GENOME WIDE IDENTIFICATION OF *NAC* FAMILY IN RICEBEAN

Dissertation report submitted in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

By

Kritika

EnrollmentNo.225111006

Under the Supervision of

Dr. Shikha Mittal

(Assistant Professor)

DEPARTMENT OF BIOTECHNOLOGY & BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY WAKNAGHAT,

SOLAN-173234, HIMACHAL PRADESH

MAY, 2024

DECLARATION

I hereby declare that work reported in the M.Sc. Biotechnology project entitled**"Genome-wide identification of *NAC* family in Ricebean (*Vigna umbellata*)"** submitted at Jaypee University of Information Technology, Waknaghat, H.P, India is an authentic record of my work carried out over a period from August 2023 to May 2024, under the supervision **of Dr. Shikha Mittal** (Assistant professor).I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my M.Sc. Project report.

Kritika

Enrollment Number: 225111006 Date:
Department of Biotechnology & Bioinformatics
Jaypee University of Information Technology
Waknaghat, India -173234

CERTIFICATE

This is to certify that the work reported in the **M.Sc.** project report **"Genome wide identification of *NAC* family in Ricebean"** submitted by **Kritika** at **Jaypee University of Information Technology, Waknaghat, H.P, India,** in the year from August 2023 to May 2024. Its 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.

Signature of Supervisor : _____

Dr.Shikha Mittal **Date:**
Assistant Professor
Department of Biotechnology &Bioinformatics
Jaypee University of Information Technology
Waknaghat, India -173234

ACKNOWLEDGMENT

My sincere gratitude to my mentor, **Dr. Shikha Mittal**, for all of her help, support, and unwavering encouragement during this project. She has assisted me with more than just my project, she has been a mentor who has cleared my doubts. She has always encouraged me to work hard and taught me how to apply various solutions to problems. She also taught me many other important lessons, like never giving up.

In addition, I would like to thank Dean of Academics & Research and Vice-Chancellor Professor **(Dr.) Rajendra Kumar Sharma** for giving me the chance to join the institute and finish my Master's degree. I also want to say that Prof. **(Dr.) Sudhir Kumar,** the Head of Biotechnology and Bioinformatics, has been a huge inspiration and motivator for my academic career as well as my personal life. I also want to express my gratefulness to the entire faculty of the JUIT BT/BI Department, who has supported me whenever I've needed it. I also want to express my gratitude to all of the lab engineers, especially **Mrs. Somlata Sharma**, who has given me a place to work and constant inspiration.

Likewise, I'd like to express my gratitude to my batchmates Anuja, Akshita and PhD scholar Priyanka for their contributions to the development of this project. They have always encouraged me and helped me through difficult times. They were there to support me whenever I needed it. Many thanks! I want to express my gratitude to the All-Powerful God for his grace throughout my life. Finally, but just as importantly, I want to express my gratitude to my parents for their unwavering motivation, support and motivation throughout my life. They have never given up on me and have always encouraged me.

Thanks to JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

KRITIKA
M.Sc. Biotechnology
JUIT, Solan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

GWI:- Genome wide identification

NAC:-NAM, ATAF1/2, and CUC2

MAFFT: -Multiple Alignment using Fast Fourier Transform

GSDS:- Gene structure display server

TF:- Transcription factors

NCBI:-National Center for Biotechnology Information

HMM:- Hidden Markov Model

BLAST:- Basic local alignment search tool

Pfam:- Protein Family database

PTFD:-Plant transcription factor database

iTOL:- Interactive tree of life

ABSTRACT

The rationale of this study was to perform a Genome wide identification of *NAC* family in Ricebean (*Vigna umbellata*). Previous research has extensively examined the entire *NAC* gene members in Arabidopsis thaliana, *Glycine max,* Rice and other plants. The whole-genome and RNA-seq data of *Vigna umbellata* were used for comprehensive analysis in this study, which included *NAC* gene family-identification, analysis of conserved motifs along with their domains, multiple sequence alignment (MSA), intron-exon distribution analysis, identification of all cis-acting elements, synteny analysis along with expression analysis of *NAC* genes. However, there is some information about the *NAC* family in *Vigna umbellata*. In this study, 115 *NAC*-gene members were identified in *Vigna umbellata* (Ricebean) . Using PFAM database, the *NAC* domains were identified. So, after removal of non *NAC* domains a total of 103 *NAC* genes of Ricebean were obtained. Features of *NAC* Proteins with sequence length ranging from 141 (LOC124830415) to 1117 (LOC124824377) amino acids, with predicted molecular weights that ranged from 36132.3(LOC12484814) to 19678.09 kDa (LOC124823072),the instability index ranged from 21.09 (LOC124824017) to 69.76 (LOC124840572) and aliphatic index ranging from 53.15(LOC124836353) to 108.89 (LOC124820079). Cis-acting regulatory elements were analyzed which included A-box, AAGAA-motif, Box4, CAAT-box, CAT-box, CCAAT-box, CGTCA motif, CTAG-motif etc. In this study, total of 103 cis regulatory elements were analyzed which included A-box, AAGAA-motif, Box4, CAAT-box, CAT-box, CCAAT-box, CGTCA motif, CTAG-motif etc. Further, Cis-acting element analysis that involves searching for cis acting elements that control the gene expression.

The RNA-seq data which was used for expression analysis were represented by heatmap which shows that all 103 *NAC* genes of Ricebean exhibited discrete expression during development stages. Synteny analysis was done across genomes of *Vigna umbellata* with *Arabidopsis thaliana*, *Vigna radiata*, *Vigna angularis*, *Vigna unguiculata*, *Glycine max* and *Cicer arietinum* to identify conserved syntenic regions and orthologous gene-pairs. It helps to understand the evolutionary relationship across different plant species. The comparative gene expression of the identified *NAC* genes was done using RNA seq data that offers how these genes are expressed through different developmental phases. Through heatmap analysis we found that 103 *NAC* genes exhibit expression during developmental stages.

CHAPTER -1

INTRODUCTION

1.1 INTRODUCTION

Ricebean (*Vigna umbellata*) is a highly nutritional legume crop that can significantly improve our diets [1]. It is rich source of minerals and proteins and it is resistant to various pest and diseases. Till now, no analysis has been done on *NAC* gene family in Ricebean. It is a warm seasonal, annual legume that bears yellow colour flowers and has little beans and, fit for human consumption. They are cultivated mainly as a dried legumes, it is also utilized as fodder and as a organic manure. It is known as a small legume since it is predominantly farmed in the grasslands of Indo-China, which runs from Bangladesh to Nepal. The dried seeds contain good quantity of nutrients , since the rice bean based protein is rich in amino acid lysine, so they make an incredible addition of nutrients to the legume -based diet. The seeds contains good amount of minerals and vitamins as thiamine, riboflavin, niacin and vitaminC. The primary tfs are *NAC* (including NAM, ATAF1/2, and CUC2), which correspond to a plant-specific common family and participate in plant growth or stress response activities [2]. It is regarded as a popular food legume in the 21st century, as it is packed with nutrients and manage well with hard growing conditions. This legume has the ability to function as a source of dietary protein to become a part of the crop that could endure abiotic and biotic stresses; therefore can be an ideal component for agricultural systems. It is a member of the Fabaceae family and is appreciated for its high nutritional content and has the adaptability to fit to a variety of environments. Ricebean genome size which is around (414 million base pairs) which in comparison is larger than the genome size of several other plants, including Arabidopsis thaliana (135 million base pairs).

These beans can be grown well in low-input environments and well in marginal soil which makes them perfect for the small farmers to cultivate. It can grow under conditions such as low fertility, water deficit, and/or acidic root zones; and also grow well even on poor soils. This makes it a vital source in sustainable farming and in crop-adapted systems for climate change tolerance. It also contains minerals and vitamins such as vitamin B1 or thiamin, vitamin B2 or riboflavin, niacin, and vitamin C or ascorbic acid. However, the other benefit that Ricebean has is that it is a versatile food with high protein, amino acids ,as well as nutritional profiles. Interest in this plant has grown due to its capability of providing food and nutritional security and reducing malnutrition. The Ricebean has been found to be able to withstand wide collection of abiotic stresses, such as inadequate water conditions, the presence of salts, and extreme heat or cold, when it comes to growth and development.

The present study focuses on the identification of *NAC* transcription factors from Ricebean, which is a valuable study to understand the diversification of this family in an economically important grain legume crop.Similarly, plant transcription factors belonging to the *NAC* family are the most considerable because of their function in response to abiotic stress and growth in plants [3]. Even though this crop is the highest-yielding and the most nutritious however many undesirable traits accompany a significant part of its germplasm accessions and varieties, later being the slow and asynchronous maturation, seed shattering, high rate of pod dehiscence an irregular development pattern, and a comparatively larger concentration of unfavorable nutritional components.

The investigation of the different families of tfs like *NAC* in plants has become an significant focus of research at genomic level because of their significant roles in controlling various biochemical processes in the plants. It should be noted that the attention to the *NAC* family of transcription factors has grown rapidly in terms of stress, growth, and development. The N-terminal conserved NAM domain and the unique transcriptional regulatory domain at the C-terminus are characteristic features of these proteins. *NAC* plays a major impact on several plant processes. *NAC*s help plants adapt to harsh environments and survive by regulating genes that react to stress. The *NAC* transcription factors have emerged as key elements in the regulation of plant response to both biotic and abiotic stress agents [4]. The comprehensive analysis of *NAC* family genes in plants opens a new prospect for analyzing the molecular functions of stress tolerance, growth, and development of plants. Such improved understanding not only advances existing knowledge of plant biology but also opens up the potential for the production of more sustainable crop types with high stress tolerance and yield capabilities. This study utilizes genomic tools like gene sequence assembly and modern approaches like annotation in order to screen and identify *NAC* transcription factors that have not been reported in Ricebean. *NAC's* functional role in this crop should be done whole time through complete genome-wide annotation of *Vigna umbellata* (Ricebean) via the available methodologies such as determination of gene structure, RNA seq data analysis, and expression profiling that may probably improve our understanding of the mechanisms resistance to plant for different conditions. In this research study, Ricebean was analyzed for the *NAC* gene family by using various sequencing and bioinformatics techniques. The first part of the comprehensive approach in this study was to extract the sequence data for all the *NAC* proteins of interest, after which the physiochemical properties of these proteins were examined.

Furthermore, the functional domains of *NAC* genes were determined by analyzing the domain

using the Pfam database, and gene structure of *NAC* genes were analyzed by using the gene structure display server (GSDS). Syntenic analysis was performed to align conserved genomic region between species and tbtools was used to reveal the evolutionary relationship across the *NAC* gene family. Furthermore, the knowledge about cis-regulating elements and the study of *NAC* gene expression patterns using RNA seq data provide insight to the regulatory elements that manage the expression of the gene and expression profile during developmental stages. The genome-wide identification as well as classification of *NAC* gene family in Ricebean provided valuable insights into their diversity, evolution, and functional specificity, making *NAC* transcription factors a subject of extensive research with significant implications for plant biology and crop improvement [5].

Origin and expansion

Ricebean belongs to the family Vigna originating from the genus Vigna and is known to be a part of the *NAC* gene, which is present in many cereals, including vegetables, legumes and fruits, and model plants. While using genetics and text, it supports the fact that it has had a historical distribution around the world and can be considered an important crop.

Nutritional Value of Ricebean

Ricebean has a higher percentage of protein (25.57%), dietary fiber, vitamins, and minerals. It is important in human diets. since they comprise of essential amino acids [6]. Particularly, Ricebeans can serve as an antioxidant, which helps them fulfill the body's dietary requirements. The Ricebean has been found to be able to withstand a broad range of abiotic stresses, such as inadequate water conditions, the presence of salts, and extreme heat or cold, when it comes to growth and development.

Significance

There is little information on the function of *NAC* gene family in Ricebean; therefore studying the family will give significant information on how the plant handles stress. Understanding Stress Responses: The various function of *NAC* genes that will help in understanding the growth and stress tolerance mechanisms in Ricebean.

In Ricebean, the biological function of the *NAC* transcription factor gene family is yet to be studied few efforts made towards researching the family will greatly help in understanding how the plant respond to stress in general.

Comprehending Stress Responses: Understanding the various identities assumed by *NAC*

genes can help in exploring the genetic control and developmental processes as well as stress resistance trait in Ricebeans.

Crop Improvement: A simple knowledge about numerous *NAC* genes can also help in understanding the impact of such genes on a particular stage of development.

Problem statement

The *NAC* gene family perform significant role in various plant species, influencing disease resistance, stress tolerance and other important functions. However, a comprehensive understanding of this gene family in Ricebean is missing. This knowledge gap inhibits the development of efficient breeding programs that could significantly improve crop resilience and adaptability for Ricebean growers facing environmental challenges and developing agricultural demands.

Objective

- To identify *NAC* family genes in *Vigna umbellata* using orthologs of vigna species.
- Expression analysis of identified *NAC* family genes using RNA-Seq data.

CHAPTER -2

REVIEWOF LITERATURE

2.1 Review of literature

Genome-wide identification studies refer to research involving large-scale investigations of the complete set of genes in the hope of searching and studying certain gene families or regulatory regions[7].These studies employ large-scale sequencing techniques, computational methods, to comprehensively investigate the content and function of DNA in an organism, in an attempt to solve the roles and relationships of the genes within a genome and are a widely used and is an important step to resolve the genetic background of various biological functions such as stress, organ development, and adaptation. In the field of plant biology, these studies have played a significant role to comprehend the gene families and regulatory systems with respect to desired plant characteristics or events in response to various stimuli.

The present study focuses on the identification of *NAC* transcription factors from Ricebean, which is a valuable study to understand the diversification of this family in an economically important grain legume crop.

Overview of *NAC* transcription factors

The *NAC* gene family is the biggest family of tfs and their role is quite important in several aspects of life including growth, development and even aging. The *NAC* family is named after three reported proteins: These DNA binding factors include NAM (No Apical Meristem), ATAF1/2 (Arabidopsis thaliana Transcription Activator Factor 1/2) and other proteins like the CUC2 (Cup-shaped Cotyledon 2) which belongs to this group and plays a role in both abiotic as well as biotic stress regulation. The *NAC* genes possess a feature two regions: N terminal which is recognized as DNA-binding domain located at the terminal end of the gene and a C terminal region known as transcription activation domain (AD) located on the other terminal end of the gene. The N-terminal regions that contain about 160 amino acid which are further classified into five subdomains which are put in the order as follows A to E. The *NAC* TFs functions in nutrient transfer, apical meristem formation, stress reactions including senescence process and the cell cycle. *NAC* TFs family, which is widely prominent for its essential role in controlling growth of plant and plant improvement . The most known and extensively studied gene families are the *NAC* tfs family. It was found that *NAC* plays an vital function in the growth and enlargement observed in Petunia embryos. Transcription factors associated with genes from the *NAC* family play a considerable role in maturing processes of heightened level of vegetation, including growth and reactions to stresses. *NAC* proteins, which are made up of sub-domains, for example A, B, and E, can form dimer through interactions .The *NAC* domain composition operates as a dimer, allowing the

formation of heterodimers and dimers with other species. The C-terminal segment of *NAC* proteins represents a transcriptional regulatory site where protein interferes with gene expression either as a repressor or an activator. Some *NAC* transcription factors (TFs) possess DNA binding motifs have transmembrane motifs which are attached to endoplasmic reticulum or the cell membrane that facilitate the response mechanisms and cell signaling. Transcriptome and genome expression studies showed that *NAC* family in Ricebean is implicated in stress tolerance pathways. This makes them great option for developing stress-resistant transgenic plants. *NAC* genes play function in stress tolerance, growth, and stress response that can provide information about how the crop is resistant to harsh conditions.

Function in Plant development and growth and Stress Responses

Plant *NAC* transcription factors are involved in a variety of biological processes that includes seed and embryo formation, shoot tip meristem generation, leaf aging, fiber formation, cell division, and lateral root development. Furthermore, numerous *NAC* gene are involved in response to abiotic stress including salinity, drought and cold stresses.

Applications in Crop Improvement

The study of *NAC* TFs has significant implications for crop improvement. Over expressing *NAC* genes has been shown to develop tolerance to salt, drought and cold stresses in the transgenic plants. Furthermore, CRISPR/Cas9 technology are used to engineer crops by knocking out negative *NAC* TFs that enhance stress tolerance or knocking in positive *NAC* TFs increases the crop resistance against various pathogens and it increases the quality and quantity of total yield.

Research highlights in different plants

Analyses of *NAC* Gene Family in Soybean

In this study, a total number of 139 *GmNACs* non-redundant *NAC* genes was analyzed in the genome of soybean plant [8]. Using MEME tool, a total of 20 conserved motifs were recognized. All Gm*NAC*s contain no less than five of the seven main motifs. Each motif was denoted using a colored box. The non-conserved sequences denoted by using black lines.

The phylogenetic tree was conducted using 1000 bootstraps by employing 139 *NAC*s from soybean, 32 *NAC*s from rice, 78 *NAC*s from Arabidopsis, which were divided into 17 separate categories [8]. In group four, members of Gm*NAC*s were absent.

The phylogenetic tree showed that Gm*NAC*100, Gm*NAC*102, Gm*NAC*103, Gm*NAC*118, Gm*NAC*108, and Gm*NAC*108 serve as crucial role in increase in growth and progression of the shoot apical meristem, additionally *OsNAC52*, *ANAC017*, *ONAC063*, *GmNAC004*, *GmNAC005*, *GmNAC116* and *GmNAC117*, which were liable for tolerating drought and oxidative stress in genetically modified plants.

Also Exon-intron analysis of Gm*NAC*s showed the numbers of introns was ranged from zero to six. With the help of data analyzed, we found 94 out of 139 Gm*NAC* genes included 2 introns. Gene structure examination representing the exon/intron distribution in Gm*NAC* genes with yellow colour representing exons (coding regions), black colour represents(non-coding regions) and blue colour representing UTR(untranslated regions) [9].

According to the findings, Gm*NAC* genes are engaged in various biological processes such as oxidative stress, drought tolerance and expansion.

Unveiling the Functional Diversity of Gm*NAC* Genes in Soybean's Drought Stress Response: *NAC* tfs are recognized for their involvement in diverse biological and physical processes such as stress responses, evolution processes and hormone signaling. In soybean, the *NAC* family is linked with the drought tolerance, a decisive trait given the escalating occurrence of drought stress caused by the climate change. A comprehensive examination of 28 Gm*NAC* genes from both drought-sensitive soybean and tolerance to dryness varieties opened a possible relation in middle of expression of particular Gm*NAC* genes and drought tolerance [10]. The functional analysis conducted on these genes under drought conditions indicated that certain Gm*NAC*s could have significant roles in enhancing drought resistance. For example, rice roots expressing Os*NAC*10 significantly enhance drought resistance and crop yield. This suggests that the possibility of Gm*NAC* genes in soybeans may be comparable. The study's evolutionary insights and structural analysis revealed that the *NAC* proteins in soybeans contained 20 conserved motifs. The distribution of these motifs among the Gm*NAC* family members suggests that the family has a diverse structure and function, with some motifs being connected to different biological processes like leaf formation and reactions to abiotic stress. To improve Gm*NAC* resilence against drought deficiency, stress response and scarcity, the researchers are concentrating on Gm*NAC* genes. This approach involves identification, characterization and analyzing the expression of Gm*NAC* genes in soybeans [11]. Differential expression of 139 Gm*NAC* genes in V6 and R2 leaves during drought stress. The genes indicate that they are either up-regulated or down-regulated by a minimum of two fold; the colors scale represents the Red color shows high intensity, green shows lower

intensity, and grey colour shows no expression. R2 and V6 were investigated at two developmental phases for the purpose to look at the gene expression of Gm*NAC* genes during drought stress. Multiple Gm*NAC* genes were discovered during the investigation [12].

The study demonstrated that large number of Gm*NAC* genes, including Gm*NAC*005, Gm*NAC*041, Gm*NAC*040, and others, showed notable upregulation in response to the drought response at both stages. This implies that the ability to withstand drought may be largely dependent on these genes. Additionally, it was shown that a subset of these genes had much higher expression levels in the roots of drought-resistant soybean cultivars, suggesting that these genes participate in the plant's ability to become adapted to drought. The functional characterisation of *NAC* transcription factors demonstrates their significance in several physiological and biotic procedures. The study highlights the potential of using these findings to develop transgenic soybeans with increased drought tolerance.

Investigation of 28 Gm*NAC* genes that are sensitive to drought, drought sensitive soybean varieties revealed correlation between the gene expression of Gm*NAC* genes and the level of water conservation [13].

Examining Mungbean Cultivation's Potential and Genetic Understanding

Overview of Mungbean

Mungbean, scientifically known as *Vigna radiata*, is a legume crop that is annually cultivated during the warm season. It belongs to the papilionoideae subfamily of the Fabaceae family. Mungbean is the largest producer worldwide in India followed by Myanmar and China [14]. This crop offers important economic and health benefits. Mungbean seeds are highly nutritious, containing amounts of iron, folate and high-quality proteins. Intercropping Mungbean with cereals has also been found to reduce pest attacks and to increase the crop yield and Plant breeders have utilized genes of interest from wild species to develop Mungbean crops. Mungbean is a pulse crop that has diverse functions but the crop were not harvested in the study area. In addition, no information is provided about the improved agro-technologies and how they can be used in that particular area. Mungbean comes from the kind of moist beans and it is a newly accepted pulse in lowlands. Agronomic technician adapts the crop to each season in low land areas and high temperature areas [15]. This crop has a source of fat, protein, carbohydrate and fiber. The agronomic technologies and potential adaptation being presented might serve as techniques to try out crops that could be used in the area. As a result, an experiment was launched to investigate the relationships between

spacing of rows, and fertilizer dosage on Mungbean varieties at Kindo Koysha district, in the year 2018 [16]. Mungbean comes from the category of moist beans and it is nothing but a newly accepted pulse in a lot of lowlands. Agronomic technician adapts the crop to each season and achieve success in high-temperature areas and low land areas. This crop has a wealth and source of carbohydrate, protein, fiber and fat. So far, the cash crop is not been extensively studied in that area. In addition, analysis of 20 conserved motifs within Vr*NAC*s revealed their functional roles in living organisms. Researchers analyzed *NAC* genes in Mung bean (*Vigna radiata*) and discovered a high degree of similarity to those in Arabidopsis. The similarity allowed them to divide the mung bean *NAC* genes into 13 groups based on shared characteristics and evolutionary relationship (paralogs and orthologs). Further investigation of the 173 putative Vr*NAC* genes showed their transcriptional network components, identifying Vra-miR165 as a key regulator of leaf development and water resistance in plants [17]. Furthermore, study of the *NAC* gene promoter sequences showed several cis-acting elements, providing insights into their diverse roles in variety of activities such as stress response, light, and hormone regulation. A precise number of 985 cis-regulatory elements were detected, which correspond to diverse mechanisms such as biotic and abiotic stresses. it gather up the different elements, including the light-responsive element, which were discovered in the promoter sequences of Vr*NAC*. Light-acting cis-regulatory element was detected in Vr*NAC*s promoter region .Through these findings, we can observe that *Vigna radiata NAC* TFs may unite various biotic and abiotic stress.

22 indiscriminately selected Vr*NAC* genes were further analyzed by qRT-PCR to determine whether they were expressed in the NM-98 genotype (a variety that is highly responsive to bacterial leaf spot disease and drought) or not.

Gene ontology (GO) analysis revealed 173 genes that were involved in transcriptional network of potential Vr*NAC* genes.

To understand the genetic associations among species Vr*NAC*s were divided in nine groups revealed by a phylogenetic analysis. Also, examination of twenty conserved motifs among the Vr*NAC*s revealed their function in various processes. Mungbean *NAC* genes were categorized into 13 groups since each gene feature paralogs and orthologs. Such characters as branching pattern, branch length were also noticed [18].

The phylogenetic analysis helps in understanding the evolutionary linkage among the species. Several works have been reported on the identification and functional investigation of *NAC* proteins in plant species, which help clarify their functions in stress response networks and developmental processes.

In identifying *NAC* proteins in *Arabidopsis thaliana,* researchers identified about105 *NAC* proteins. To determine *NAC* proteins in Arabidopsis thaliana the research investigated , which focused on their participation in stress response signaling [19]. The study designed to classify the function of *NAC* tfs in the development of plant and in stress responses by employing bioinformatics and protein expression. The methodology used was followed by data collection where *NAC* protein sequences were reported from the Arabidopsis Information Resource (TAIR) and UniProt databases by performing a systematic search. Followed by bioinformatics analysis to classify and characterize the identified *NAC* proteins, multiple sequence alignment, and phylogenetic analysis were conducted. The expression profiling was done for the purpose, RNA-seq data under various stress conditions, and to analyze the expression of *NAC* genes, various developmental stages were utilized. Overall, 105 *NAC* proteins were detected in the *Arabidopsis thaliana* and genome were separated into particular subfamilies in accordance to the phylogenetic tree analysis. Consequently, expression patterns of *NAC* genes indicated their involvement in stress responses, where their expression was found to be differentially regulated under various abiotic and biotic stresses. Other than that gene ontology analysis was performed, where the *NAC* proteins are implicated in stress signaling, hormone regulation, and developmental pathways [20].

Another study of *NAC* gene family in cowpea

Cowpea (Vigna unguiculata) is likely to be a food legume plant in early 21st century with numerous nutritional and economic benefits. Relative to other grain legumes, cowpea is moderately tolerant to abiotic stresses and can be produced in low P availability and non-rains Cassava grain production can occur in areas with low rainfall and high temperatures, and cowpea grains continue to be produced even at 300 mm of rainfall . The cowpea genetic variation and phenotypic flexibility has placed the crop into biotic and abiotic stress environments that is bad for soil and climate. The cowpea genome was recently assembled which provides a platform for high throughput genomic and functional analysis of the crop. Perhaps, the exploration till now is unexplored. Studying *NAC* family in this drought tolerant legume crop may help to understand strategies for improving multiple stress tolerance, growth and yield can be exploited and identify new avenues for developing this crop plant.In this study, we aimed to understand the annotation of *NAC* Tfs (Vu*NAC*s) in cowpea and predict their role and significance. This objective of this study was to analyze: (i) Vu*NAC* Transcription factor family, (ii) conserved motifs and others unique to protein and gene

characteristics and (iii) Vu*NAC* TFs regulatory pathways of stress signaling, developmental and metabolic processes.

Results of this analysis showed that cowpea plant has 130 *NAC* proteins, or Vu*NAC*01-130 placed into 8 phylogenetic clusters [21]. None of the 27 putative cowpea-specific members closer to any other species hence implying new functional roles for this single clade. Multipartite nuclear signals and distinct transactivation domains similar in sequences were presented in Vu*NAC* proteins. These included the following proteins: It was determined that 18 proteins contained non-*NAC* domains. The genes had a different regulatory region organization of a well-defined promoter. Furthermore, there was a high degree of segmental and chromosomal duplication within the family and consequently, large paralogous groups and a high number of stress—responsive genes [22]. This supports the suggestion derived from hows the genes are regulated for the existence and interact between stress and growth regulating signals For Multi-tier regulation which is processed through light hormone and transcription factor (*NAC*/MYB/WRKY/ERF and Dof/TCP).

A comprehensive analysis of *NAC* gene family members in chickpea (*Cicer arietinum*): identification and expression profiling

Chickpea (*Cicer arietinum)* is the most significant legume crops cultivated around the world, particularly in Afro-Asian regions. Which utilized abundantly around the globe because it is affordable and its quality of protein, carbohydrate, mineral, vitamin and lipids is far better for direct human consumption than other legumes? Chickpeas by-products such as chickpea hay and straw, chickpea pod husks and so on, they are also appropriate for the animal fodders [23]. But, drought is one of the biggest threats for the production of chickpeas, so it is necessary to develop a variety of drought-resistant chickpea and it is the most crucial agenda in the current chickpea breeding programs.

To analyze Ca*NAC*genes in chickpea the subsequent methods were employed; all the predicted Ca*NAC* genes was retrieved from the sequence data of PlantTFDB and iTAK [24]. Seventy one potential Ca*NAC* genes were predicted for chickpea. To determine where each of the Ca*NAC* genes were located on the chromosomes, a blast was done against the assembled "kabuli genome".Therefore, we sought to know the Ca*NAC* genes that are annotated in the desi ICC4958 chickpea variety. In the identified genome of "desi" ICC4958, there were about 62 Ca*NAC*s; however, the "kabuli" type CDC Frontier genome contained 9 more Ca*NAC*s. This could be because desi chickpea genome contain less number of protein encoding genes

about 27, 571 genes while the kabuli chickpea genome contain 28, 269 genes [25]. For ease of reference, we suggest a nomenclature system for the identified Ca*NAC* members after the present study. This system provided the name Ca*NAC*01 up to Ca*NAC*71 according to their chromosome position in the "kabuli" origin chickpea genome. out, 65 of the Ca*NAC* genes could be mapped on the 8 different chromosomes of chickpea based on the available chickpea genome databases.. Such 65 Ca*NAC*s were located on the eight chromosomes and are distributed at an unequal rate. It is also evident that more number of Ca*NAC*s are encoded in the chr 6 with 13 members contributing ~ 20% of the Ca*NAC*s identified on the expenditure of members encoded in chr 7, which only contributed 4 of the 65 members of the mapped Ca*NAC* genes accounting only for ~ 6% . Out of the 65 identified Ca*NAC*s, applying the criterion "greater than 60% homology at nucleotide level", it was found that there were 8 the genes which were present in pairs, none of which were tandem duplications. The following findings are noteworthy when comparing chickpea with soybean: Thirteen sets of 2 or more Gm*NAC*s were found duplicated in soybean out of 152 analyzed Gm*NAC*s. On the basis of phylogenetic relationships Ca*NAC*s are easily categorized into 12 groups along with A*NAC* homologs [26]. Expression patterns of Ca*NAC* genes across different tissues revealed total of 44 Ca*NAC* genes in these tissues, and the relative expression data of these genes could be found in the CTDB using this keyword. With respect to the data, it can be conclude that the Ca*NAC*s have highly diverse tendencies connected with transcript expression. For instance, the expression of Ca*NAC*01, Ca*NAC*49, or Ca*NAC*63 production was very low in each of the tissues among the members of Ca*NAC*s. Among the stress-related Ca*NAC*s (S*NAC*s) which are putatively predicted from the gene characteristics or the membrane-linked CaNTLs, genes showing a high level of transcript abundance in the tissues. It was also observed that certain Ca*NAC* genes showing differential expression pattern which included Ca*NAC*16, Ca*NAC*20 and Ca*NAC*50 were identified to be specific to certain tissues alone although a lot of them showed ubiquitous expression in the tissues analyzed at various developmental stages. Through RT-qPCR, expression levels of 23 Ca*NAC* genes were mapped in this study. Based on phylogenetic tree analysis ,15 stress-related Ca*NAC*s were predicted and all the 8 CaNTLs containing a membrane-anchoring domain in dehydration-stressed chickpea leaves and roots [27]. Out of twenty-three genes of Ca*NAC*s, fourteen were up-regulated, while four genes which was namely CaNTL1/Ca*NAC*04 were down-regulated by minimum of two fold in the leaves under dehydration stress. The expression levels of these genes exceeding P>200 and P>300, and Ca*NAC*06 and Ca*NAC*67 were checked to be the most strongly induced genes respectively. In contrast, Ca*NAC*02 and Ca*NAC*04 were the most significantly down-

regulated genes; they show (23. 8 fold and 28. 6 fold) down-regulation after 5 hours of depletion in chickpea leaves. In the roots 12 genes, out of which CaNTLs (CaNTL2/Ca*NAC*19 and CaNTL6/Ca*NAC*44) showed up-regulation, and 3 genes, which includes CaNTLs (CaNTL1/Ca*NAC*04) were down-regulated at least 2 folds following dehydration for 2 or 5hours [28] .Collectively the data showed that out of 23 *CaNAC*s examined, 19 genes were dehydration responsive in either leaves or roots that were representing 62.5% of the CaNTLs. Dehydration-responsive Ca*NAC*s overlap in the roots and leaves, with 10 and 3 genes up-regulated and down-regulated in each organ, according to Venn diagram analysis. In case of dehydrated leaves 3 genes (Ca*NAC*05, 21, and 57) were only upregulated, whereas incase of dehydrated roots two genes (Ca*NAC*24 and 44) were only up-regulated under the experimental circumstances [29].

In terms for down-regulation, the *CaNAC24* were being identified to be down-regulated in the leaf tissues by dehydration. It was observed that out of 15 phylogenetically predicted stress-related *CaNAC*s, 14 genes were dehydration-responsive. On basis of phylogenetic based analysis method,gave 93.33% accuracy rate which is a good rate for a prediction. Our recent study, which described the full discovery and characterisation of the Ca*NAC* family in chickpea, shed light on the Ca*NAC* family's functional variety. Additionally, expression analyses of Ca*NAC* genes' during growth, depletion, and ABA treatments have provided scientists studying chickpeas with significant information that made them select candidate genes and the associated tissue-specific, dehydration-, and/or ABA-responsive promoters for further study in plant functional studies, that leads to the production of modified chickpea cultivars that are more drought-tolerant [30].

Another study based on structural and Functional examination of *NAC* Proteins in Maize Plant. In a study on *NAC* proteins in maize plants, molecular biology techniques and physiological assays were used to identify and analyze their function and potential in controlling drought stress responses. The study aimed to understand the function of *NAC* tfs in regulating maize plants during water-deficit conditions [31]. Maize plants were transformed with *NAC* genes, which could potentially improve drought tolerance via Agrobacterium-mediated transformation and overexpression. Drought stress tests were done to assess the impact of *NAC* in enhancing drought tolerance, monitoring measures such as relative water content, stomatal conductance, and chlorophyll fluorescence [32].

Therefore, real-time PCR method was used for Gene Expression Analysis to know how stress-responsive stress genes are expressed in *NAC* overexpressing maize plants under

drought condition for getting a Improved Drought Tolerance plant. Having seen the effects of overexpression of selected *NAC* genes in the maize plants it was observed that the plants overexpressing the particular genes had better drought tolerance as compared to the normal plants in terms of relative water content and rate of transpiration water loss. This raised the expression of osmotic stress related genes and the antioxidant systems in the plants by overexpressing *NAC* genes transgenically. The transgenic plants that overexpress the *NAC* gene exhibited higher quantum yield of PSII, stomatal conductance to water vapor, and less wilting under drought stress, which implied that the water status in the plant was optimal [33].

CHAPTER – 3

MATERIAL & METHODS

3.1 Materials and methods

Sequence Data retrieval

The protein sequences of plants like ,*Cicer arietinum, Glycine max, Vigna unguiculata, Vigna angularis, Arabidopsis thaliana* and *Vigna radiata* were searched for and retrieved from the Plant transcription factor database. Ricebean protein and gene sequences were retrieved from the NCBI website. The BLASTX search was conducted using protein sequences of *Vigna radiata, Cicer arietinum, Glycine max, Vigna unguiculata, Vigna angularis* and *Arabidopsis thaliana* by setting the parameters like expected value to 1e -10 that represents the number of hits.

Domain analysis

Domain analysis refers to the process of locating conserved protein domains within gene sequences. According to this analysis was conducted using (Pfam) protein family database of identified 113 Ricebean *NAC* genes, it was analyzed and identified that the non *NAC* domains in *NAC* genes were 10 in total and after the removal of non *NAC* domains a total of 103 *NAC* genes of Ricebean were obtained.

To analyze Physicochemical properties of proteins

Analysis of different parameters like the molecular weight (MW), isoelectric point (pI), theoretical pi and GRAVY by inputing the protein sequences of *NAC* genes and also the chromosome location of *NAC* gene family was done using Protparam tool https://web.expasy.org/protparam/. Protparam helps to analyze the physical and chemical properties of a protein. Analysis of all *NAC* genes ,physiochemical properties is essential in understanding the function, stability and behavior of proteins in the biological processes.

Comparative phylogenetic analysis

Through Phylogenetic tree construction, Scientists can visualize and interpret the evolutionary relationship among the species. It gives a clear understanding of how different species are related or not related to each other. For multiple sequence alignment, (MAFFT) Multiple alignment using fast fourier transform tool was used for the obtained protein sequences from *Vigna radiata, Vigna unguiculata, Cicer arietinum, Glycine max, Vigna angularis* and *Arabidopsis thaliana* that were taken as query. All protein sequences of these genes that includes *Vigna radiata, Vigna*

unguiculata, *Cicer arietinum*, *Glycine max*, *Vigna angularis* and *Arabidopsis thaliana* genes were aligned using (MSA) Multiple sequence alignment. Further to visualize the phylogenetic tree, iTOL (https://itol.embl.de/) (Interactive tree of life) software was used.

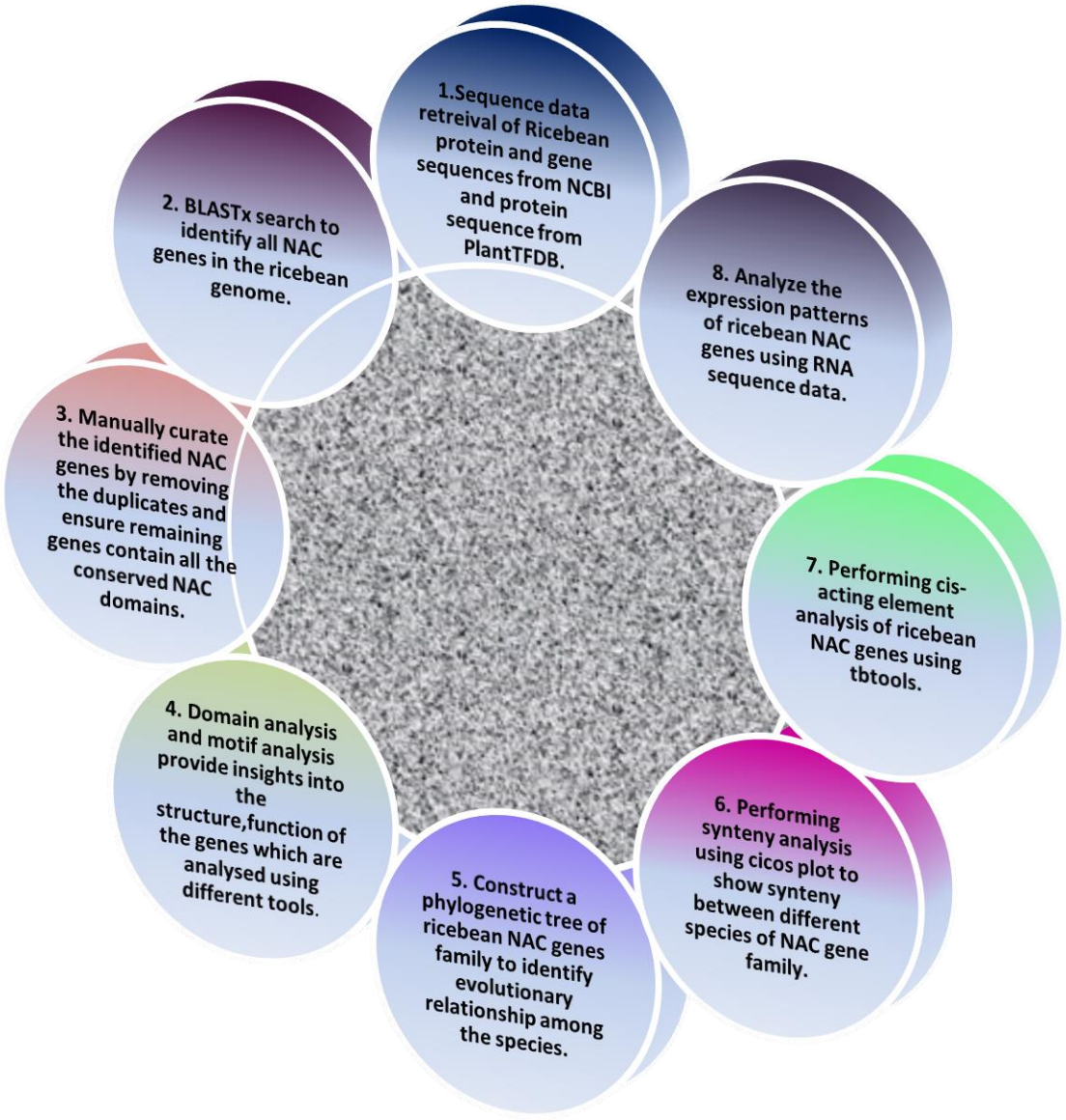


Figure 3.1 Flowchart for the Genome wide identification of *NAC* family in Ricebean.

Conserved Motif analysis

The conserved protein motifs found in the *NAC* genes were identified and characterized using the Multiple Em for motif Elicitation (MEME) version (http://meme-suite.org/tools/meme) by setting the parameters with the number of searches that were set to 5 motifs. The analysis demonstrates the possible function of *NAC* genes in plant growth, development and stress response as well as their evolutionary relationships and regulatory mechanisms.

Gene Structure analysis

To identify the gene structure of all *NAC*s , Gene Structure Display Server 2. 0 (http:// gsds. cbi. pku. edu. cn/) was used. Analysis of structure of exons and introns has had studies on how genes are characterized. Based on the information, the Gtf file of Ricebean was taken and their graphical representation was made by TBtools (https:(<link/ http://github. com/CJ-Chen/TBtools/releases).Researchers can discover patterns in exon and intron length and quantity and its sequence for explanation of gene expression, protein diversity. The exon-intron structures of Ricebean *NAC* genes were examined using GSDS2.0 server. These exons and introns give a clear exposure of how *NAC* gene families are undergoing progression.

Synteny evolutionary analysis

Syntenic analysis was carried out by utilizing TBtools (http://circos.ca/) that was used for analyzing genomic data. Syntenic evolutionary analysis between different plant species were performed using the genome file of each *NAC* species as inputs in the fasta stat. The synteny analysis was conducted with tbtools using multiple chromosome layouts file, gene links files and GFF files between *Vigna umbellata* and other query species *Vigna radiata*, *Vigna unguiculata*, *Cicer arietinum*, *Glycine max*, *Vigna angularis* and *Arabidopsis thaliana were obtained* by performing MCScanx with default parameters. Synteny analysis used to determine the evolutionary links among the species. The collinear relationship between other query species and genome of *Vigna umbellata* were visualized using these files through advanced circos tbtools.

Putative Cis-acting Element Analysis of *NAC* genes

Cis regulatory elements plays necessary function in controlling gene expression. They differ

from diffusible substances that are known as trans-acting regulatory elements, which have the ability to control genes that are found in other regions of the genome. The complete length sequence of the transcriptional start site of *NAC* genes were extracted and submitted to PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ to retrieve and observe the cis -acting elements of the *NAC* genes. Tbtools was used to create a heatmap from the data that was obtained from the PlantCARE website.

Gene expression analysis using RNA-seq data

The expression analysis using RNA-seq data of identified *NAC* genes were analysed and interpreted by heatmap through tbtools. RNA-seq is a very popular technique in making the transcriptome more readable in determining the quantity of transcripts, produced in population of cells at every instance in time. Gene expression level analysis and detection of (DE) differentially expressed genes or molecular pathways are one of the most frequently studied aims in an RNA-Seq experiment. Expression analysis is an essential tool that helps researchers decode the intricate regulatory networks that underpin biological processes.

CHAPTER-4

RESULTS

4.1 Results

Identification of *NAC* Gene family in Ricebean

Based on the study, 115 *NAC* genes of Ricebean were identified after filteration, manual curation and removal of duplicates through blastX.

Domain analysis

According to analysis conducted using (Pfam) databse of identified 115 Ricebean *NAC* genes, it was analyzed and identified that the non *NAC* domains in *NAC* genes were 10 in total and after the removal of non *NAC* domains a total of 103 *NAC* genes of Ricebean were obtained.

Physicochemical properties

Each Ricebean *NAC* protein were analyzed as shown in table 4.1, with its molecular weight, no of amino acids, stability index, theoretical pi, gravy score, aliphatic index and chromosome location were also taken in consideration through NCBI. The *NAC* gene family in Ricebean length ranging from 141 (LOC124830415) to 1117 (LOC124824377) amino acids, with predicted molecular weights that ranged from 36132.3(LOC12484814) to 19678.09 kDa (LOC124823072) . The instability index ranged from 21.09(LOC124824017) to 69.76(LOC124840572) and Aliphatic index ranging from 53.15(LOC124836353) to 108.89 (LOC124820079).

Table 4.1 Physiochemical characteristics of *NAC* genes at different LOC IDs

LOC ID	NO OF AMINO ACIDS	MOLECULAR WEIGHT	THEORETICAL PI	INSTABILITY INDEX	GRAVY SCORE	ALIPHATIC INDEX
LOC124824377	1117	121309.77	5.41	47.07	-0.163	94.62
LOC124826937	352	39824.93	8.2	34.82	-0.635	65.34
LOC124846879	189	21962.49	5.26	35.91	-0.767	67.51
LOC124847884	681	74647.81	6.87	51.22	-0.684	56.01
LOC124820079	405	44831.93	5.98	42.01	0.622	108.89
LOC124823064	172	20139.38	9.64	23.93	-0.692	69.13
LOC124823072	177	19678.09	6.01	37.79	-0.435	52.94
LOC124824027	595	65166.77	4.6	32.96	-0.381	76.08

LOC124824029	442	49185.8	4.39	32.78	-0.438	79.62
LOC124824017	392	43688.7	4.85	21.09	-0.268	81.07
LOC124824018	262	29083.12	6	29.14	-0.363	79.2
LOC124824028	462	51679.89	6.62	49.16	-0.731	64.83
LOC124824690	351	40111.23	7.72	40.8	-0.715	64.79
LOC124824712	359	40975.37	8.99	47.15	-0.65	64.37
LOC124824895	242	27627.94	6.01	57.31	-0.623	64.09
LOC124825303	693	75883.76	6.17	44.23	-0.759	56.91
LOC124826129	336	37251.98	9.07	33.64	-0.607	56.58
LOC124826397	336	38391.51	9.92	40.35	-0.506	80.95
LOC124827647	371	42077.64	5.13	50.19	-0.655	61.7
LOC124827646	481	54262.68	5.42	44.9	-0.509	79.13
LOC124827681	379	41926.06	8.77	35.63	-0.68	64.56
LOC124828265	413	46809.45	6.34	45.5	-0.59	63.78
LOC124828266	413	46809.45	6.34	45.5	-0.59	63.78
LOC124828365	346	39175.28	6.05	43.14	-0.685	65.72
LOC124828779	245	27948.56	8.81	36.93	-0.768	65.59
LOC124828776	279	31893.05	8.71	39.16	-0.77	65.63
LOC124829482	339	38134.95	8.11	34.38	-0.547	67.2
LOC124829580	263	30076.73	8.97	55.72	-0.665	67.07
LOC124829577	296	33798.75	8.22	60.34	-0.59	70.14
LOC124829607	585	64705.01	6.1	44.29	-0.815	56.72
LOC124829730	662	72985.31	6.75	52.03	-0.646	59.41
LOC124829978	280	31864.93	8.75	46.78	-0.816	66.43
LOC124830175	355	39917.73	9.13	51.53	-0.696	60.14
LOC124830395	278	32016.83	9.63	27.93	-0.608	75.72
LOC124830415	141	16308.76	9.32	31.25	-0.587	72.55
LOC124830660	286	32948.2	6.13	38.82	-0.761	65.45
LOC124831180	900	99170.63	6.31	42.49	-0.424	67.18
LOC124831228	336	38087.82	6.26	38.34	-0.568	67.13
LOC124831910	329	36652.22	7.01	44.24	-0.527	69.33
LOC124832504	178	20241.25	5.43	64.84	-0.779	56.52
LOC124832542	334	38170.62	5.6	52.56	-0.785	50.6

LOC124832819	276	31163.14	5.94	53.5	-0.455	79.78
LOC124832950	309	35416.47	7.6	40.84	-0.823	60.52
LOC124833182	310	35173.59	6.71	45.03	-0.755	65.39
LOC124833425	257	29657.9	5.67	55.84	-0.96	54.63
LOC124833512	573	63974.56	5.03	40.66	-0.718	65.5
LOC124833900	353	39229.68	8.67	45.42	-0.745	57.17
LOC124834403	295	34126.56	5.95	32.56	-0.76	63.53
LOC124834615	190	21642.98	5.73	58.78	-0.182	91.79
LOC124834948	365	41074.5	6.63	41.15	-0.54	71.37
LOC124836002	305	35442.28	8.78	31.03	-0.684	63.87
LOC124836059	451	51071.86	6.66	49.71	-0.916	59.2
LOC124836076	103	11377.87	4.84	38.63	-0.507	76.7
LOC124836145	285	32855.76	5.83	38.88	-0.628	62.25
LOC124836318	234	26466.85	8.7	41.59	-0.645	57.86
LOC124836353	343	38732.87	5.14	44.23	-0.734	53.15
LOC124836354	229	25152.35	5.51	36.49	-0.308	66.94
LOC124836800	369	41351.28	6.27	37.87	-0.287	90.41
LOC124836801	237	27402.77	4.9	54.77	-0.721	59.24
LOC124836887	329	36636.29	4.87	38.83	-0.445	74.98
LOC124836896	204	23006.77	4.96	58.11	-0.552	63.58
LOC124837510	681	74647.81	6.87	51.22	-0.684	56.01
LOC124838002	559	62955.96	4.73	46.79	-0.485	71.82
LOC124838355	288	33243.23	6.62	32.6	-0.653	63.61
LOC124838420	422	48428.36	4.83	51.44	-0.401	74.71
LOC124839229	323	36110.28	6.4	35.4	-0.757	58.24
LOC124839918	282	32384.75	8.95	27.2	-0.649	70.53
LOC124839920	204	23725.04	9.01	40.25	0.567	114.71
LOC124839968	256	29443.5	7.6	31.14	-0.684	63.98
LOC124840066	363	40446.34	6.83	50.31	-0.549	58.93
LOC124840140	383	43307.6	6.61	45.16	-0.579	74.1
LOC124840572	395	45256.16	6.78	69.76	-0.873	58.48
LOC124841545	290	32653.02	6.42	49.24	-0.73	58.21
LOC124841542	509	57219.12	6.13	56.01	-0.463	78.33

LOC124841649	363	41797.74	6.41	48.93	-0.682	65.26
LOC124841644	363	41867.87	6.41	48.88	-0.668	66.58
LOC124841682	434	49266.2	5.86	36.45	-0.657	59.1
LOC124841792	455	50788.96	5.76	41.81	-0.762	58.75
LOC124841794	455	50788.96	5.76	41.81	-0.762	58.75
LOC124841875	434	49007.38	5.4	51.28	-0.715	66.24
LOC124842565	437	49226.47	8.06	36.46	-0.483	69.86
LOC124842566	437	49226.47	8.06	36.46	-0.483	69.86
LOC124843561	589	66968.26	4.73	46.8	-0.383	71.97
LOC124843625	411	46587.97	5.4	37.4	-0.647	65.5
LOC124843904	344	38678.15	6.18	47.32	-0.55	67.21
LOC124844055	449	50481.03	5.04	48.06	-0.696	71.22
LOC124844176	387	43496.72	6.45	38.43	-0.495	65.97
LOC124844204	400	46538.77	7.05	37.83	-0.635	62.17
LOC124844400	366	42749.38	6.08	46.43	-0.664	67.95
LOC124845973	425	48093.56	5.99	37.94	-0.708	67.48
LOC124846897	274	31053.25	8.28	52.8	-0.43	73.98
LOC124847171	370	42486.57	4.9	47.89	-0.606	70.86
LOC124847174	370	42503.42	4.95	48.58	-0.616	70.86
LOC124847699	381	43621.26	4.88	49.95	-0.595	61.23
LOC124847816	325	37215.91	5.85	54.35	-0.62	63.35
LOC124848414	313	36132.3	4.22	45.24	-0.549	69.17
LOC124848550	223	25839.56	4.78	48.79	-0.654	70.36
LOC124819432	333	37659.58	7.16	40.3	-0.482	65.89
LOC124819490	366	42232.09	6.06	49.18	-0.707	62.4
LOC124819533	381	44339.36	5.68	46.42	-0.815	61.42
LOC124820250	480	54227.18	6.37	48.17	-0.787	64.81
LOC124820546	383	43086.27	7.21	46.75	-0.494	64.46
LOC124820786	404	46199.72	5.97	39.55	-0.626	59.43

Motif analysis: Total of 5 conserved motifs were investigated within *NAC* genes using MEME tool. Figure 2 constitutes number of motifs in each *NAC* gene and Figure 3 depicts the motif location of each Ricebean *NAC* protein members.

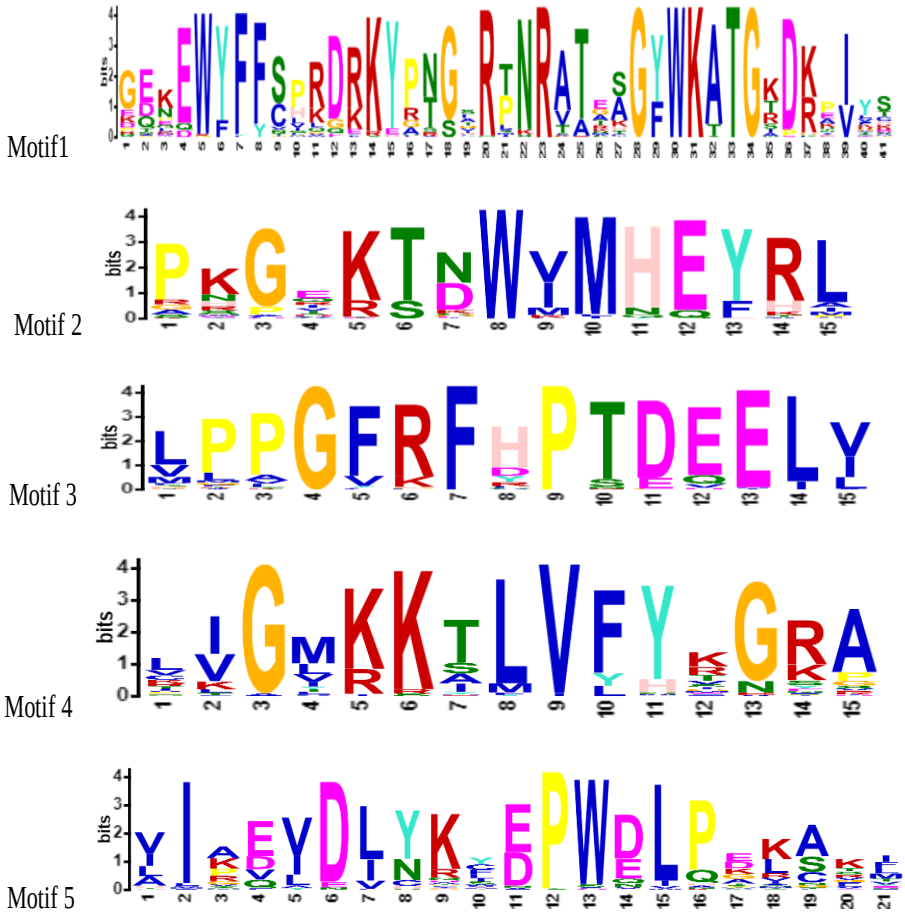
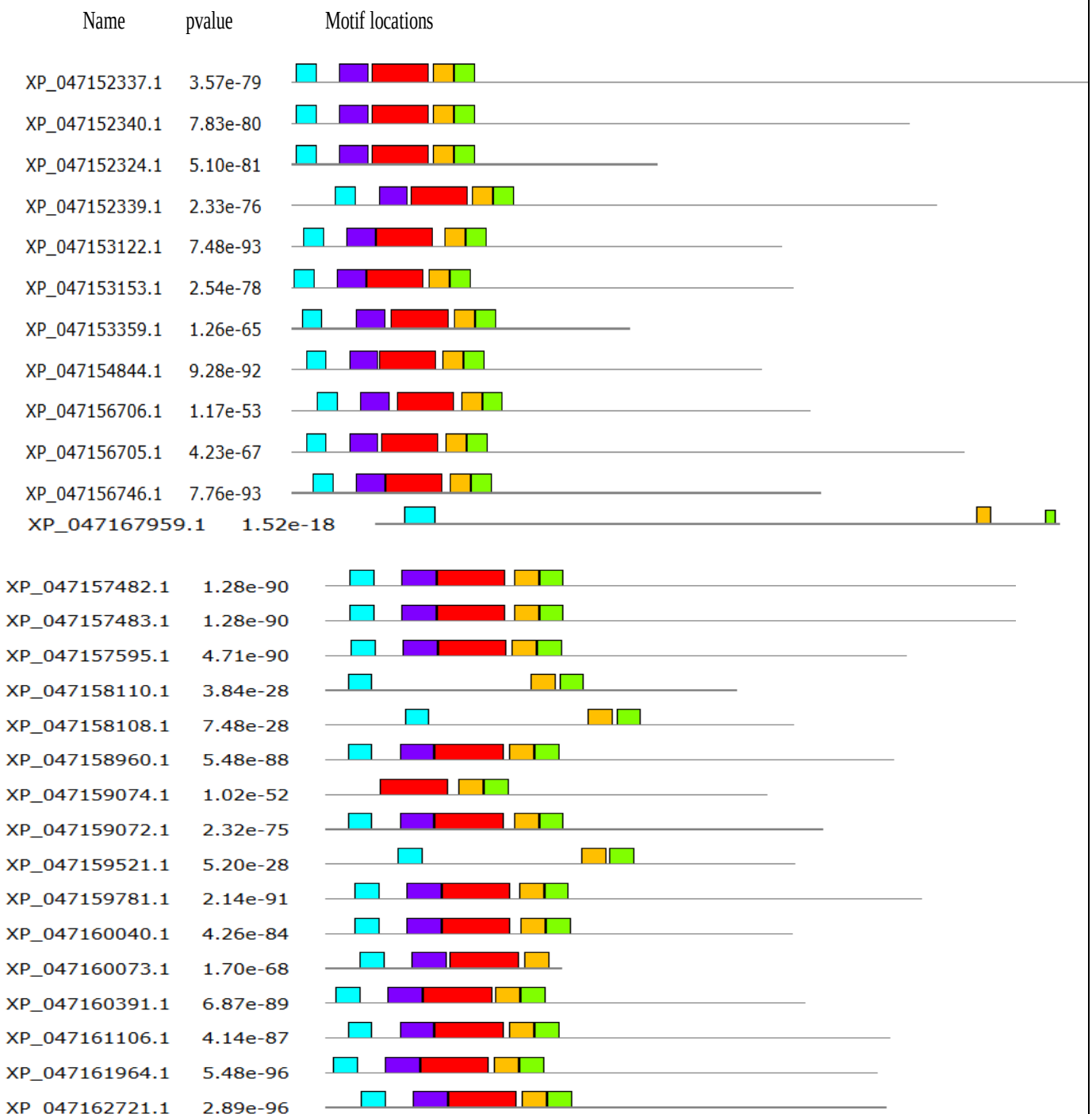
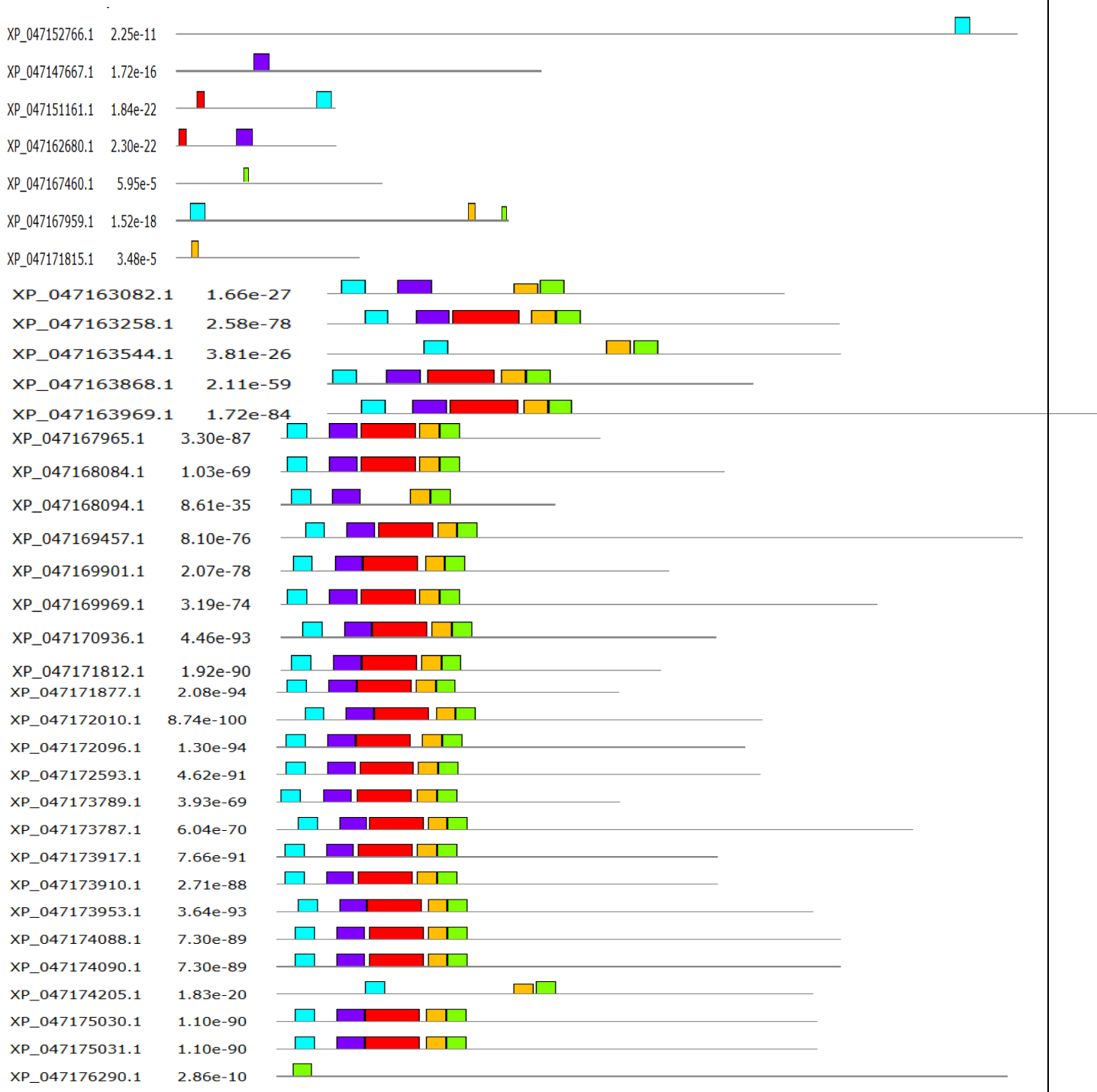


Figure 4.1 Motif logos


Name
pvalue
Motif locations
XP_047152337.1 3.57e-79
XP_047152340.1 7.83e-80
XP_047152324.1 5.10e-81
XP_047152339.1 2.33e-76
XP_047153122.1 7.48e-93
XP_047153153.1 2.54e-78
XP_047153359.1 1.26e-65
XP_047154844.1 9.28e-92
XP_047156706.1 1.17e-53
XP_047156705.1 4.23e-67
XP_047156746.1 7.76e-93
XP_047167959.1 1.52e-18
XP_047157482.1 1.28e-90
XP_047157483.1 1.28e-90
XP_047157595.1 4.71e-90
XP_047158110.1 3.84e-28
XP_047158108.1 7.48e-28
XP_047158960.1 5.48e-88
XP_047159074.1 1.02e-52
XP_047159072.1 2.32e-75
XP_047159521.1 5.20e-28
XP_047159781.1 2.14e-91
XP_047160040.1 4.26e-84
XP_047160073.1 1.70e-68
XP_047160391.1 6.87e-89
XP_047161106.1 4.14e-87
XP_047161964.1 5.48e-96
XP_047162721.1 2.89e-96


XP_047152766.1 2.25e-11
XP_047147667.1 1.72e-16
XP_047151161.1 1.84e-22
XP_047162680.1 2.30e-22
XP_047167460.1 5.95e-5
XP_047167959.1 1.52e-18
XP_047171815.1 3.48e-5
XP_047163082.1 1.66e-27
XP_047163258.1 2.58e-78
XP_047163544.1 3.81e-26
XP_047163868.1 2.11e-59
XP_047163969.1 1.72e-84
XP_047167965.1 3.30e-87
XP_047168084.1 1.03e-69
XP_047168094.1 8.61e-35
XP_047169457.1 8.10e-76
XP_047169901.1 2.07e-78
XP_047169969.1 3.19e-74
XP_047170936.1 4.46e-93
XP_047171812.1 1.92e-90
XP_047171877.1 2.08e-94
XP_047172010.1 8.74e-100
XP_047172096.1 1.30e-94
XP_047172593.1 4.62e-91
XP_047173789.1 3.93e-69
XP_047173787.1 6.04e-70
XP_047173917.1 7.66e-91
XP_047173910.1 2.71e-88
XP_047173953.1 3.64e-93
XP_047174088.1 7.30e-89
XP_047174090.1 7.30e-89
XP_047174205.1 1.83e-20
XP_047175030.1 1.10e-90
XP_047175031.1 1.10e-90
XP_047176290.1 2.86e-10

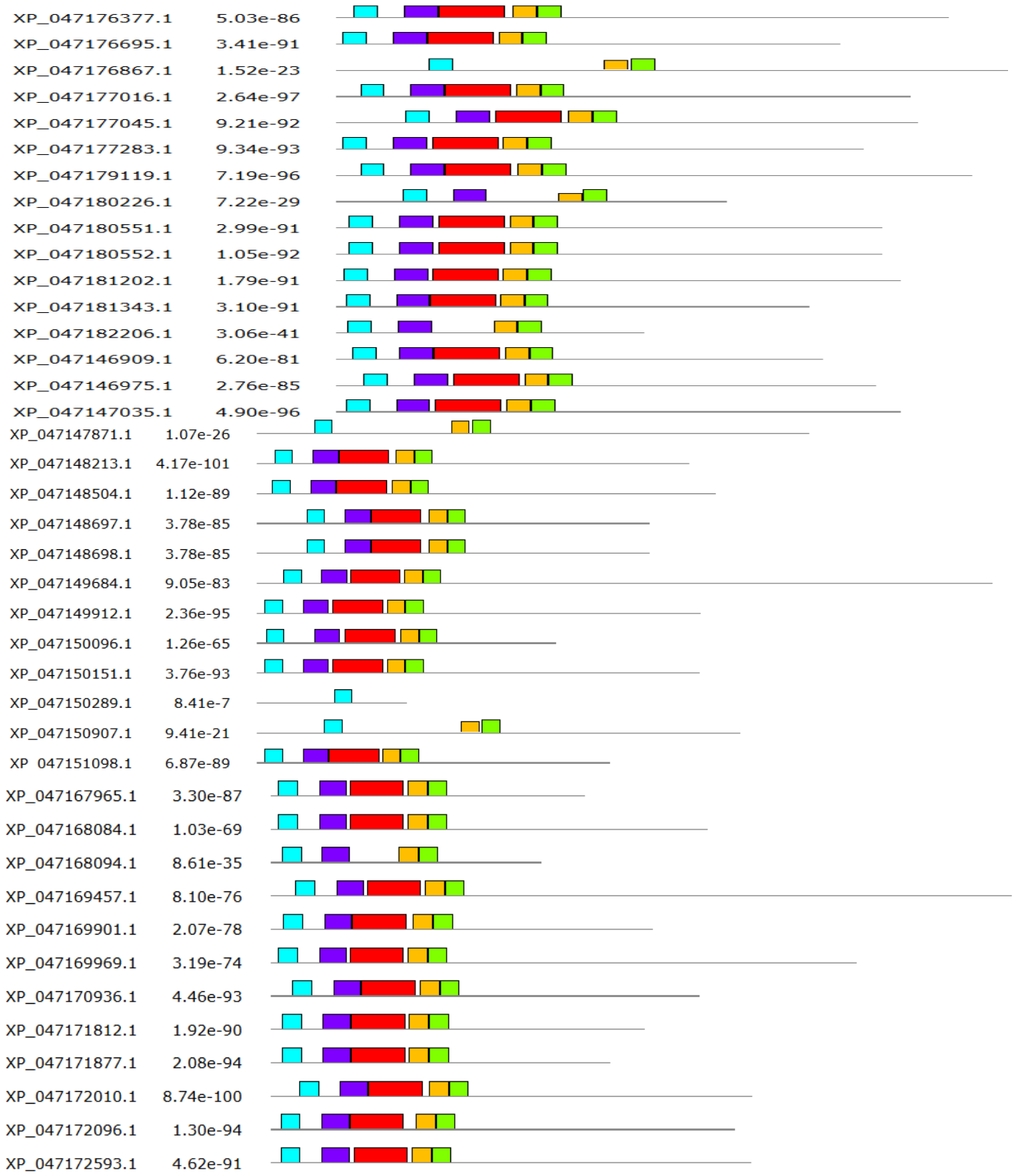


Figure 4.2 Motif locations of Ricebean *NAC* proteins using MEME tool.

Phylogenetic analysis of *NAC* genes

In order to perform Multiple Sequence Alignment the protein sequences of *Vigna unguiculata, Vigna angularis, Cicer arietinum ,Glycine max, Arabidopsis thaliana* and *Vigna radiata* as a query and were aligned using MAFFT as well as NJ/UPGMA phylogeny. Phylogenetic tree help us to understand the evolutionary relationship among species. Next, to create, visualize and annotate the phylogenetic tree iTOL (Interactive tree of Life) was used.

Comparative phylogenetic study based on *NAC* protein sequences are shown in figure 4.3 an the unrooted tree.Phylogenetic tree shows the evolutionary relationship among the species (figure 4.4).

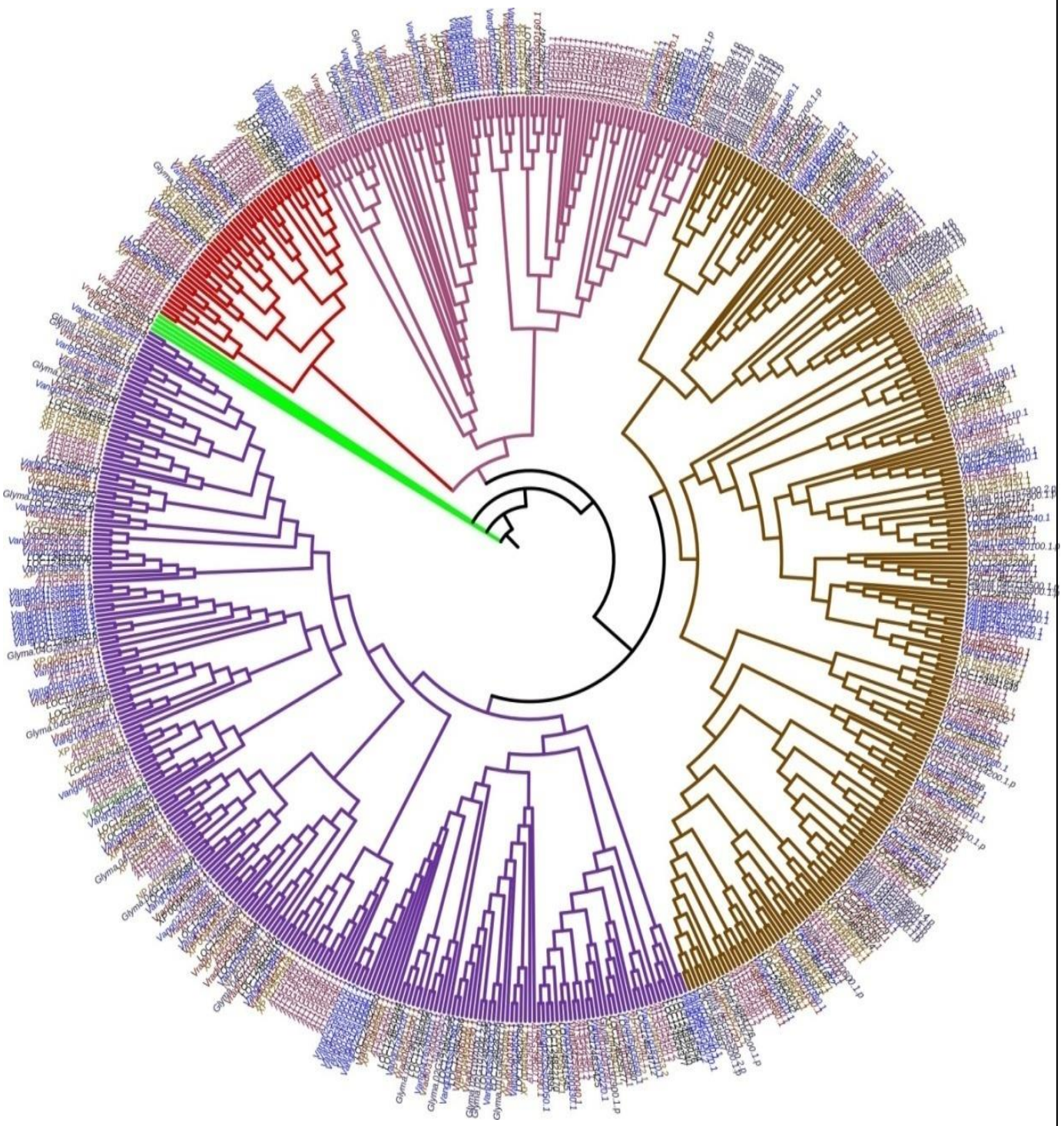


Figure 4.3 Comparative phylogenetic analysis based on *NAC* protein sequences of ***Vigna unguiculata, Vigna angularis, Cicer arietinum, Glycine max, Arabidopsis thaliana*** **and** ***Vigna radiata*** which are represented by different colors.

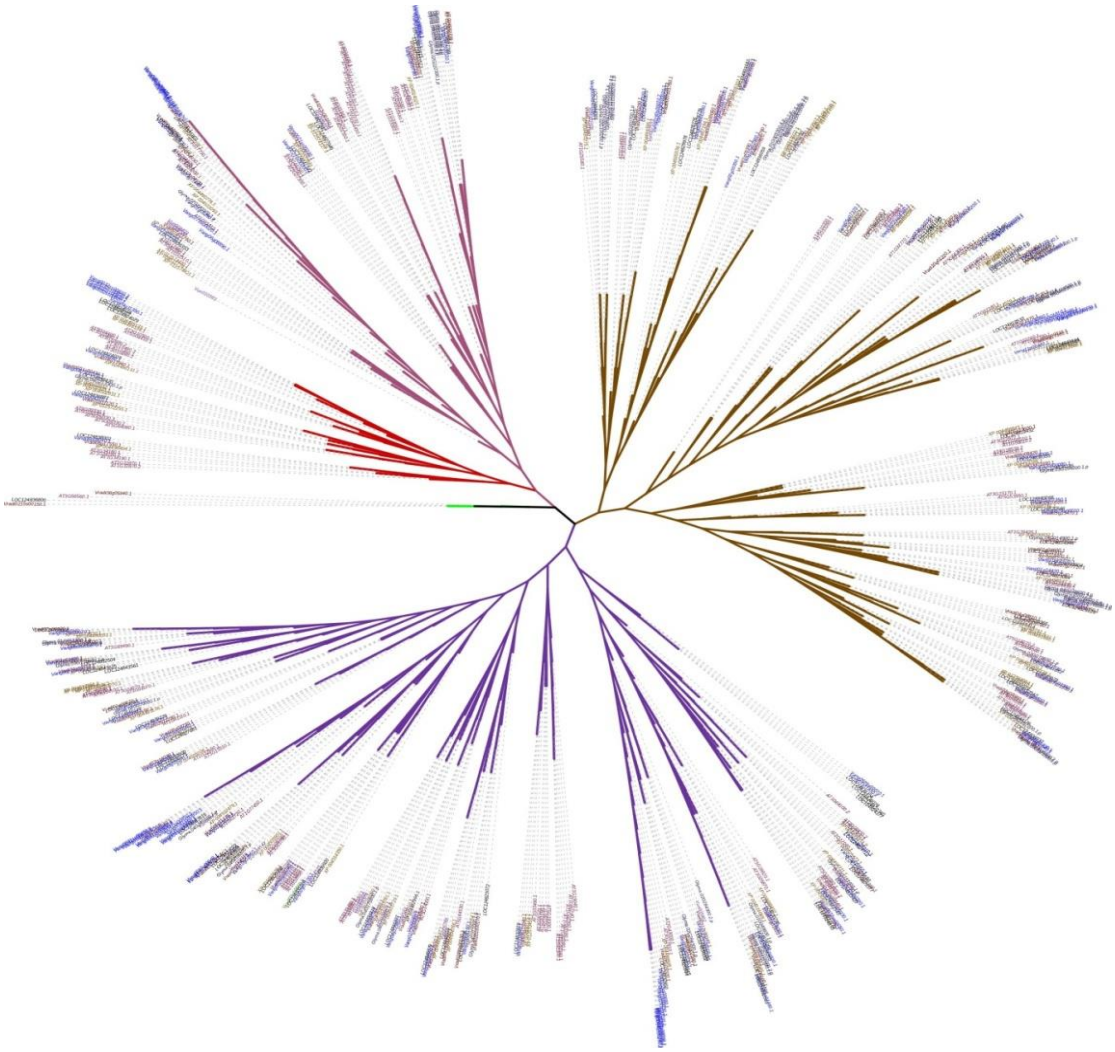


Figure 4.4 Phylogenetic tree in unrooted type

Gene structure analysis of *NAC* genes family

To understand the gene structure of *NAC* gene family, we have look at the information of introns and exons of the genes. The exon-intron structures of Ricebean *NAC* genes were examined using GSDS2.0 server. These exons and introns give a clear exposure of how *NAC* gene families are

undergoing progression. Analysis of structure of exons and introns has had studies on how genes are characterized.

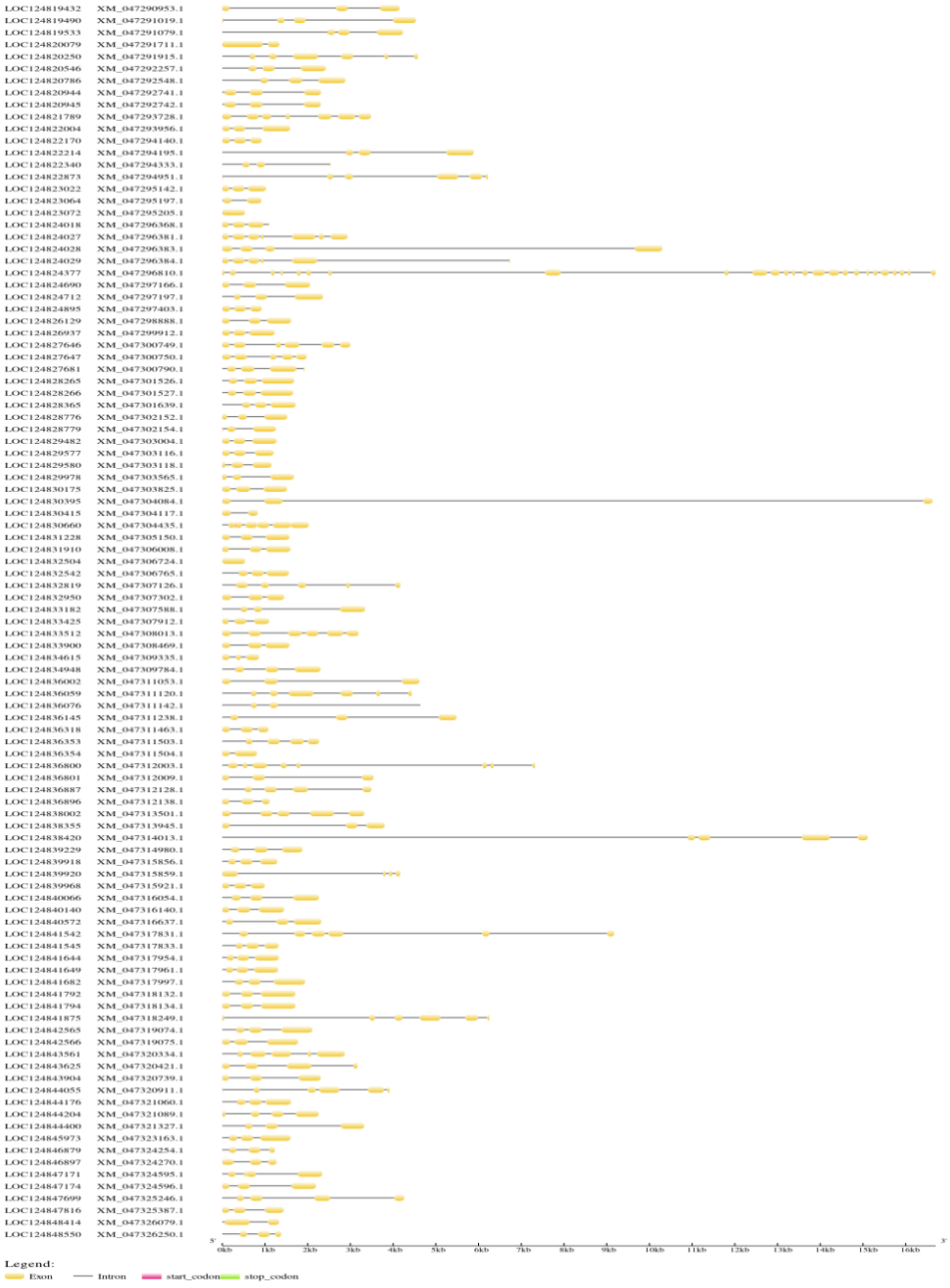


Figure 4.5 Gene structure examination representing the exon/intron distribution in *NAC* genes with yellow colour representing exons (coding regions), black colour represents(non- coding regions).

Synteny analysis of *NAC* genes with Ricebean

Synteny analysis was employed to classify homologous genes and evolutionary correlation amongst genes. To discover the evolutionary connection of Ricebean with various other plant species, genomes for each *NAC* protein sequences like *Vigna unguiculata* ,*Vigna angularis, Cicer arietinum* ,*Glycine max, Arabidopsis thaliana* and *Vigna radiata* were taken to perform synteny between Ricebean with other plant species .Synteny analysis shows an evolutionary relationship and least synteny is between Ricebean with *Vigna unguiculata* ,*Vigna angularis, Cicer arietinum Glycine max, Arabidopsis thaliana* and *Vigna radiata.*

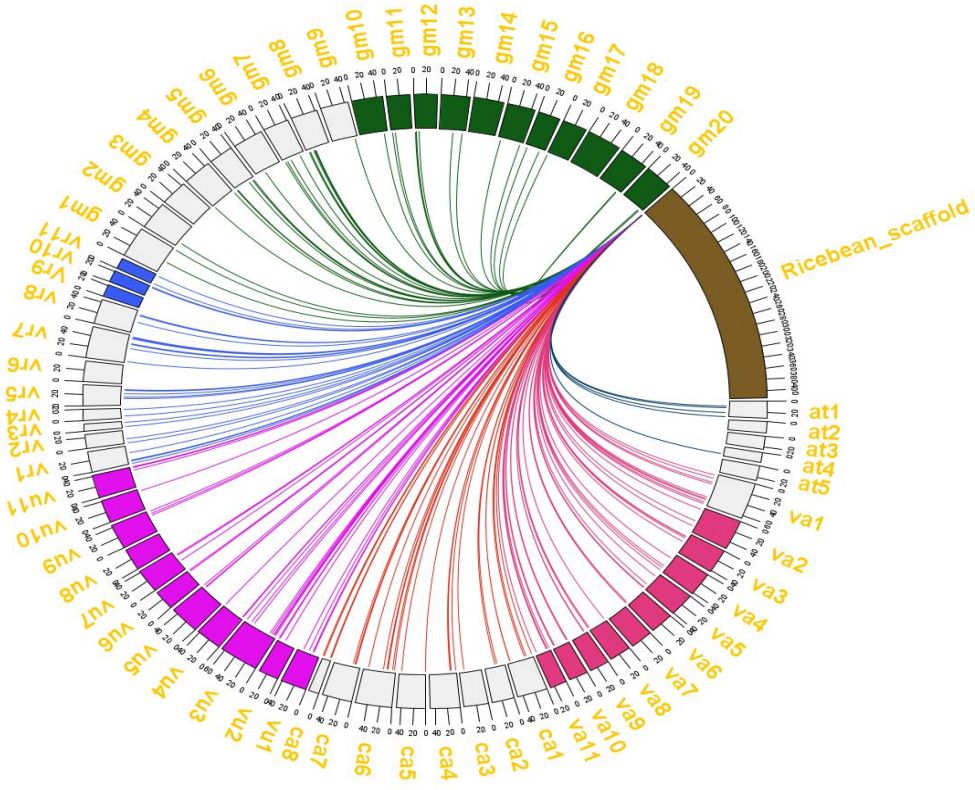


Figure 4.6 Circos plot illustrating the syntenic block between the scaffolds of Ricebean with other plant species such as *Vigna unguiculata, Vigna angularis, Cicer arietinum, Glycine max, Arabidopsis thaliana* and *Vigna radiata*.

Analysis of Promoter Cis-Acting Elements

We analyzed cis-acting elements in Ricebean *NAC* through PlantCARE website to identify specific types and distribution of these elements. GT-1 motif, I box, LAMP element, MBS, MRE, MYB recognition site, STRE, Wbox, WRE3, AuRR cre, TATA box, TATC- box and TGA element were some of the cis regulatory elements.

Figure 4.7 represents the Heatmap of Ricebean *NAC* genes representing cis regulatory elements, and a total of 103 cis acting regulatory elements were observed and represented through heatmap by tbtools.

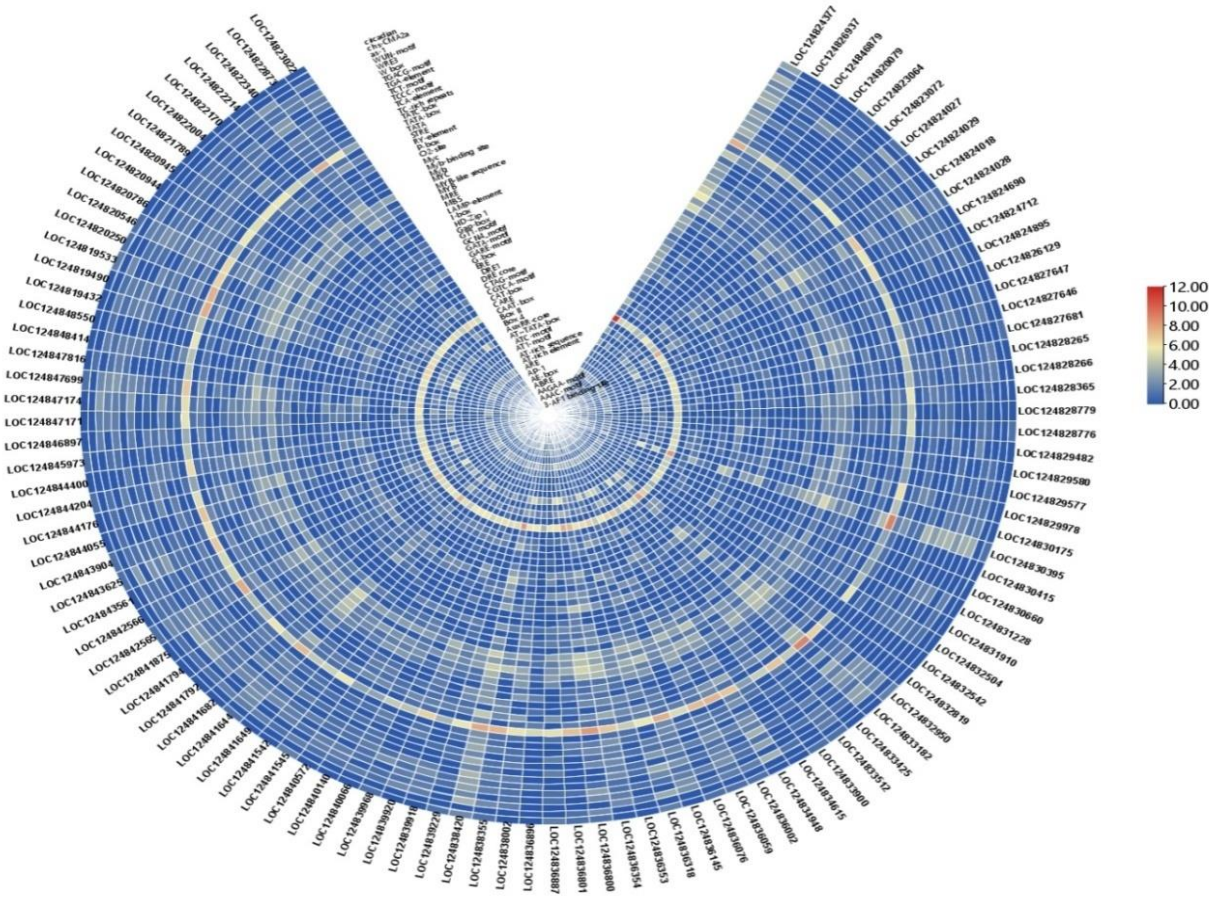


Figure 4.7 Ricebean *NAC* genes representing cis-acting elements were analyzed using heatmap. Colour scale constitutes intensity of cis acting elements.Red colour indicates higher abundance of cis regulatory elements and yellow colour or blue represents lower abundance.

Expression analysis of Ricebean *NAC* genes

Expression profiles of genes using RNA-seq data was conducted using tbtools. RNA-seq data shows gene expression levels of 103 *NAC* genes of Ricebean (*Vigna umbellata*) in two Ricebean genotypes with different pod sizes (B5_B10) bold genotype and (S5_S10) small genotype at two distinct stages of seed development:

5 days post-anthesis (DPA) and 10 DPA: Inclusive Transcriptome analysis data displayed differentially expressed genes were found in small genotype and bold genotype of the two Ricebean genotypes at 5-DPA and 10 DPA. Figure 4.8 represents the RNA-seq data represented by heatmap which shows that the expression profiles revealed that all 103 *NAC* genes of Ricebean exhibited discrete expression during developmental stages.

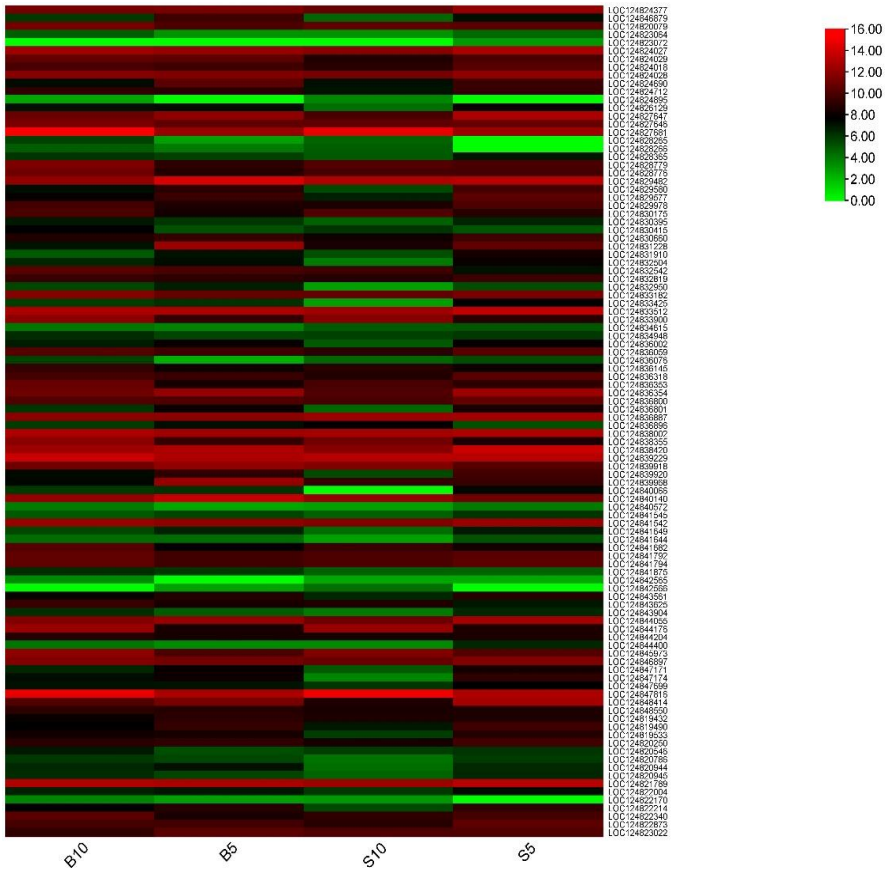


Figure 4.8 Expression pattern of *NAC* genes using Heatmap.Colour scale represents the relative expression levels of multiple genes across different samples or conditions. Red color shows high intensity, green shows lower intensity, and black colour shows intermediate gene expression level.

CHAPTER -5

CONCLUSION

5.1 Conclusion

In conclusion we presented a detailed genetically analyzed work for the investigation of the *NAC* gene family in *Vigna umbellata*. In the beginning, we overviewed the parameters characterize these gene families in the line of their physicochemical properties, (MSA) Multiple sequence alignment, conserved motif analysis, Phylogenetic evolutionary linkage, gene structure, cis acting element analysis, synteny analysis and gene expression using RNA-se data were all identified by means of comprehensive analyses. Phylogenetic and synteny analysis that were conducted could be used in identifying evolutionary relationship among the species.In this study, 103 *NAC* genes were analysed in Ricebean and comparison was made between other legumes to prove that Ricebean has independent evolutionary lineages and contains conserved motifs. Consequently, this study systematically investigates the involvement of phylogenetic, conserved motif and cis acting element analyses to achieve an extensive analysis of the *NAC* gene family, which is an key finding for additional studies on the functional diversification of *NAC* genes in members of legumes and possibly other plant species. It offers theoretical support for cultivating higher-stress-tolerant Ricebean and plants with increased productivity. Based on findings from comparative investigations of *NAC* genes in other plants, complete genome-wide identification in Ricebean using modern techniques will be important for identifying and understanding the regulatory mechanisms of individual *NAC* genes, creating stress-resistant transgenic plants, conducting thorough transcriptome analysis, genetic enhancement use in the study of frontier research in comparable related gene families.

5.2 FUTURE PROSPECTIVE

Genetic Tools and Techniques nowadays are the most Progressive. CRISPR-Cas9 and Gene Editing: Tomorrow's research will certainly take advantage of CRISPR-Cas9 technology to cut and paste the *NAC* gene sequences which will be able to carry out precise modifications on the Ricebean that increase resistance to stress and crops yield. This method can be considered a promising tool for ensuring the precise editing of gene functions and possible examination of the consequences it might have on the plants. Advanced Phenotyping and Genotyping Technologies: A pipeline that combines high-throughput phenotyping and genotyping technology will speed up the process of identifying and distinguishing *NAC* gene variants of Ricebean that manifest beneficial characteristics. The technologies will come in handy in the faster pinpointing of pioneering genetically selected plants . Bioinformatics and Computational Modeling: The more wide used bioinformatics tools and computational models are, the more one can estimate the influence of genetic modifications on the functioning of *NAC* genes. This type of analysis will help us in getting knowledge about gene network and their complex interactions with different environmental stimuli. Genomic Selection: Employing genomic selection techniques would allow breeders to predict a plant breeding value by analyzing *NAC* gene traits. It is hoped that this approach would aid in the selection of phenotypes that would be resilient to climatic stresses. Hybrid Breeding Techniques: Investigating hybrid breeding methods which carry *NAC* gene properties from different Ricebean varieties could result in the introduction of advanced Ricebean cultivars featuring exceptional resilience and high yield. Policy and Regulatory Frameworks Genetic Modification Regulations: With the revolution in the genetic engineering technologies, it will be indispensable to work on the regulation and ensure the true application of genetically modified Ricebeans. Regulations should definitely strive for balancing security with the specific crop improvement associated with *NAC* genes . International Collaboration: Enhancing international cooperation in studies on the *NAC* gene family will allow the sharing of knowledge and resources, thus increasing the rate at which stress-resilient Ricebean varieties appropriate for a variety of environments are developed.

References

1. Atta, K., Kousik, et al., "A review on stress physiology and breeding potential of an underutilized, multipurpose legume: *Ricebean* (*Vigna umbellata*)," *Developing climate resilient grain and forage legumes*, pp. 235-253, 2022.
2. H. Su, et al., "Genome-wide analysis and identification of stress-responsive genes of the NAM–ATAF1, 2–CUC2 transcription factor family in apple," *Plant Physiol. Biochem.*, vol. 71, pp. 11-21, 2013.
3. M. Nuruzzaman, A. M. Sharoni, and S. Kikuchi, "Roles of *NAC* transcription factors in the regulation of biotic and abiotic stress responses in plants," *Front. Microbiol.*, vol. 4, p. 248, 2013.
4. Z. Wang, and F. Dane, "*NAC* (NAM/ATAF/CUC) transcription factors in different stresses and their signaling pathway," *Acta Physiol. Plant.*, vol. 35, pp. 1397-1408, 2013.
5. C. Zhong, et al., "Genome-wide analysis reveals regulatory mechanisms and expression patterns of TGA genes in peanut under abiotic stress and hormone treatments," *Front. Plant Sci.*, vol. 14, p. 1269200, 2023.
6. R. Katoch, et al., "Nutritional and nutraceutical potential of *Ricebean* (*Vigna umbellata*)–a legume with hidden potential," *Front. Nutr.*, vol. 10, p. 1126544, 2023.
7. F. W. Lohoff, "Overview of the genetics of major depressive disorder," *Curr. Psychiatry Rep.*, vol. 12, pp. 539-546, 2010.
8. R. M. Hussain, et al., "The essence of *NAC* gene family to the cultivation of drought-resistant soybean (*Glycine max* L. Merr.) cultivars," *BMC Plant Biol.*, vol. 17, p. 1, 2017.
9. L. Wang, et al., "Genome-wide identification and characterization of GRAS genes in soybean (*Glycine max*)," *BMC Plant Biol.*, vol. 20, p. 1, 2020.
10. H. Arya, M. B. Singh, and P. L. Bhalla, "Towards developing drought-smart soybeans," *Front. Plant Sci.*, vol. 12, p. 750664, 2021.
11. S. Singh, H. Koyama, K. K. Bhati, and A. Alok, "The biotechnological importance of the plant-specific *NAC* transcription factor family in crop improvement," *J. Plant Res.*, vol. 134, no. 3, pp. 475-495, 2021.

12. C. Van Ha, et al., "The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress," *DNA Res.*, vol. 20, no. 5, pp. 511-524, 2013.
13. N. P. Thao, et al., "Differential expression analysis of a subset of drought-responsive Gm*NAC* genes in two soybean cultivars differing in drought tolerance," *Int. J. Mol. Sci.*, vol. 14, no. 12, pp. 23828-23841, 2013.
14. S. C. Mogali, and G. M. Hegde, "Recent advances in mungbean breeding: a perspective," *Accelerated Plant Breeding, Volume 3: Food Legumes*, pp. 235-282, 2020.
15. R. Singh, "Development of iron and zinc enriched mungbean (*Vigna radiata* L.) cultivars with agronomic traits in consideration," Wageningen University and Research, 2013.
16. M. Baza, D. Shanka, and M. Bibiso, "Agronomic and economic performance of mung bean (*Vigna radiata* L.) varieties in response to rates of blended NPS fertilizer in Kindo Koysha district, Southern Ethiopia," *Open Life Sci.*, vol. 17, no. 1, pp. 1053-1063, 2022.
17. R. Tariq, et al., "Genome-wide analyses of the mung bean *NAC* gene family reveals orthologs, co-expression networking and expression profiling under abiotic and biotic stresses," *BMC Plant Biol.*, vol. 22, no. 1, p. 343, 2022.
18. S. Zhang, et al., "Cloning and expression analysis of Vr*NAC*13 gene in mung bean," *Open Life Sci.*, vol. 18, no. 1, p. 20220627, 2023.
19. T. Zhu, et al., "Phylogenetic analyses unravel the evolutionary history of *NAC* proteins in plants," *Evolution*, vol. 66, no. 6, pp. 1833-1848, 2012.
20. R. Srivastava, and L. Sahoo, "Genome-wide analysis of cowpea *NAC* transcription factor family elucidating the genetic & molecular relationships that interface stress and growth regulatory signals," *Plant Gene*, vol. 31, p. 100363, 2022.
21. C. V. Ha, et al., "Genome-wide identification and expression analysis of the Ca*NAC* family members in chickpea during development, dehydration and ABA treatments," *PLoS One*, vol. 9, no. 12, p. e114107, 2014.
22. K. Jukanti, et al., "Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review," *Br. J. Nutr.*, vol. 108, suppl. 1, pp. S11-S26, 2012.
23. H. D. Chu, et al., "Identification, structural characterization and gene expression analysis of members of the Nuclear Factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments," *Int. J. Mol. Sci.*, vol. 19, no. 11, p. 3290, 2018.

24. P. Ruperao, "Analysis of the chickpea genome using next generation sequencing data," 2016.
25. W. Hu, et al., "Genome-wide identification and expression analysis of the *NAC* transcription factor family in cassava," *PLoS One*, vol. 10, no. 8, p. e0136993, 2015.
26. S. Singh, et al., "Comprehensive analysis and identification of drought-responsive candidate *NAC* genes in three semi-arid tropics (SAT) legume crops," *BMC Genomics*, vol. 22, p. 1, 2021.
27. H. Peng, et al., "Characterization of a chickpea (*Cicer arietinum* L.) *NAC* family gene, Car*NAC*5, which is both developmentally- and stress-regulated," *Plant Physiol. Biochem.*, vol. 47, nos. 11-12, pp. 1037-1045, 2009.
28. R. N. Kimotho, E. H. Baillo, and Z. Zhang, "Transcription factors involved in abiotic stress responses in Maize (Zea mays L.) and their roles in enhanced productivity in the post genomics era," *PeerJ*, vol. 7, p. e7211, 2019.
29. H. Shao, H. Wang, and X. Tang, "*NAC* transcription factors in plant multiple abiotic stress responses: progress and prospects," *Front. Plant Sci.*, vol. 6, p. 156056, 2015.
30. H. Shao, H. Wang, and X. Tang, "*NAC* transcription factors in plant multiple abiotic stress responses: progress and prospects," *Front. Plant Sci.*, vol. 6, p. 156056, 2015.
31. H. Mao, et al., "Zm*NAC*55, a maize stress-responsive *NAC* transcription factor, confers drought resistance in transgenic Arabidopsis," *Plant Physiol. Biochem.*, vol. 105, pp. 55-66, 2016.
32. M. Lu, Q. P. Sun, D. F. Zhang, T. Y. Wang, and J. B. Pan, "Identification of 7 stress-related *NAC* transcription factor members in maize (Zea mays L.) and characterization of the expression pattern of these genes," *Biochem. Biophys. Res. Commun.*, vol. 462, no. 2, pp. 144-150, 2015.
33. X. Peng, et al., "Genomewide identification, classification and analysis of *NAC* type gene family in maize," *J. Genet.*, vol. 94, pp. 377-390, 2015.

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 03/06/2024

Type of Document (Tick): PhD Thesis | M.Tech/M.Sc. Dissertation ✓ | B.Tech./B.Sc./BBA/Other

Name: Kritika Department: MSc Biotechnology Enrolment No 225111006

Contact No. 7876782535 E-mail. 225111006@juitsolan.in

Name of the Supervisor: Dr. Shikha Mittal

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): GENOME WIDE IDENTIFICATION OF NAC FAMILY IN RICE (OsA)

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 51
- Total No. of Preliminary pages = 7
- Total No. of pages accommodate bibliography/references = 3

KRITIKA
(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at 8 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) 04/06/2024

04/06/2024
Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
04/06/2024	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String	05%	Word Counts	9,288
			Character Counts	50,163
Report Generated on		Submission ID	Page counts	42
04/06/2024		2395281115	File Size	2.62 MB

Checked by 04/06/2024
Name & Signature

04.06.24
Librarian

LIBRARIAN
LEARNING RESOURCE CENTRE
Jaypee University of Information Technology
Waknaghat, Distt. Solan (Himachal Pradesh)
Pin Code - 173234

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at plagcheck.juit@gmail.com