity (GRAVY). Only *MedWRKY*54, *MedWRKY*82 and *MedWRKY*96 proteins had lower instability indices. Gene replication events were the primary mechanism for the evolution of the *MedWRKY* gene, according to the collinearity study. The ripe fruit and roots of *M. dodecandrum* appeared to have greater expression levels of *MedWRKY* genes, according to the RT-qPCR and transcriptome data analysis [18].

In the case of *Asparagus***,** according to the findings corresponding to the data implying that altogether 70 genes are scattered in 10 chromosomes and one remained unidentified chromosome. Transcriptome analysis identified 11 *WRKY* subgroups (C1–C9, U1, U2), among which the conserved *WRKY* subgroups were specifically found in *Arabdiopsis thaliana*. The total 11 elements, divided into the abscisic acid responsiveness elements (232), defense and stress (29) responsiveness elements were distributed over 70 *AoWRKY* gene superfamily members,51 gibberellin-responsive elements,124 light-responsive elements, 25 auxin-responsive elements, fifty four MYB binding sites are engaged into drought conditions and stresses,60 salicylic acid responsiveness elements, six wound responsive elements,129 anaerobic induction elements ,41 low temperature responsiveness elements along with 294 MeJA (Methyl Jasmonate) regulatory responsive elements. The *WRKY* gene's family

members in *Asparagus officinalis* was identified and characterized genome-wide as the main focus of the study's analysis. The *WRKY* gene's family was then through examined in study using whole genome and salt-stress (RNA-seq) transcriptomic data. Significant new information about the regulation and function of *WRKY* genes in asparagus under salt stress was provided by the results. Analysis of transcriptome data revealed the critical role that *WRKY* family genes play in controlling plant development and growth in salt-stressed environments. The *WRKY* genes' responses to salt stress were shown to cause both upregulation and downregulation of gene expression [19].

Volcanic maps which were created through tbtools, which showed that three and fifteen *AoWRKY* genes were up regulated as well as downregulated in NI&NI+S along with AMF&AMF+S.

The physicochemical properties revealed that: protein sequences of *WRKY*'s with the highest number amino acid(aa) along with molecular weight (MW) was *AoWRKY32* (674amino acid and 72,655.57Da), structure instability are identified and examined by an instability index value more than 40. The remaining *AoWRKY* sequence members are all unstable except *AoWRKY*11. Out of *AoWRKY*'s family only, *AoWRKY* 44 founded in the chloroplast region, while 22 present in the extracellular space and 47 in the nucleus. *AoWRKY*4 and *AoWRKY*34's theoretical isoelectric point(pI) along with molecular weight (MW) were not predictable, while the PI (theoretical isoelectric point) of other *AoWRKY*'S genes which varied from 4.66 (*AoWRKY*29) to 10.24 (*AoWRKY* 38). From all total number of protein sequences, 6 sequences were comparatively neutral, 33 were alkaline and 29 were acidic in nature.

The same subfamily's *AoWRKY* genes were found to have nearly identical conserved motifs upon analysis, indicating that they were involved in related regulatory processes. While motif15 is exclusive to the C5 subfamily, motif one and motif two are present nearly in all the *AoWRKY* genes. The only domain owned by the *AoWRKY*61 in C3 group is the *WRKY* superfamily.

The *AoWRKY* family has 7conserved domains and all *AoWRKY* family have *WRKY* domains. Only *WRKY* domains are present in C8, U1, and U2, C5 and C7 but plant Zn clust domains are also present in C6. *AoWRKY* 29 is located in C4 and *AoWRKY* 59 is present in C1 [19].

Additionally, an analysis was also addressed that performed the comparison of gene-wide analysis as well as identification of *WRKY* transcriptional factors drowning in 2 Asian legume

crops. Adzuki bean is known scientifically as *Vigna angularis* and Mung bean is known scientifically as *Vigna radiata*. Mung bean had a greater number of *WRKY* proteins, which were 91 proteins with 71 W-box binding domains when comparing to Adzuki bean 85 proteins with 71 W-box binding domains. A phylogenetic tree analysis showed that genes can be divided in approximately 3 groups: 15 in Adzuki bean, 16 in Mung bean in Group-I, 56 found in both species in Group II as well as in 13 in both species in Group-III.

The number of genes in the protein distribution was relatively similar in both species. As shown in Table S2, the pI values ranged from 4.74 for *SaWRKY*55, 9.99 for *SaWRKY*57 in the case of Mung bean, whereas in Adzuki bean, the pI value ranged from 4.96 for *SaWRKY*73 to 9.99 for *SaWRKY*57 The amino acid sequence comparison indicating that *VrWRKY* and *VaWRKY* proteins had an average length of about 340 residues, and *VrWRKY*7 is the largest length protein with 746 aa.

Exon/intron analysis of *VaWRKY* genes showed that all had two to five introns while Vr *WRKY* genes had two to six introns. Two IIIc members have exactly 2/2 introns. *VaWRKY* members in group IIIc have 0–3 introns, and *VaWRKY*36 (Vang08g01570) does not have any introns which are an example restricted. In contrast, the group IIc *VrWRKY* gene has one to five introns. The *VaWRKY* genes in Group IId have two and five introns, and the *VrWRKY* genes have two and three introns with less variation. Similarly, 1-5 and 2-4 were the introns of *VaWRKY* and *VrWRKY*. Mutiple sequence alignment (MSA) of the *VaWRKY* as well as *VrWRKY* domain shows that the conserved *WRKYGQK* hepta-peptide has mutation in W, R as well as Q aa. Phylogenetic analyses of conserved *WRKY* domain identified eight clade for its members in *VaWRKY* as well as *VrWRKY*, respectively.

Furthermore, promoter analysis of a collection of 17 *VaWRKY* genes and 18 *VrWRKY* genes revealed significant numbers of cis-regulatory elements in reaction to biotic or envirinmemtal as well as abiotic stress. According to this analysis, genes have an essential roles in stress tolerance mechanisms, which may open up opportunities for genetically modifying agronomic traits in related crops like mung beans and Adzuki beans [20].

The elements indicated the occurrence of stress-inducible in homologs even in Mung bean and Adzuki bean as homologs in Rice (*Oryza sativa*) and *Arabdiopsis thaliana*. Stress-responsive

regulatory mechanisms are conserved, as evidenced by the identification of stress-inducible elements in both Adzuki bean as well as Mung bean, such [21].

In addition, the *WRKY* gene identification and classification of *Glycine max* were analyzed. In these analysis, it was reported that sum total of 188 *WRKY*'s gene were analyzed in soybean, which furthur can be grouped into total three types, referred to as Group one, Group two, and Group three, with Group two containing five subtypes, namely IIa-IIe. There were 130 sequences in group II which were further divided into subgroups 26 sequences formed group III. Thirty-two members from group I possessed a single N-terminal *WRKY* domain, including *GmWRKY*65 and -72. In soybeans, these genes are spread all throughout in all twenty chromosomes. *WRKY* gene family's expansion in the soybean species is mainly accomplished by duplication of the whole genomic sequence. In order to figure out *GmWRKY*s' roles in the conservation and classification of *Glycine max*, attributed motifs were predicted for putative *GmWRKY*s using MEME software. The result was 16 original motives after all investigation were done [22].

Soya bean *WRKY* genes and salt stress:49 *WRKY* genes were expressed in the plant's aerial sections, the majority of which were downregulated. Under salt stress, various patterns of expression were then shown by the detailed RT-qPCR analysis. From this analysis, twelve genes exhibiting no significant change, 35 decreased, and nineteen were induced. The majority of sixty-six *WRKY* gene in soybean roots that responded to salt stress were upregulated, according to RNA-seq analysis.

Functional characterization and gene expression: Though many *WRKY* genes have been identified, only a small number have been the subject of functional studies, suggesting a great deal of untapped potential for further investigation. And because of the genes sensitivity to salt stress, soybean crops may be genetically modified to increase their tolerance to salt [22].

To gain a better understanding of roles, played by the uncharacterized *WRKY*'s gene in soybean stresses responses, more functional studies are required. The results of the study add to our knowledge of plant physiology in general and stress responses in particular. According to these results, the *WRKY* transcription factors in soybeans are important for the plant's reaction to salt stress. Consequently, modifying these genes may help create soybean varieties that are resistant to salt. The cited literature bolsters the argument that additional

investigation is necessary to completely comprehend these genes' functions and potential uses in crop improvement.

Similarly, another study analyses identification as well as classification of *WRKY*-Genes in Chickpea. Results of this analysis shows that genome of the chickpea includes 70 non-redundant *WRKY*-encoding genes that are interestingly spread over all chromosomes with the exception of chromosome 8 [23]. Group I, group II, and group III are the three main groups into which these genes are further divided[24].

Additional information can be obtained from gene structure analysis using tools (GSDS). Chickpeas' major chromosomes contain the Car *WRKY* 56 gene, which is one prominent example. Its patterns of gene expression under stressful conditions might be affected by its lack of introns. The identification and characterization of chickpea's *WRKY* genes led to the discovery of more HD-Zip (I) family members, including CaHDZ12. The *WRKY* transcription factors' adaptive roles in the face of environmental challenges are highlighted by the abiotic stress responses associated with this gene [25].

In addition to advancing our knowledge of the biological functions of these genes, this systematic method of discovering and categorizing *WRKY* genes in chickpea highlights their potential to increase crop resilience to abiotic challenges. A detailed Grouping Group II is distinguished by its intricacy and is further separated into five subgroups, which together represent the diversity found in this family [26], namely IIa, IIb, IIc, IId and IIe. This thorough categorization facilitates comprehension of the subtle structural variations in these proteins and lays the groundwork for future research into their functional roles in stress reactions.

Chickpea gene duplication research revealed that segmental duplications played a major role in the growth of *WRKY* genes, and purifying selection was observed during the evolution of these gene families [27]. The functional diversification and adaptation of *WRKY* genes in response to environmental stressors are largely dependent on this evolutionary mechanism.

Under salt, drought, and cold stress conditions in chickpeas, in silico transcriptome data analysis showed differential expression of *CarWRKY* genes in root and shoot tissues. Abiotic stress-response pathways that are conserved may be associated in the expression patterns of several *CarWRKY* genes, which showed consistent expression patterns under all stress settings [27]. In both susceptible and tolerant chickpea genotypes, drought stress

markedly elevated the expression of *WRKY* genes, highlighting the significance of these genes for drought resistance.

Differential regulation during stress circumstances was revealed by analyzing the transcript levels of chickpea *WRKY* group-III genes under pathogenic stress as well as treatments with abscisic acid, jasmonic acid, and salicylic acid [28].

The plant nucleus contains the gene *CaWRKY*50, which binds to the W-box and has a C-terminal transactivation domain. It is activated by both *Ascochyta rabiei* infection and salicylic acid therapy. Early flowering as well as senescence were examined in tobacco plants with overexpressed *CaWRKY*50, highlighting the significance of this protein for plant growth and stress responses [29].

Some analysis shows expression analysis under biotic along with abiotic stress and identification of *WRKY* genes such as in barley (*Hordeum vulgare* L.). As a crucial crop for food, feed, and brewing, barley's resilience and adaptability are of paramount importance, especially considering its natural tolerance to challenges such as drought, salinity, and fungal diseases [30]. Through genome-wide analysis, we found 86 candidate genes in this study that contain the *WRKY* domain. Based on their location on the barley chromosome, we named these genes *HvWRKY*1 through *HvWRKY*86. Seven linkage groups of barley correspond to 82 of the 86 *HvWRKY* genes, according to the results of chromosome mapping. Of them, chr 3 had greatest number of *HvWRKY* genes—21—while chr 6 had the lowest numbers just five.

Furthermore, it was not possible to precisely map four genes to any linkage group. The *HvWRKY* proteins were categorized into three groups according to the kind of zinc-finger motifs and the quantity of *WRKY* domains. By using MEME motif analysis, ten motifs in total were found in the barley *WRKY* proteins. With the exception of seven genes, the majority of *HvWRKY*s had at least one intron, according to gene structure analysis, which also revealed that the number of exons varied from two to seven. With three exons, a *HvWRKY* gene accounted for 47.7% of the total, making it the most common type. The genes encoding group I and subgroup IIb proteins had the greatest number of introns, according to the statistical analysis of the number of introns in each group.

Additional context for the function and evolution of barley *WRKY* genes has been provided by comparative analysis with other species, including rice and *Arabidopsis*. Comparative expression analysis provides additional evidence for the similarity in gene function between species suggested by orthologous relationships. Understanding the conserved and distinct facets of *WRKY* gene functions in barley, especially in connection to stress responses and developmental processes, depends on these comparisons [31].

Comprehensive gene expression profiling analyses that when 15 *HvWRKY* genes were subjected to salt, cadmium, or drought stress, it was discovered that the majority of these genes were responsive to a variety of abiotic stresses [32].

Expression Specific to Genotype: Genes such as HvDRFI.3, HvCBF6, HvCO5, and *HvWRKY*42 were significantly upregulated during drought, while HvDRF1.3 was specifically induced in a different genotype.

Developmental Stage Specificity: Different sets of non-redundant genes were found at both the vegetative and reproductive stages of barley's response to drought stress, indicating developmental stage-specific gene roles [33].

Transcription factors and stress tolerance: Transcription factors such as *HvWRKY*38, HvDREB1/CBF and HvNAC6 have been cloned and characterized from barley, showing potential for enhancing stress tolerance in transgenic approaches. The efficient DNA-binding specificity of these barley transcription factors makes them promising candidates for enhancing abiotic stress tolerance in other crops.

Through transgenic techniques, the HVA1 gene from barley has been effectively utilized to increase abiotic stress tolerance in a variety of crops, including wheat, rice, and maize [34].

Some studies also say that *WRKY* genes throughout the entire genome in the Desert Poplar *Populus euphratica* were identified and that The *WRKY* Genes Adaptive Evolution Facing Salt Stress. The investigation into *WRKY* genes in this desert species of *Populus euphratica* provides valuable clues for understanding the molecular evolution of high salt tolerance mechanisms in this poplar. A total of 107 *PeWRKY* genes from the *P. euphratica* genome were

identified and phylogenetically analyzed with *WRKY* genes from *P. trichocarpa*, representing a less salt-tolerant poplar [35].

The main findings of the study included the following:

1. Ten *PeWRKY* genes were identified as specific *to P.euphratica*, and five of them were found to have differential expression levels under salt stress, suggesting that those genes possibly have roles to play in adapting to a saline environment.

2. Two pairs of orthologous *WRKY*-genes between *P.euphratica* as well as salt-sensitive *P. trichocarpa* were detected as experiencing positive evolution, indicated by dN/dS ratios exceeding 1. These genes also significantly change expression in reaction to salinity stresses. The study suggests the genesis of new genes along with the adaptive evolution of some pre-existing orthologs has been critical in *P.euphratica* acquiring high salt tolerance [36].

Evolutionary Implications

These results provide evidence that selection pressure has led to the attainment of high salt endurance in *P.euphratica*. This study therefore lays a strong basis for understanding gene family expansion and functional divergence response to environmental stresses, and especially to the major problem of salinity stress threatening the survival and productivity of plants [35].

Gene expression response to salt stress. Certain *WRKY*'s significantly upregulated in acknowledgement to salt stress, as shown by a deep transcriptome sequencing, underscoring their function in abiotic stress response mechanisms. The *P. euphratica* genome was found to have expanded gene families such as Myb, ERF, Bzip and *WRKY*, which are associated with abiotic stress response and are heavily involved in gene regulation under salt stress [36].

Significance of plant breeding and conservation

Breeding techniques for salt-tolerant poplar and other related species can benefit from an understanding of genetic basis of salt tolerance in *P.euphratica*. The findings of this study could help genetically modify cultivated poplars to grow in saline soils, which would be

advantageous for afforestation and reforestation initiatives in areas affected by saline and dry conditions [37].

All things considered, conservation and plant breeding are essential for a sustainable future. Together, these approaches can guarantee that we have the resources required to feed everyone on the planet and keep it healthy.

CHAPTER 3-MATERIALS AND METHODS

Flowchart of *WRKY* gene identification

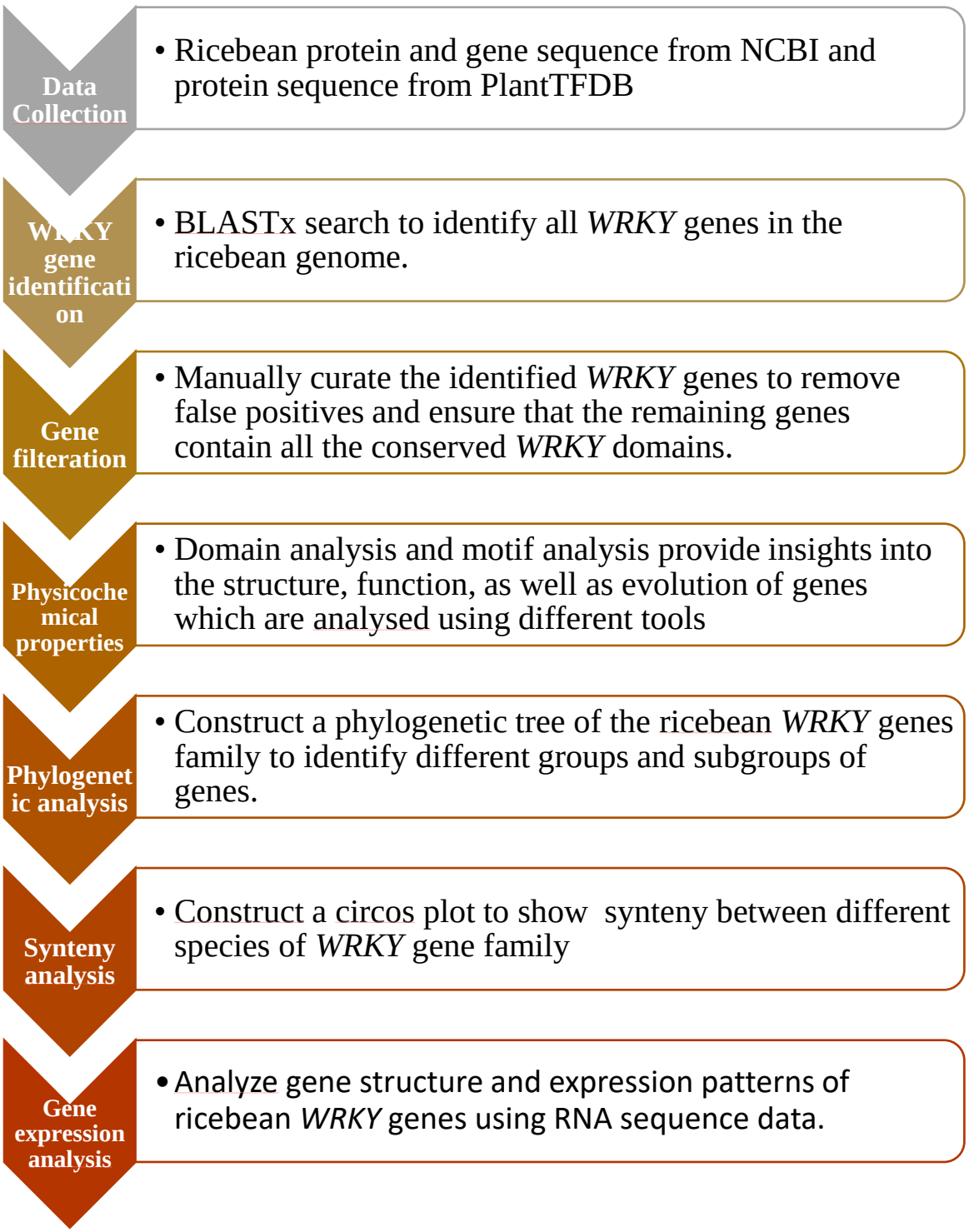


Figure 3.1 Flowchart of *WRKY* gene identification

3.1 Database sequence searches

The National Center for Biotechnology Information (NCBI)(https://www.ncbi.nlm.nih.gov/) provided the ricebean protein and gene sequence. PlantTFDB website (http://planttfdb.gao-lab.org/) provided *WRKY*'sprotein sequences of many plants, including *Glycine max, Cicer arientinum, Vigna radiata, Vigna angularis, Vigna unguiculata and Arabdiopsis thaliana*. BLASTx was used to find for all members of the *Vigna umbellata WRKY*'s family using protein sequences *from Arabdiopsis thaliana, Vigna radiata, Vigna angularis, Vigna unguiculata, Glycine max, and Cicer arietinum*. E-value (Expected value) was set at a parameter of 1e-10 in order to get produce precise results.

Figure 3.2 *WRKY* genes in other 6 species

3.2 Domain analysis

In genome-wide investigations, domain analysis refers to the process of locating and describing conserved protein domains within gene sequences. These domains are the structural along with functional components of the proteins that carry out particular tasks such enzymatic activity, protein-protein interactions, and DNA binding [38].

The protein family database (Pfam) has the file for the structural conserved domains (PF03106) and (PF10533) of the *WRKY*'s. The *WRKY* gene in rice beans is then questioned using HMMER 3.0. All of these genes have the intact *WRKY* structural domain, according to the Pfam analysis, which also removed the sequences that did not have the *WRKY* domain.

3.3 Physicochemical properties

Analysis of molecular weight, length of protein, theoretical index (PI), instability index, gravy score, aliphatic-index (AI), chromosomal location of *WRKY* gene family of ricebean is done using protparam tool(https://web.expasy.org/protparam/). Analysis of all *WRKY*'s physicochemical properties is essential in understanding the stability, interactions, and potential functions of these proteins in healthy and malformed tissues, as well as in different developmental stages or under stress conditions. By examining these characteristics scientists can learn important details regarding the makeup of *WRKY* genes and how they function in different biological systems [39].

3.4 Structures of *WRKY* genes

Exon/Intron structure

Examination of the structure of exons as well as introns has had studies on how genes are organized and characterized. This includes an analysis of the characteristics and organization of exons, that is, the coding regions, and introns, which are non-coding regions in genes. This will provide clear understanding of the evolutionary history and the gene regulatory and functional diversity of genes in species. Researchers can discover patterns in exon and intron length and quantity and its sequence for explanation of gene expression, alternative splicing, and protein diversity [40]**.** The exon along with intron structures of ricebean *WRKY*'s genes weas then examined by the GSDS2.0 website tool(http://gsds.cbi.pku.edu.cn/index.php) from Center for Bioinformatics at Peking University.

Motif analysis

Using motif analysis, conserved protein motifs found in the *WRKY*-gene are identified and characterized. The *WRKY* gene family's conserved motifs and variants highlight the structural along with functional diversity of *WRKY*-transcription factors. The analysis demonstrates the possible function of *WRKY*'s in plant growth, development, and stressful responses as well as their evolutionary relationships and regulatory mechanisms [41]. Using the maximum output

of five motifs as a parameter, the conserved motifs were found by using (MEME) website (http://meme-suite.org/index.html).

3.5 Phylogenetic analysis

Phylogenetic analysis helps in providing insights into the evolutionary relationships and development of different species or groups of organisms. Here, phylogenetic analysis would be used to study genetic data to know how different species are related or not related to each other over time. Through the construction of phylogenetic trees, scientists can visualize and interpret the evolutionary history of organisms, further identifying common ancestors and understanding patterns in genetic changes that have occurred during evolution [42]. For multiple sequence alignment, it is done using MAFFT (mafft.cbrc.jp - MAFFT alignment and NJ / UPGMA phylogeny) for obtained protein sequences from *Vigna umbellata* and from other species including (*Glycine max, Cicer arietinum, Vigna radiata, Vigna angularis, Vigna unguiculata and Arabdiopsis thaliana*) taken as query. Then the phylogenetic tree was made, visualized, and annotated using iTOL (Interactive Tree Of Life).

3.6 Synteny analysis

In order to examine how gene order and orientation are conserved among various species or genomes, a basic technique in bioinformatics is synteny analysis. In order to find areas of conserved synteny, which can reveal information about evolutionary relationships, genome structure, and functional conservation, it compares how genes or genetic elements are arranged in the genomes of related organisms [43].

The synteny analysis is performed with Tbtools, and the multiple chromosome layouts file, gene links files and GFF files between *Vigna umbellata* and other query *species Arabdiopsis thaliana, Vigna radiata, Vigna angularis, Vigna unguiculata, Glycine max, and Cicer arietinum* are obtained by performing One Step MCScanx tool with the default parameters. The collinear relationships amongst the genomes of *Vigna umbellata* and other query species are then visualized using these files through the use of advanced circos in tbtools.

3.7 Analysis of cis-acting regulatory elements

Cis acting regulatory elements are DNA sequences that work on same DNA molecule to control the transcription of genes that are close by. Promoters, enhancers, silencers, insulators, and other regulatory modules are examples of cis-acting elements. They bind transcription factors and other regulatory proteins to the same DNA strand in order to control the expression of a gene. They differ from diffusible substances (often proteins) known as trans-acting regulatory elements, which have the ability to control genes found in other regions of the genome [44]. Tbtools is used to create heatmap interpreting and analyzing data which is obtained from (http://bioinformatics.psb.ugent.be/webtools/plantcare/html), the plantCARE website.

3.8 Expression analysis using RNA-seq data

The study of gene expression patterns analysis to determine the location, timing, and intensity of a given gene's activity is known as expression analysis. This analysis sheds light on the functions of genes in various biological processes, including disease, development, and reactions to environmental cues.

In this**,** RNA-Sequencing (RNA-Seq) data was used and analysed because, it makes the transcriptome more readable and makes it possible to identify new transcripts and splice variants.

The expression analysis using RNA-seq data of the obtained *WRKY* genes were analysed and interpreted by heatmap created by tbtools. In molecular biology and genetics, expression analysis is an essential tool that helps researchers decipher the intricate regulatory networks that underpin biological processes. It offers insightful information about how genes work and what roles they play in different physiological and pathological situations and in developmental stages.

CHAPTER 4-RESULTS

4.1 Identification of *WRKY* gene family in ricebean

After manual curation, filtration and removal of duplication, total 84 *WRKY* genes of ricebean were obtained after performing blastX. Filtrations process is done on the basis of Bit-score, percent identity and E-value (Expected value).

4.2 Domain analysis

According to the analysis conducted using Pfam database of obtained 84 ricebean *WRKY* genes, it was identified and analyzed that all the *WRKY* gene contained the intact *WRKY* domain (PF03106) and (PF10533), the non *WRKY* domains were 5 in total.

And the non *WRKY* domains were removed. So, after removal of non *WRKY* domains and their sequences, total 79 *WRKY* genes of ricebean were obtained.

4.3 Physicochemical properties

Each ricebean *WRKY* protein is thoroughly analyzed in this table, with the molecular weight, amount of amino acids(aa), theoretical PI, stability index, gravy score, aliphatic index (AI), and chromosomal location all being taken into consideration. Figure 4.1 shows that the *WRKY* gene in ricebean has lengths ranging from 121 (LOC124844535) to 746 (LOC124824305) amino acid, theoretical PI ranging from 4.3 (LOC124844535) to 11.36 (LOC124831060), aliphatic index ranging from 42.54 (LOC124823866) to 84.42 (LOC124835419), molecular weight ranging from 12907.79 kDa (LOC124844535) to 80872.58 kDa (LOC124824305) and instability index ranging from 27.59 (LOC1248358590) to 77.38 (LOC124824607).

Table 4.1 Physicochemical properties of *WRKY* gene

LOC ID	NO. OF AMINO ACID	MOLECULAR WEIGHT	THEORITICAL PI	INSTABILITY INDEX	GRAVY SCORE	ALIPHATIC INDEX
LOC124826558	289	32881.81	5.52	57.51	-0.772	67.54
LOC124836200	238	25993.52	9.58	47.45	-0.512	66.01
LOC124822494	234	26368.48	8.96	44.81	-0.822	64.23
LOC124823659	327	35318.95	9.94	51.22	-0.576	63.91
LOC124824305	746	80872.58	5.85	51.71	-0.822	54.53
LOC124831756	588	63915.83	5.95	48.05	-0.72	59.46
LOC124833308	192	21288.7	9.38	38.03	-0.892	53.23
LOC124833741	387	42818.18	6.25	39.53	-0.931	50.9
LOC124838521	570	61258.1	7.2	48.87	-0.605	63.72
LOC124840356	431	47218.19	5.36	60.15	-0.662	59.56
LOC124840347	431	47218.19	5.36	60.15	-0.662	59.56
LOC124841641	351	38714.25	5.86	61.28	-0.981	44.26
LOC124841879	163	19117.74	5.04	37.69	-1.331	34.54
LOC124843453	168	17658.14	4.39	46.13	-0.544	55.71
LOC124848213	586	63772.83	6.31	50.85	-0.73	62.44
LOC124820436	570	62005.42	6.27	45.14	-0.788	56.98
LOC124821399	593	63859.27	5.55	48.68	-0.718	61.08
LOC124823866	539	60202.72	8.04	62.53	-1.043	42.54
LOC124824197	361	40957.55	5.32	50.36	-0.764	63.8
LOC124824607	282	32648.43	5.94	77.38	-0.839	67.77
LOC124824985	286	32523.09	5.54	70.71	-0.834	62.73
LOC124825560	518	56584.97	7.3	56.11	-0.766	60.08

LOC124826241	**292**	**31869.57**	**6.1**	**61**	**-0.678**	**56.2**
LOC124826279	**228**	**25945.16**	**9.11**	**55.68**	**-0.688**	**57.28**
LOC124826358	**150**	**16774.75**	**9.66**	**40.2**	**-1.057**	**58.47**
LOC124826423	**402**	**43717.3**	**5.83**	**61.59**	**-0.77**	**48.88**
LOC124826474	**695**	**76023.13**	**5.99**	**48.3**	**-0.752**	**60.89**
LOC124826753	**328**	**35627.71**	**9.61**	**43.07**	**-0.509**	**71.4**
LOC124827292	**304**	**34368.47**	**5.6**	**41.03**	**-0.697**	**61.91**
LOC124827612	**397**	**43396.66**	**6.62**	**48.13**	**-0.837**	**54.11**
LOC124829436	**527**	**57705.5**	**6.95**	**50.72**	**-0.599**	**65.71**
LOC124831060	**279**	**31428.01**	**11.36**	**64.23**	**-1.012**	**54.87**
LOC124832601	**492**	**54423.56**	**8.62**	**44.62**	**-0.817**	**59.07**
LOC124833946	**345**	**37647.64**	**5.64**	**64.77**	**-0.621**	**56.55**
LOC124833937	**183**	**21024.86**	**9.19**	**34.01**	**-0.779**	**67.6**
LOC124835435	***273***	**30726.64**	**8.35**	**55.76**	**-0.659**	**66.7**
LOC124835419	**269**	**30487.74**	**8.35**	**52.17**	**-0.594**	**84.42**
LOC124835536	**324**	**36484.77**	**6.86**	**46.54**	**-0.691**	**63.24**
LOC124835911	**318**	**35483.94**	**8.83**	**40.11**	**-0.795**	**62.8**
LOC124835859	**154**	**16893.58**	**9.62**	**27.59**	**-0.383**	**74.74**
LOC124836286	**126**	**15191.2**	**9.69**	**54.84**	**-1.12**	**54.05**
LOC124836452	**240**	**27108.63**	**9.11**	**64.85**	**-0.701**	**60.08**
LOC124836539	**277**	**31027.68**	**7.08**	**40.31**	**-0.71**	**68.95**
LOC124837079	**317**	**35773.57**	**6.14**	**58.82**	**-0.883**	**53.72**
LOC124837553	**299**	**32119.28**	**8.83**	**51.34**	**-0.389**	**65.28**
LOC124837667	**503**	**54822.77**	**9.22**	**55.56**	**-0.852**	**62.6**
LOC124838075	**471**	**52063.36**	**6.39**	**45.74**	**-0.654**	**67.6**

LOC124838177	**270**	**30122.43**	**5.07**	**60.47**	**-0.847**	**58.11**
LOC124838807	**269**	**28759.71**	**9.9**	**59.15**	**-0.565**	**59.89**
LOC124838992	**294**	**33824.86**	**7.15**	**58.9**	**-0.963**	**50**
LOC124839422	**290**	**33330.59**	**6.76**	**53.59**	**-0.644**	**71.31**
LOC124839419	**298**	**33875.9**	**5.62**	**59.7**	**-0.647**	**68.36**
LOC124840695	**361**	**40667.34**	**4.88**	**54.75**	**-0.654**	**61.77**
LOC124841415	**573**	**62605.52**	**6.95**	**62.77**	**-0.867**	**49.2**
LOC124841446	**320**	**35153.8**	**9.66**	**54.51**	**-0.658**	**63.72**
LOC124841697	**222**	**25320.94**	**7.57**	**42.86**	**-0.612**	**79.46**
LOC124842044	***293***	**32737.64**	**5.73**	**59.74**	**-0.635**	**65.22**
LOC124842421	**140**	**16178.01**	**8.85**	**37.07**	**-0.971**	**61.21**
LOC124842639	**217**	**24893.43**	**6.71**	**49.45**	**-0.917**	**55.21**
LOC124842646	**164**	**18567.49**	**6.42**	**41.09**	**-0.829**	**62.99**
LOC124842987	**487**	**53789.95**	**6**	**48.23**	**-0.759**	**64.52**
LOC124843167	**369**	**40352.29**	**6.96**	**56.49**	**-0.824**	**49.73**
LOC124843282	**304**	**33962.15**	**9.08**	**50.63**	**-0.907**	**56.41**
LOC124844535	**121**	**12907.79**	**4.3**	**42.08**	**-0.627**	**55.54**
LOC124844712	**320**	**35881.43**	**6.25**	**57.03**	**-0.966**	**45.47**
LOC124844834	**352**	**39697.25**	**6.35**	**56.74**	**-0.742**	**59.83**
LOC124844880	**354**	**39778.89**	**9.74**	**51.45**	**-0.8**	**66.64**
LOC124844927	**555**	**61150.7**	**7.58**	**46.77**	**-0.474**	**72.05**
LOC124845289	**371**	**41768.66**	**5.39**	**50.27**	**-0.611**	**66.5**
LOC124847246	**125**	**14040.17**	**9.28**	**42.24**	**-0.673**	**60.88**
LOC124848095	**461**	**50510.45**	**7.22**	**53.01**	**-0.584**	**69.02**
LOC124848283	**706**	**78686.37**	**5.31**	**37.07**	**-0.812**	**67.03**

LOC124849084	**496**	**53694.01**	**5.69**	**54.3**	**-0.709**	**53.55**
LOC124819734	**146**	**15796.56**	**6.57**	**51.82**	**-0.436**	**60.21**
LOC124820208	**528**	**57905.28**	**6.39**	**57.84**	**-0.874**	**50.97**
LOC124820330	**353**	**39325.79**	**5.47**	**60.43**	**-0.622**	**69.89**
LOC124820402	**255**	**29293.39**	**4.93**	**58.53**	**-1.054**	**55.84**
LOC124820768	**255**	**28069.34**	**5.54**	**66.96**	**-0.724**	**55.49**
LOC124821552	**511**	**55235.72**	**6.49**	**51.67**	**-0.521**	**69.75**

4.4 Analysis of structure of *WRKY* genes

Exon/Intron structure

The variety of *WRKY*'s gene-structures can reveal information about evolutionary background of *WRKY* gene family members. Thus, as illustrated in figure 4.2, we examined the exon/intron distribution as well as the total amount of coding exons for every ricebean *WRKY* gene family. Within the exon/intron distribution, the total number of coding exons for the ricebean *WRKY*'s genes ranges between2 to 10. The majority of *WRKY* genes had three coding exons, six genes had varying numbers of UTRs (ranging from two to three), and there were between two to ten introns.

Only six genes has 2 to 3 UTRs regions (untranslated region) and the remaining genes does not have any UTRs. All 79 *WRKY* genes contains more than 2 introns and exons.

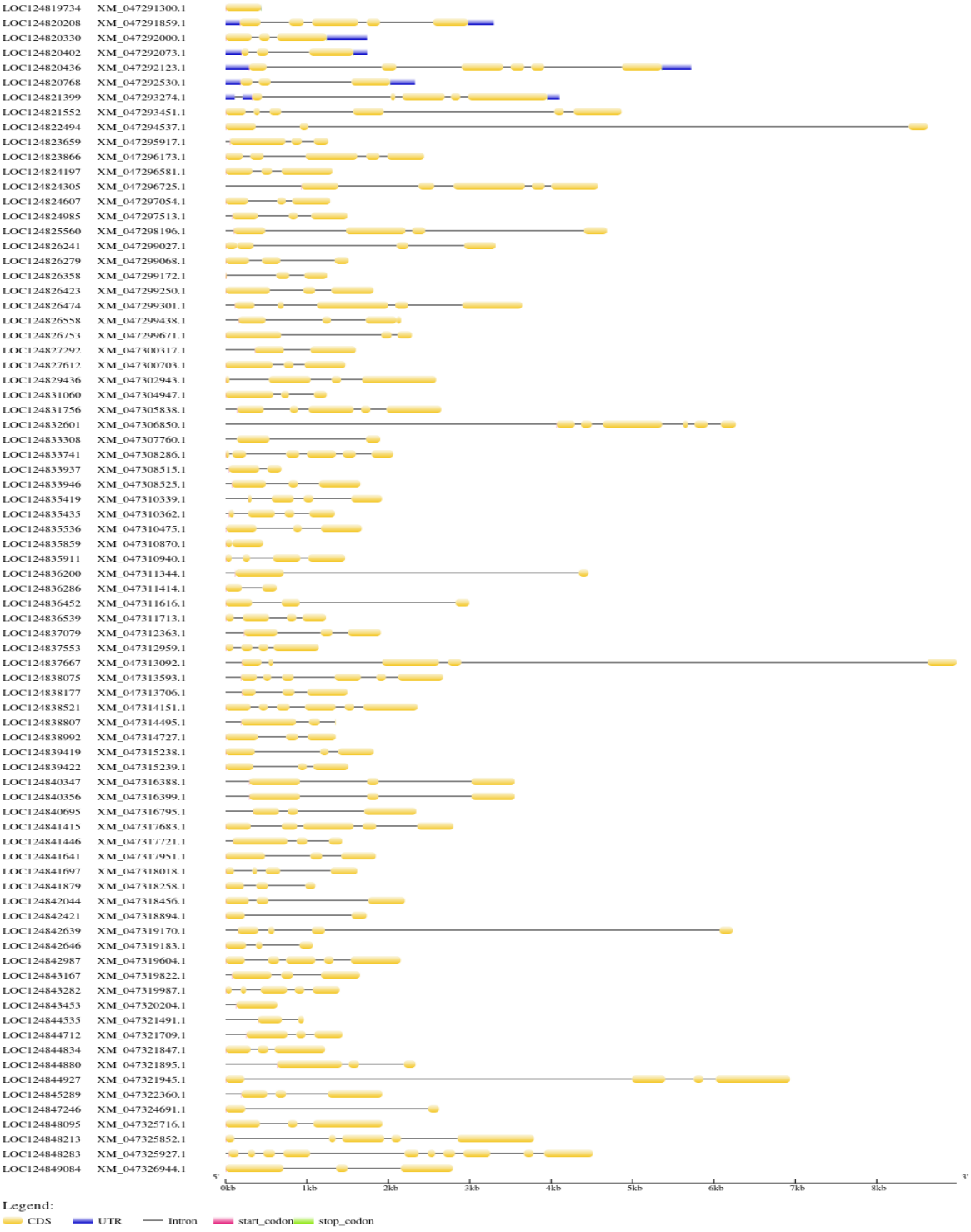
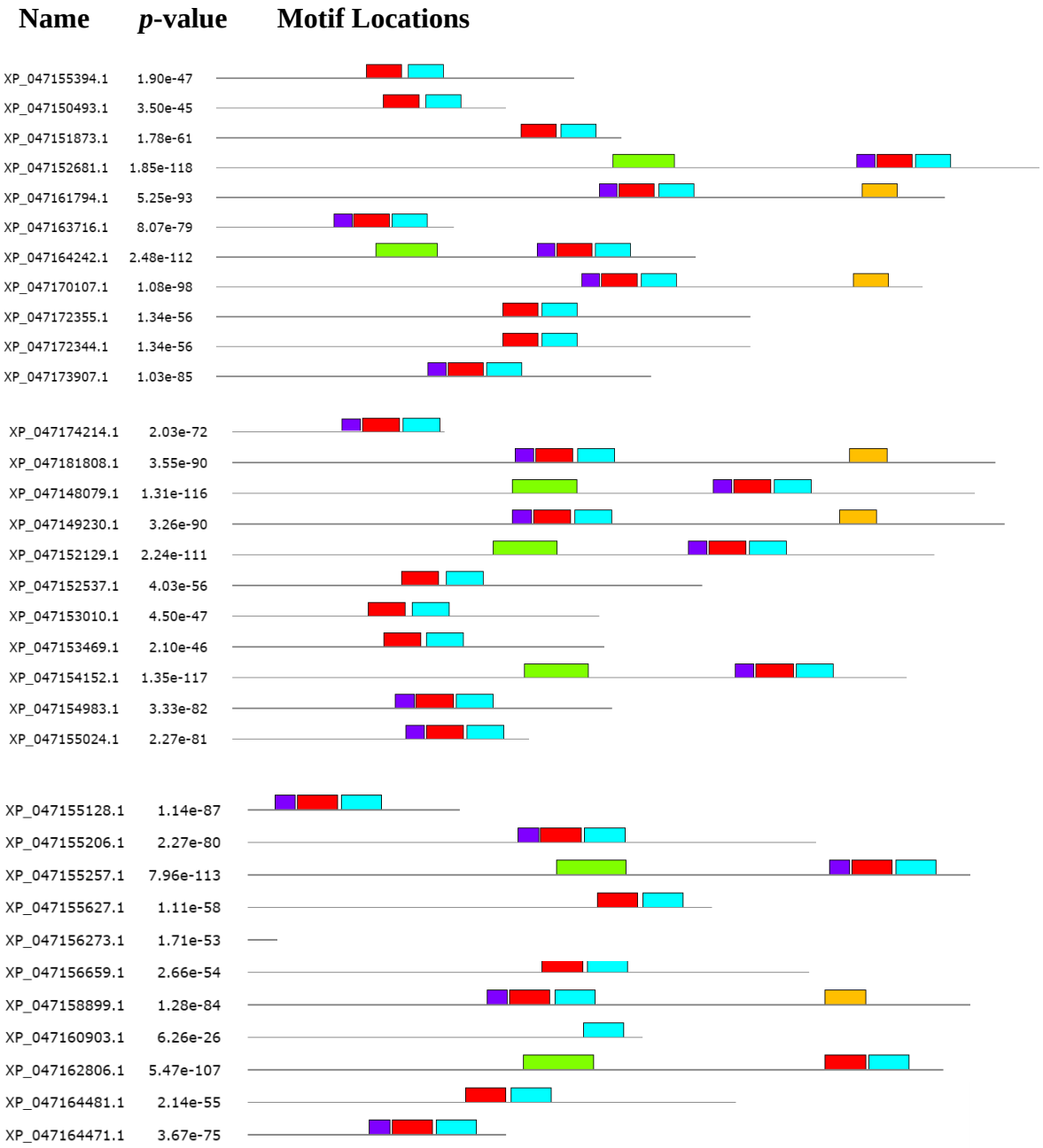


Figure 4.2. Exon/intron distribution: yellow colour depicts the exons (coding region), blue colour represents the UTRs and black represents the introns (noncoding region).

Motif analysis

Total 5 conserved motifs were investigated within *WRKY* genes by using the MEME website in order to acquire or get a better understanding of the diversity as well as similarity of gene motifs in various genes. Figure 4.3 and 4.4 depicts that among the 79 ricebean *WRKY* members, each gene contain 2 to 4 motifs. Motif 1 as well as motif 2 are present in almost all the genes.



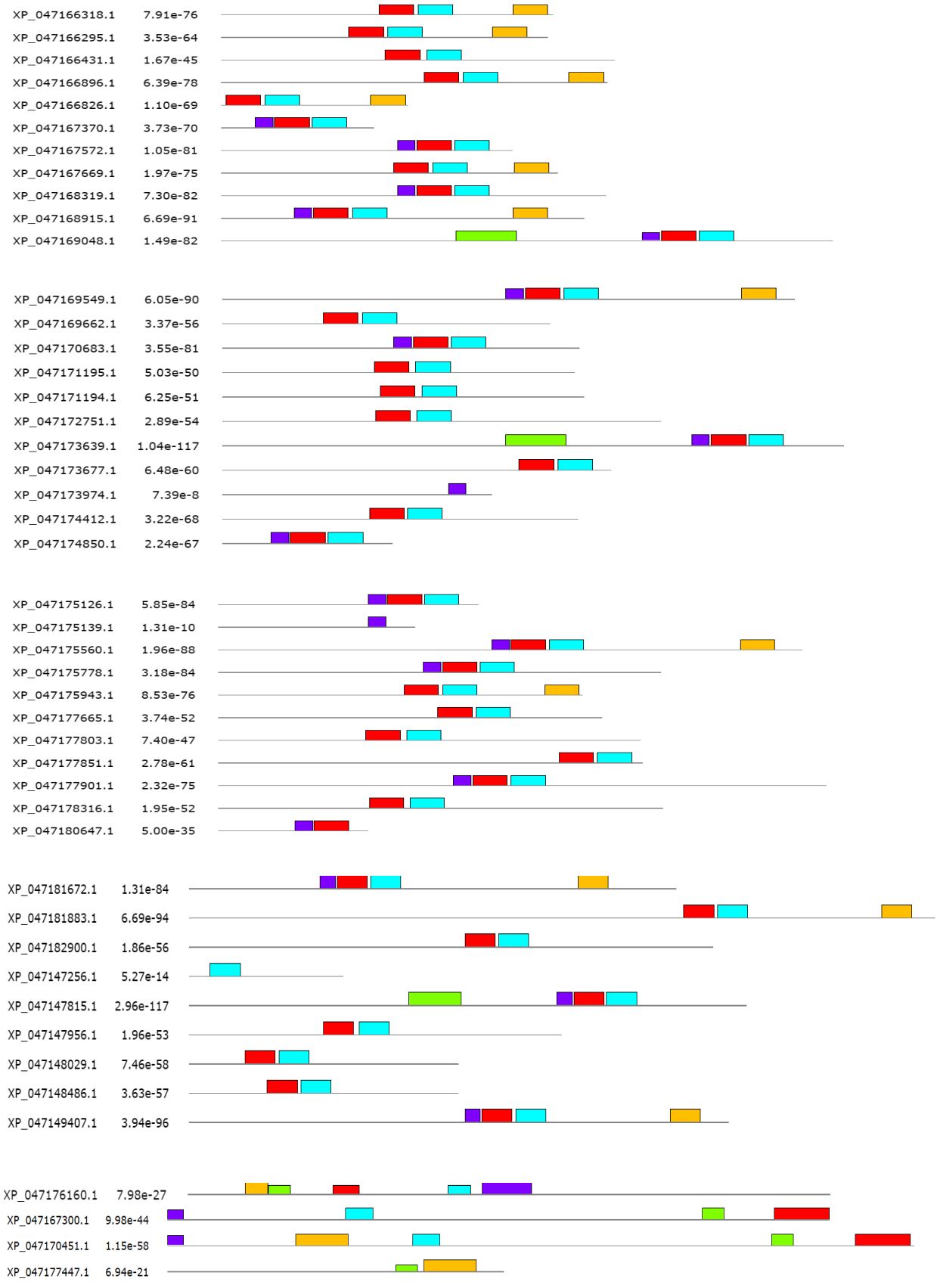
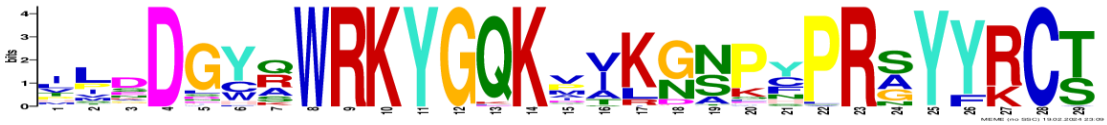
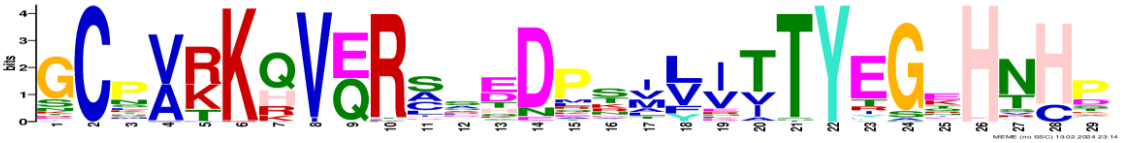


Figure 4.3 Motif locations

Motif Symbol	Motif Consensus
	1.ILDDGYQ*WRKY*GQKVVKGNPYPRSYYRCT
	2.GCPVRKQVZRSSEDPSILITTYEGEHNHP
	3.DGYN*WRKY*GQKQVKGSEYPRSYYKCTHPNCPVKKKVERSLDGQITEIIYK
	4.TVREPRVAVQTRSEV
	5.DLVEAATAAJTSDPNFTAALAAAISSIIG



Motif 1



Motif 2

Motif 3

Motif 4

Motif 5

Figure 4.4 Motif logos

4.5 Phylogenetic analysis

In order to perform multiple sequence alignment (MSA) analysis, protein sequences of *WRKY* genes of *Vigna umbellata* and other species of *Glycine max, Cicer arietinum, Vigna radiata, Vigna angularis, Vigna unguiculata and Arabdiopsis thaliana* that were used as queries were aligned using MAFFT0(mafft.cbrc.jp - MAFFT alignment) as well as NJ / UPGMA phylogeny.

Next, **iTOL**(Interactive Tree of Life) online tool was used to create, visualize, and annotate the phylogenetic tree. In figure 4.5(a)and (b), based upon the phylogenetic analysis examination ,79 *WRKY* genes was divided into 3 groups based on their conserved motifs: Group1 which consisted of 48 *WRKY* genes and has conserved motif1 and motif 2, Group 2 consisted of 14*WRKY* genes and has conserved motif 1, 2 and 4 and Group 3 consist of 17 *WRKY* genes which has conserved motif 1, 2, 3 and motif 5.

Overall, the study of gene evolution, gene family expansion, and the discovery of conserved functional domains among various species is aided by phylogenetic tree grouping based on conserved motifs, which offers insightful information on the evolutionary dynamics and functional implications of gene families.

Figure 4.5 Rooted Phylogenetic tree of WRKY proteins from Vigna umbellata along with other plant species Glycine max, Cicer arientinu, Vigna radiata, Vigna angularis,Vigna unguiculata and Arabdiopsis Thaliana . Here "LOC ID" represents WRKY genes of Vigna umbellata, "Glyma" represents Glycine max, "AT" represents Arabdiopsis Thalian, "XP" represents Cicer arietinum, "Vradi" represents Vigna radiata," Vang" represents Vigna angularis, "Vun" represents Vigna unguiculata.

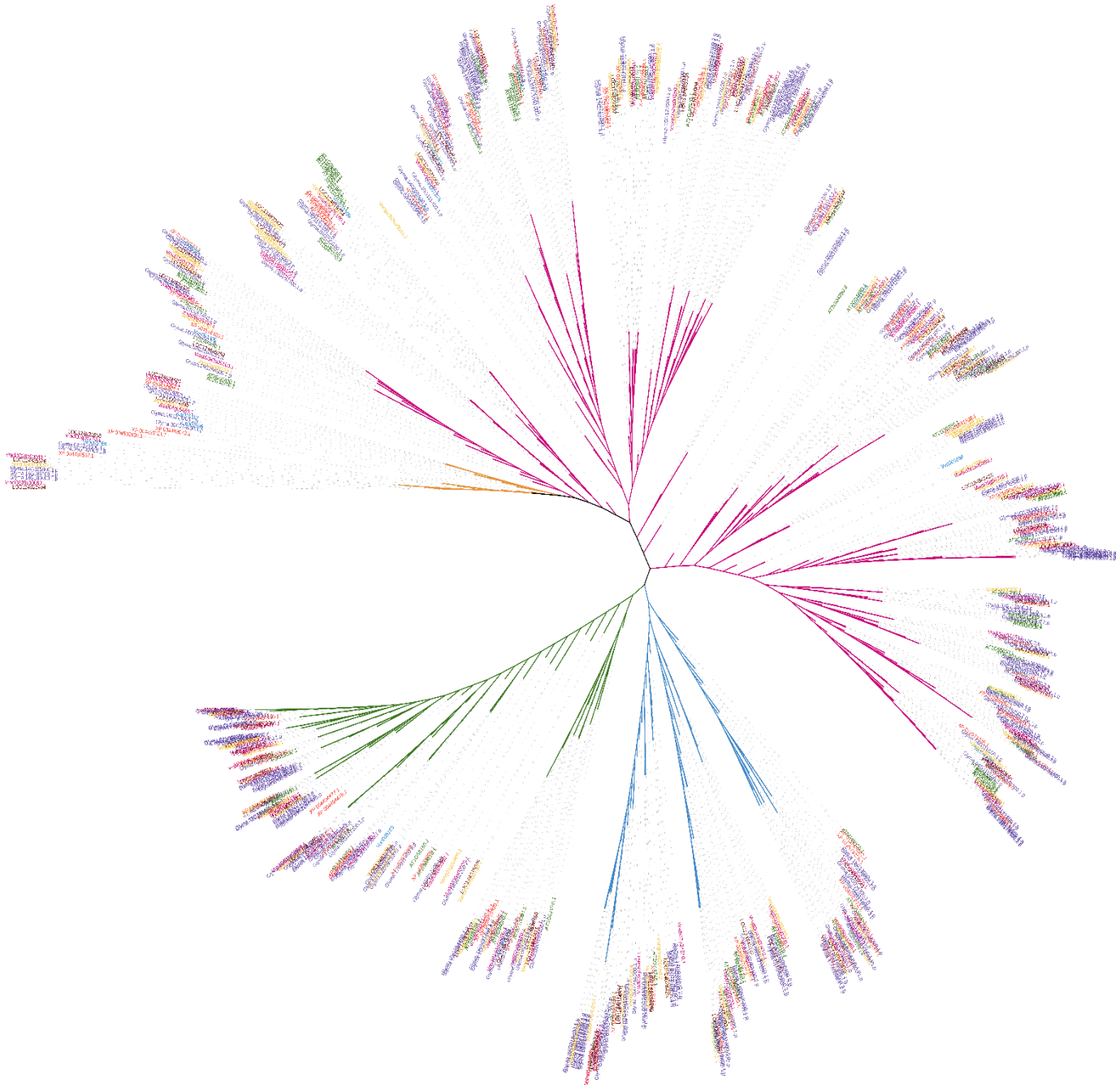


Figure 4.6 Unrooted phylogenetic tree

4.6 Synteny analysis

In this, we conducted a synteny analysis across genomes of *Vigna umbellata* with *Arabdiopsis thaliana, Vigna radiata, Vigna angularis, Vigna unguiculata, Glycine max and Cicer arietinum.* to identify conserved syntenic regions and orthologous gene pairs.

Figure 4.7 depicts maximum synteny of *Vigna umbellata with Vigna angularis and Vigna unguiculata* among all the other species which shows a closer evolutionary relationship and least synteny is between *Vigna umbellata* and *Arabdiopsis thaliana*

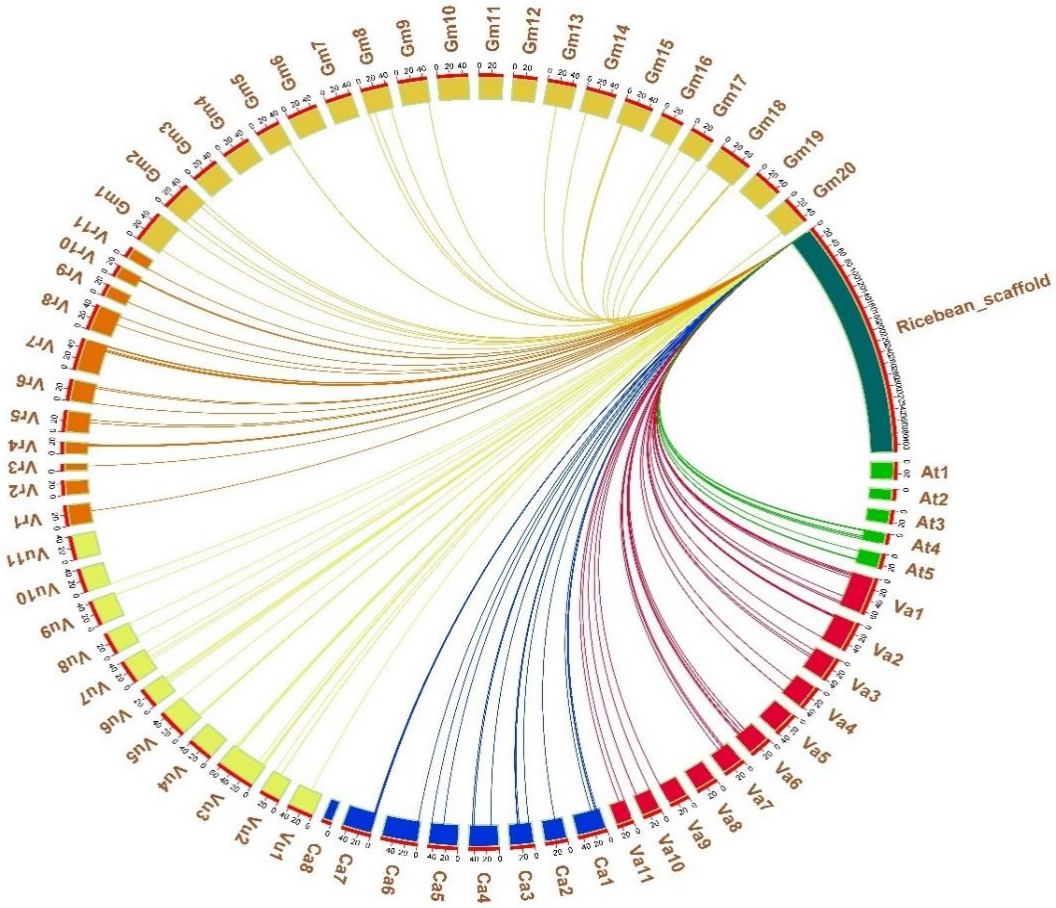


Figure 4.7 Circos plot depicting the syntenic blocks between the chromosomes and/or scaffolds of Vigna umbellata with Arabdiopsis thaliana, Vigna radiata, Vigna angularis, Vigna unguiculata, Glycine max and Cicer arientinum which are represented with different colours . 'At' represents Arabdiopsis thaliana 'Va' represents Vigna angularis, 'Ca' represents Cicer arientnum,'Vu' represents Vigna unguiculata,'Vr'represents Vigna radiata and 'Gm' represents Glycine max.

4.7 Analysis of cis-acting regulatory elements

PlantCARE website was employed to analyze the cis acting elements in ricebean *WRKY*s in order to identify the specific types and distribution of these elements. Figure 4.8 Heatmap of ricebean *WRKY* genes representing cis-acting regulatory element, total 66 cis-acting regulatory elements were observed and are represented through heatmap by tbtools.

- AP-1, ARE, ACE, AE-box, MYB, MYC, Sp1, GATA-motif, GCN4_motif Myb binding site, Myc are some of Transcription Factor Binding Sites (TFBSs)
- A-box, AAGAA-motif, Box4, CAAT-box, CATbox, CCAAT-box, CGTCAmotif, CTAG motif, GCmotif, GT1-motif, I box, LAMP element, MBS, MRE, MYB recognition site, MYB like sequence, RY-element , STRE, TC rich repeats, TCCC-motif, TCT motif, TGA element, TGACG-motif, Wbox,WRE 3, WUN motif, as-1, chsCMA1a, AuxRR core, TATC-box, AT-rich element, AT1-motif, CARE, ATCT-motif are some of Cis-Regulatory Elements (CREs).
- ABRE, ABRE3a, ABRE4, GARE-motif, TCA-element are some hormone/stress eesponse Elements
- CCGTCCmotif, CCGTCC-box,ERE,F-box,G-Box,G-box,LTR,TATA,TATA box,TCA,Circadian are some other regulatory elements and have some function.

The most common *cis*-acting regulatory elements in ricebean *WRKY* genes that are found in almost all genes were AAGAA motif, ABRE, ARE,AT~TATA-box,Box4, CAAT-box, CAT-box, MYB and TATA-box.

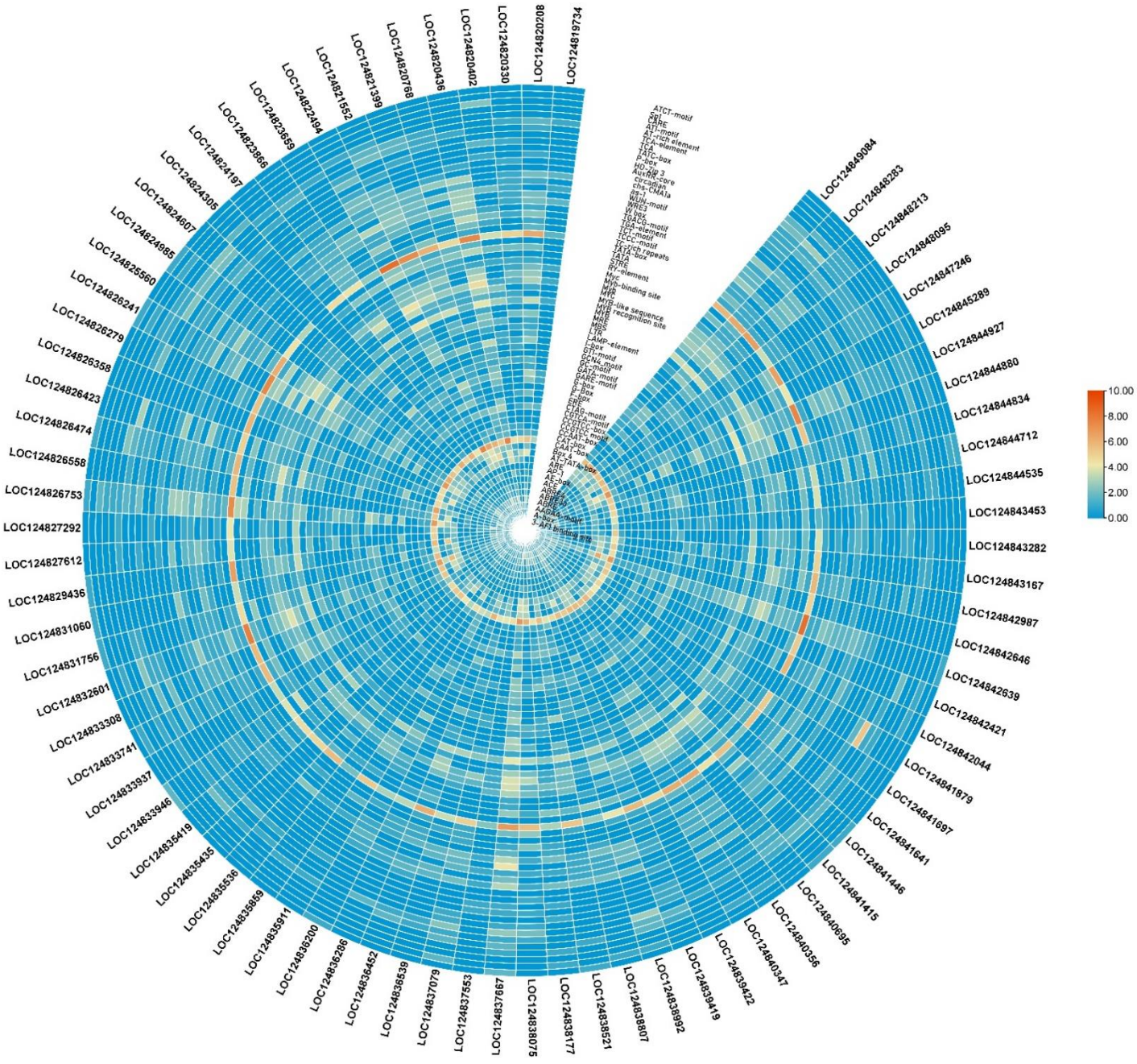


Figure 4.8 Heatmap of ricebean *WRKY* genes representing cis-acting regulatory element. Colour scale represent the intensity of cis acting elements. Red colour indicate higher abundance of cis-regulatory elements and lighter colour like blue and yellow represent lower abundance.

4.8 Expression analysis of ricebean *WRKY* genes

Tbtools was used to create heatmap to show expression profile of genes using RNA-seq data. The given RNA-sequence data shows read counts (gene expression levels) of the 79 *WRKY* genes of ricebean (*Vigna umbellata*) in two ricebean genotypes with different pod sizes (bold and small) at two distinct stages of seed development:
5 days post-anthesis (DPA) and 10 DPA

Transcriptome analysis data exhibit differentially expressed genes (DEGs) were found in bold genotype and small genotype of the two ricebean genotypes at 5-DPA and 10-DPA. The intricate transcriptome dynamics that occur during ricebean seed development are highlighted, as are possible regulators of ricebean seed size as well as other associated features. Figure 4.9 shows the RNA-seq data is represented by heatmap which shows that the expression profiles revealed that all 79 *WRKY* genes of ricebean exhibited discrete expression during development stages. FPKM is a measure used in RNA sequencing experiments to quantify gene expression levels. Higher FPKM values typically indicate higher levels of gene expression, while lower values suggest lower expression levels. Figure8, shows the colour intensity which indicates expression level of two different developmental stages. Red colour has higher intensity of gene expression as compared to green colour which has low intensity and low gene expression level and black colour has intermediate expression level.

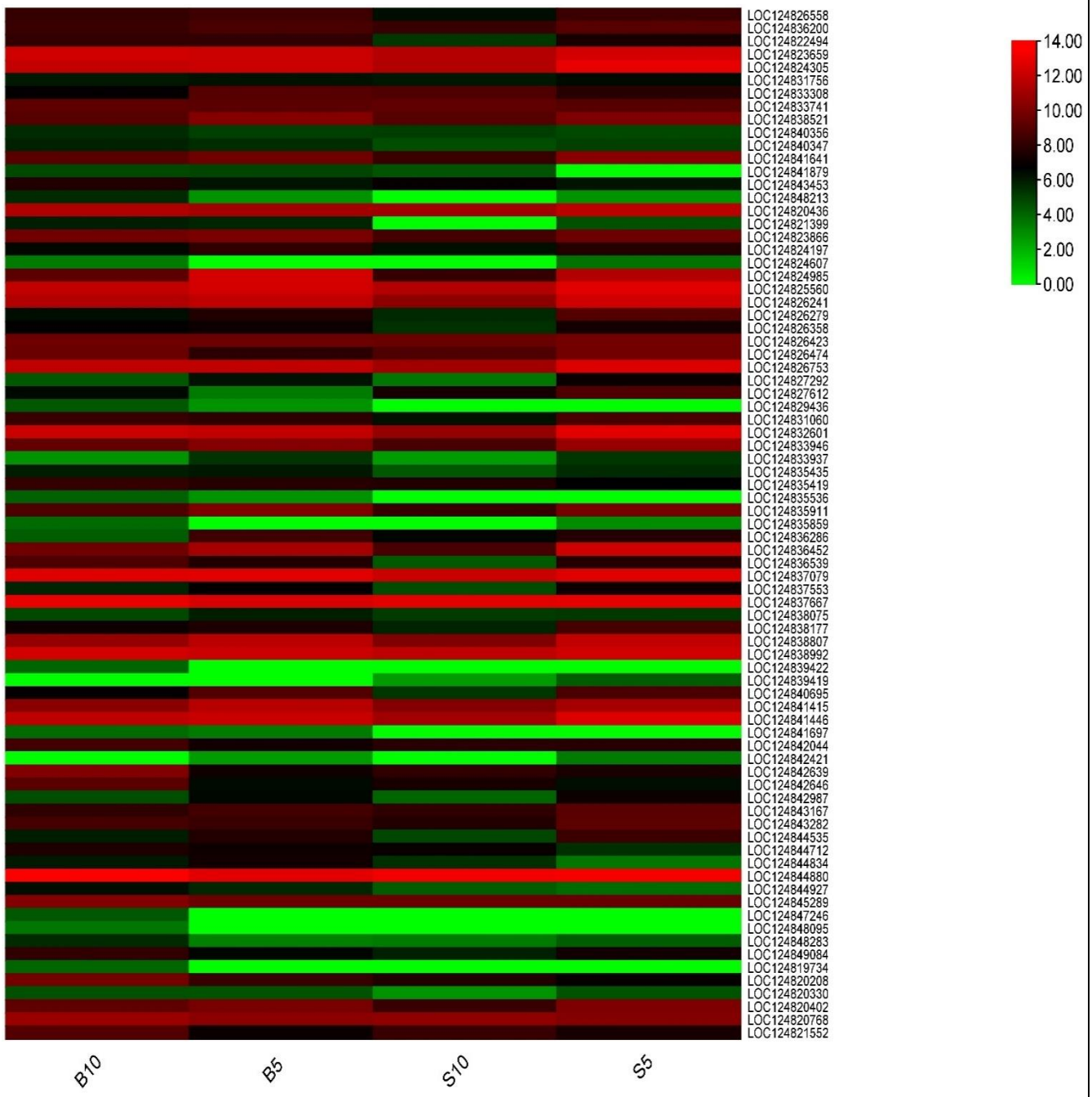


Figure 4.9 Heat map showing the expression pattern of WRKY genes across different developmental stages. In this, red colour has the higher colour intensity and has higher gene expression level. Green colour shows lower intensity which has lower gene expression and black colour has intermediate gene expression level.

CHAPTER 5-

CONCLUSION AND FUTURE PROSPECTIVE

CONCLUSION AND FUTUTRE PROSPECTIVE

In conclusion, given all the above findings and their role, therefore, we presented a detailed genetically analyzed work analysis of the *WRKY* gene family existing in the *Vigna umbellata*. At the beginning, we overviewed the parameters characterize these gene families in the line of their physicochemical properties, multiple sequence alignment, phylogenetic construct analysis, conserved motifs found in the selected groups, gene structure, cis-acting regulatory elements and synteny analysis and gene expression analysis by using the RNA-seq date. Ricebean *WRKY* protein members were distributed in three major structural gene resemblance groups, and all putative conserved Motifs that were found in the same group.

Phylogenetic and synteny analyses that were conducted could be used as high-throughput biomarkers in disease research among other applications once more researches and experiment will confirm the methods. Furthermore, biomarkers from the outcomes of the current study can be used for making predictions concerning genetic triggering of genes in disease or diagnosis. This study was important because the existence of RNAseq data enabled an analysis of the ricebean *WRKY*s from the gene expression perspective on developmental stages. The results of the present study are most valuable, as they will provide a background on where to base future research on the ricebean *WRKY* gene family as a functional and structural annotation.

Future barcodes for genomic-wide identification of the *WRKY* gene family in ricebeans, based on the accumulated information of the importance of the gene family from comparative research in other plant species, will be instrumental in functional characterization, regulatory mechanisms, transgenic studies, application in crop improvement, transcriptomic analysis, comparative genomics, genetic improvement, biotechnological and the use in the study of frontier research in comparable related gene families.

CHAPTER 6 - REFERENCES

REFERENCES

[1] He X, Li JJ, Chen Y, Yang JQ, Chen XY. Genome-wide Analysis of the *WRKY* Gene Family and its Response to Abiotic Stress in Buckwheat (*Fagopyrum Tataricum*). Open Life Sci. 2019 Mar 20;14:80-96. doi: 10.1515/biol-2019-0010. PMID: 33817140; PMCID: PMC7874777.

[2] Todeschini AL, Georges A, Veitia RA. Transcription factors: specific DNA binding and specific gene regulation. Trends Genet. 2014 Jun;30(6):211-9. doi: 10.1016/j.tig.2014.04.002. Epub 2014 Apr 26. PMID: 24774859.

[3] Zhang C, Wang D, Yang C, Kong N, Shi Z, Zhao P, Nan Y, Nie T, Wang R, Ma H, Chen Q. Genome-wide identification of the potato *WRKY* transcription factor family. PLoS One. 2017 Jul 20;12(7):e0181573. doi: 10.1371/journal.pone.0181573. PMID: 28727761; PMCID: PMC5519183.

[4] Banerjee A, Roychoudhury A. *WRKY* proteins: signaling and regulation of expression during abiotic stress responses. ScientificWorldJournal. 2015;2015:807560. doi: 10.1155/2015/807560. Epub 2015 Mar 24. PMID: 25879071; PMCID: PMC4387944.

[5] Garrido-Gala J, Higuera JJ, Rodríguez-Franco A, Muñoz-Blanco J, Amil-Ruiz F, Caballero JL. A Comprehensive Study of the *WRKY* Transcription Factor Family in Strawberry. Plants (Basel). 2022 Jun 15;11(12):1585. doi: 10.3390/plants11121585. PMID: 35736736; PMCID: PMC9229891.

[6] Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The *WRKY* superfamily of plant transcription factors. Trends Plant Sci. 2000 May;5(5):199-206. doi: 10.1016/s1360-1385(00)01600-9. PMID: 10785665.

[7] Phukan UJ, Jeena GS, Shukla RK. *WRKY* Transcription Factors: Molecular Regulation and Stress Responses in Plants. Front Plant Sci. 2016 Jun 3;7:760. doi: 10.3389/fpls.2016.00760. PMID: 27375634; PMCID: PMC4891567.

[8] Li W, Pang S, Lu Z, Jin B. Function and Mechanism of *WRKY* Transcription Factors in Abiotic Stress Responses of Plants. Plants (Basel). 2020 Nov 8;9(11):1515. doi: 10.3390/plants9111515. PMID: 33171689; PMCID: PMC7695288.

[9] Guo, X.; Ullah, A.; Siuta, D.; Kukfisz, B.; Iqbal, S. Role of *WRKY* Transcription Factors in Regulation of Abiotic Stress Responses in Cotton. *Life* **2022**, *12*, 1410. https://doi.org/10.3390/life12091410

[10] Pandey SP, Somssich IE. The role of *WRKY* transcription factors in plant immunity. Plant Physiol. 2009 Aug;150(4):1648-55. doi: 10.1104/pp.109.138990. Epub 2009 May 6. PMID: 19420325; PMCID: PMC2719123.

[11] Liu G, Zhang D, Zhao T, Yang H, Jiang J, Li J, Zhang H, Xu X. Genome-wide analysis of the *WRKY* gene family unveil evolutionary history and expression characteristics in tomato and its wild relatives. Front Genet. 2022;13:962975. doi: 10.3389/fgene.2022.962975. Published September 15, 2022.

[12] Zhang C, Wang D, Yang C, Kong N, Shi Z, Zhao P, Nan Y, Nie T, Wang R, Ma H, Chen Q. Genome-wide identification of the potato *WRKY* transcription factor family. PLoS One. 2017 Jul 20;12(7):e0181573. doi: 10.1371/journal.pone.0181573. PMID: 28727761; PMCID: PMC5519183.

[13] Wu KL, Guo ZJ, Wang HH, Li J. The *WRKY* family of transcription factors in rice and Arabidopsis and their origins. DNA Res. 2005 Feb 28;12(1):9-26. doi: 10.1093/dnares/12.1.9. PMID: 16106749.

[14] Li, MY., Xu, ZS., Tian, C. *et al.* Genomic identification of *WRKY* transcription factors in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants. *Sci Rep* **6**, 23101 (2016). https://doi.org/10.1038/srep23101

[15] Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, Xie B. Genome-wide analysis of *WRKY* gene family in Cucumis sativus. BMC Genomics. 2011 Sep 28;12:471. doi: 10.1186/1471-2164-12-471. PMID: 21955985; PMCID: PMC3191544.

[16] Chen, C., Chen, X., Han, J. *et al.* Genome-wide analysis of the *WRKY* gene family in the cucumber genome and transcriptome-wide identification of *WRKY* transcription factors that respond to biotic and abiotic stresses. *BMC Plant Biol* **20**, 443 (2020). https://doi.org/10.1186/s12870-020-02625-8

[17] Tang R, Zhu Y, Yang S, Wang F, Chen G, Chen J, Zhao K, Liu Z, Peng D. Genome-Wide Identification and Analysis of *WRKY* Gene Family in *Melastoma dodecandrum*. Int J Mol Sci. 2023 Oct 5;24(19):14904. doi: 10.3390/ijms241914904. PMID: 37834352; PMCID: PMC10573167.

[18] Poh W. H., Ruhazat N. S., Yang L. K., Shivhare D., Lim P. K., Kanagasundaram Y., Rice S. A., Mutwil M. (2023) Transcriptomic and metabolomic characterization of antibacterial activity of Melastoma dodecandrum. Front. Plant Sci. 14:1205725. doi: 10.3389/fpls.2023.1205725. Published: September 13, 2023.

[19] Chen J, Hou S, Zhang Q, Meng J, Zhang Y, Du J, Wang C, Liang D, Guo Y. Genome-Wide Identification and Analysis of the *WRKY* Gene Family in *Asparagus officinalis*. Genes (Basel). 2023 Aug 27;14(9):1704. doi: 10.3390/genes14091704. PMID: 37761844; PMCID: PMC10530708.

[20] Srivastava, R., Kumar, S., Kobayashi, Y. *et al.* Comparative genome-wide analysis of *WRKY* transcription factors in two Asian legume crops: Adzuki bean and Mung bean. *Sci Rep* **8**, 16971 (2018). https://doi.org/10.1038/s41598-018-34920-8

[21] Yamaguchi-Shinozaki, K., & Shinozaki, K. (1994). A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell, 6(2), 251-264.

[22] Yu Y, Wang N, Hu R, Xiang F. Genome-wide identification of soybean *WRKY* transcription factors in response to salt stress. Springerplus. 2016 Jun 29;5(1):920. doi: 10.1186/s40064-016-2647-x. PMID: 27386364; PMCID: PMC4927560.

[23] Waqas, M., Azhar, M.T., Rana, I.A. *et al.* Genome-wide identification and expression analyses of *WRKY* transcription factor family members from chickpea (*Cicer arietinum* L.) reveal their role in abiotic stress-responses. *Genes Genom* **41**, 467–481 (2019)

[24] Yan, Y., Yan, Z. & Zhao, G. Genome-wide identification of *WRKY* transcription factor family members in *Miscanthus sinensis* (*Miscanthus sinensis Anderss*). *Sci Rep* 14, 5522 (2024). https://doi.org/10.1038/s41598-024-55849-1

[25] Senjuti Sen, Joydeep Chakraborty, Prithwi Ghosh, Debabrata Basu, Sampa Das, Chickpea *WRKY*70 Regulates the Expression of a Homeodomain-Leucine Zipper (HD-Zip) I Transcription Factor *CaHDZ12*, which Confers Abiotic Stress Tolerance in Transgenic Tobacco and Chickpea, *Plant and Cell Physiology*, Volume 58, Issue 11, November 2017, Pages 1934–1952, https://doi.org/10.1093/pcp/pcx126

[26]Yan, Y., Yan, Z. & Zhao, G. Genome-wide identification of *WRKY* transcription factor family members in *Miscanthus sinensis* (*Miscanthus sinensis Anderss*). *Sci Rep* **14**, 5522 (2024). https://doi.org/10.1038/s41598-024-55849-1

[27] Waqas, M., Azhar, M.T., Rana, I.A. *et al.* Genome-wide identification and expression analyses of *WRKY* transcription factor family members from chickpea (*Cicer arietinum* L.) reveal their role in abiotic stress-responses. *Genes Genom* **41**, 467–481 (2019).https://doi.org/10.1007/s13258-018-00780-9

[28] Kamal Kumar, Vikas Srivastava, Savithri Purayannur, V. Chandra Kaladhar, Purnima Jaiswal Cheruvu, Praveen Kumar Verma, *WRKY* domain-encoding genes of a crop legume chickpea (*Cicer arietinum*): comparative analysis with *Medicago truncatula WRKY* family and characterization of group-III gene(s), *DNA Research*, Volume 23, Issue 3, June 2016, Pages 225–239, https://doi.org/10.1093/dnares/dsw010

[29] Kamal Kumar, Vikas Srivastava, Savithri Purayannur, V. Chandra Kaladhar, Purnima Jaiswal Cheruvu, Praveen Kumar Verma, *WRKY* domain-encoding genes of a crop legume chickpea (*Cicer arietinum*): comparative analysis with *Medicago truncatula WRKY* family and characterization of group-III gene(s), *DNA Research*, Volume 23, Issue 3, June 2016, Pages 225–239, https://doi.org/10.1093/dnares/dsw010

[30] Gürel F, Öztürk ZN, Uçarlı C, Rosellini D. Barley Genes as Tools to Confer Abiotic Stress Tolerance in Crops. Front Plant Sci. 2016 Aug 3;7:1137. doi: 10.3389/fpls.2016.01137. PMID: 27536305; PMCID: PMC4971604.

[31] Mangelsen, E., Kilian, J., Berendzen, K.W. *et al.* Phylogenetic and comparative gene expression analysis of barley (*Hordeum vulgare*) *WRKY* transcription factor family reveals putatively retained functions between monocots and dicots. *BMC Genomics* **9**, 194 (2008). https://doi.org/10.1186/1471-2164-9-194

[32] Zheng, J.; Zhang, Z.; Tong, T.; Fang, Y.; Zhang, X.; Niu, C.; Li, J.; Wu, Y.; Xue, D.; Zhang, X.Genome-Wide Identification of *WRKY* Gene Family and Expression Analysis under Abiotic Stress inBarley. *Agronomy* **2021**, *11*, 521.https://doi.org/10.3390/agronomy11030521

[33] Javadi, S.M., Shobbar, ZS., Ebrahimi, A. *et al.* New insights on key genes involved in drought stress response of barley: gene networks reconstruction, hub, and promoter analysis. *J Genet Eng Biotechnol* **19**, 2 (2021). https://doi.org/10.1186/s43141-020-00104-z

[34] Gürel F, Öztürk ZN, Uçarlı C, Rosellini D. Barley Genes as Tools to Confer Abiotic Stress Tolerance in Crops. Front Plant Sci. 2016 Aug 3;7:1137. doi: 10.3389/fpls.2016.01137. PMID: 27536305; PMCID: PMC4971604.

[35] Ma, J., Lu, J., Xu, J., Duan, B., He, X., Liu, J. (2015). Genome-wide Identification of *WRKY* Genes in the Desert Poplar *Populus euphratica* and Adaptive Evolution of the Genes in Response to Salt Stress. *EvolBioinform Online*, 11(Suppl 1), 47-55. doi: 10.4137/EBO.S22067.

[36] Cadavid IC, Balbinott N, Margis R. Beyond transcription factors: more regulatory layers affecting soybean gene expression under abiotic stress. Genet Mol Biol. 2023 Jan 23;46(1

Suppl 1):e20220166. doi: 10.1590/1678-4685-GMB-2022-0166. PMID: 36706026; PMCID: PMC9881580.

[37] Ma Y, Xu T, Wan D, Ma T, Shi S, Liu J, Hu Q. The salinity tolerant poplar database (STPD): a comprehensive database for studying tree salt-tolerant adaption and poplar genomics. BMC Genomics. 2015 Mar 17;16(1):205. doi: 10.1186/s12864-015-1414-7. PMID: 25881271; PMCID: PMC4372326

[38] Wang C, Chen M, Shao Y, Jiang M, Li Q, Chen L, Wu Y, Cen S, Waterfield NR, Yang J, Yang G. Genome wide analysis revealed conserved domains involved in the effector discrimination of bacterial type VI secretion system. Commun Biol. 2023 Nov 24;6(1):1195. doi: 10.1038/s42003-023-05580-w. PMID: 38001377; PMCID: PMC10673891.

[39] Sun, S., Chen, H., Yang, Z. *et al.* Identification of *WRKY* transcription factor family genes in *Pinus massoniana* Lamb. and their expression patterns and functions in response to drought stress. *BMC Plant Biol* **22**, 424 (2022). https://doi.org/10.1186/s12870-022-03802-7

[40] Kriventseva EV, Gelfand MS. Statistical analysis of the exon-intron structure of higher and lower eukaryote genes. J Biomol Struct Dyn. 1999 Oct;17(2):281-8. doi: 10.1080/07391102.1999.10508361. PMID: 10563578.

[41] Tiika RJ, Wei J, Ma R, Yang H, Cui G, Duan H, Ma Y. Identification and expression analysis of the *WRKY* gene family during different developmental stages in *Lyciumruthenicum* Murr. fruit. PeerJ. 2020 Oct 28;8:e10207. doi: 10.7717/peerj.10207. PMID: 33194409; PMCID: PMC7602686.

[42] Munjal G, Hanmandlu M, Srivastava S. Phylogenetics Algorithms and Applications. Ambient Communications and Computer Systems. 2018 Dec 10;904:187–94. doi: 10.1007/978-981-13-5934-7_17. PMCID: PMC7123334.

[43] Veltri D, Wight MM, Crouch JA. SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. Nucleic Acids Res. 2016 Jul 8;44(W1):W41-5. doi: 10.1093/nar/gkw330. Epub 2016 May 3. PMID: 27141960; PMCID: PMC4987899.

[44] Kaur A, Pati PK, Pati AM, Nagpal AK. In-silico analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabdiopsis thaliana* and Oryza sativa. PLoS One. 2017 Sep 14;12(9):e0184523. doi: 10.1371/journal.pone.0184523. PMID: 28910327; PMCID: PMC5598985.