Transcriptome-Wide identification and characterization of regulatory landscape of NAC Genes in Drimia indica.

Vivek Shit
Central University of Jharkhand
Mahesh Kumar Dhakar
ICAR-RCER: ICAR Research Complex for Eastern Region
Manoj Kumar (manoj@cuj.ac.in)
Central University of Jharkhand https://orcid.org/0000-0001-7945-7713

Research Article

Keywords: NAC transcription factors, Gene characterization, Phylogenetic classification, Motif analysis, Stress-responsive elements

Posted Date: August 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3202066/v1

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on November 30th, 2023. See the published version at https://doi.org/10.1007/s10722-023-01814-2.

Abstract

Drimia indica, a medicinal plant with promising therapeutic potential, aimed to understand the role of the NAC gene family in plant development and stress responses to enhance its medicinal properties and agronomic traits. We identified and characterized sixty-one non-redundant putative NAC genes, analyzing their physicochemical properties, which exhibited variations in amino acid composition, length, molecular weight, and isoelectric points. Subcellular localization prediction revealed diverse protein distributions, mainly within the nucleus. Phylogenetic analysis classified the NAC genes into 17 subgroups, showing distribution differences between *Drimia indica* and *Arabidopsis*. Gene structure analysis unveiled a conserved intron-exon organization within each subfamily. Motif analysis identified ten conserved motifs, with 'Motif 5' being the most prevalent. Promoter analysis detected cis-elements responsive to light, abscisic acid, methyl jasmonate, and MYB transcription factors. Additionally, transcription factor binding site analysis revealed several families potentially regulating NAC gene expression. MiRNA target analysis highlighted the significant role of miRNAs in the post-transcriptional regulation of NAC genes. Our findings offer valuable insights into the structural characteristics, regulatory mechanisms, and potential functions of the NAC gene family in *Drimia indica*, advancing our understanding of plant stress responses and suggesting future research avenues for stress tolerance and secondary metabolite production.

1. Introduction

Drimia indica, is a perennial herbaceous plant belongs to the Asparagaceae family. It has a long history of use as a traditional medicine,for its diverse therapeutic properties in treating respiratory problems and reducing inflammation and microbial infections. (Binny et al. 2012). Understanding the molecular mechanisms underlying the therapeutic potential of *Drimia indica*, necessitates an extensive investigation of its transcriptome, particularly the NAC domain gene family (Hu et al. 2010; Puranik et al. 2013). The NAC (NAM, ATAF1/2, and CUC2) domain transcription factors represent a diverse and important family of plant-specific transcription factors that play crucial roles in various essential biological processes, including plant development, growth, and stress responses. (Hong 2016; De Siqueira Pinto et al. 2019; Yan et al. 2021).

Transcription factors regulate the transcription of genetic information; these proteins can bind to specific DNA sequences in promoter regions, thus controlling the synthesis of messenger RNA. These factors encompass various families in plants, including WKRY, bZIP, MYB, DREB, AP2/EREBP, C2H2, MADS, and NAC, each characterized by a distinct DNA-binding domain (De Bodt et al. 2003; Gujjar et al. 2014). A typical NAC transcription factor (TF) consists of a conserved DNA-binding domain located at the N-terminus and a highly variable C-terminal region. The DNA-binding domains of NAC TFs typically consist of several subdomains, commonly referred to as A-E (Shen et al. 2020). These subdomains are comprised of 150 amino acid residues and have the ability for DNA binding and protein-protein interactions. The C and D subdomains within the DNA-binding domain contain net positive charges, enabling the NAC TF to bind specifically to cis-acting DNA sequences. This binding specificity allows NAC

TFs to regulate the expression of target genes by directly interacting with their promoter regions. The A subdomain allows it to form dimers with other NAC TFs or regulatory proteins, which is crucial for gene regulation. The "B" and "E" subdomains likely have diverse functions, possibly by interacting with other regulatory proteins.These DNA-binding domain recognizes and binds to specific genes. NAC TFs, also possesse a highly variable C-terminal region" known as the "transcriptional regulatory domain." This part of the protein can either activate or repress the expression of target genes. Some NAC TFs also have a transmembrane motif in their C-terminal region, allowing them to anchor to cell membranes and participate in signal transduction pathways.This localization facilitates their involvement in signal transduction pathways, connecting extracellular signals with transcriptional responses (Mohanta et al. 2020). The transcription factors are vital for multiple aspects of plant growth and development, including the formation of organs, secondary metabolism, hormone communication, and adaptation to environmental challenges (Wang et al. 2016). These proteins have been extensively studied in several plant species, but their comprehensive identification and characterization in *Drimia indica*, a medicinally important plant, remain largely unexplored (Kumar et al. 2021). Transcriptome-wide identification and characterization of the NAC domain gene family in *Drimia indica* can provide crucial insights into the regulatory mechanisms and potential roles of these transcription factors in plant development, stress responses and secondary metabolite production. By employing high-throughput RNA sequencing (RNA-seq) technologies and advanced bioinformatics tools, it is possible to decipher the entire repertoire of NAC genes in *Drimia indica* and unravel their putative functions.

Hence, the study endeavors to investigate the NAC domain gene family to bridge the existing knowledge gap in *Drimia indica*. By mining and analyzing the transcriptomic data, we aimed to analyse their sequence and structural characteristics, predict their protein domains and conserved motifs, perform phylogenetic analysis to determine their evolutionary relationships, and examine the presence of cis-acting elements in their gene promoters (Lv et al. 2016; Bar-Lev et al. 2021). The thorough investigation of the NAC domain gene family in *Drimia indica* will not only contribute to our comprehension of the regulatory networks that govern the growth and development of this medicinal plant, but also provide insights on the plant's adaptive responses to environmental challenges. Furthermore, the identified NAC genes that hold promise as potential candidates for further functional studies, including gene manipulation and breeding programmes aimed to enhance the medicinal properties and agronomic traits of *Drimia indica*. The results obtained from this research will serve as a significant asset for future studies in the field of plant molecular biology, medicinal plant research, and biotechnological applications.

2. Materials and Methods

2.1 Identification of NAC Family Genes in Drimia indica

The genomic sequences of *Drimia indica* were obtained from the NCBI SRA database (Not yet public). To identify the NAC genes, we employed the BLASTX program against the UniProt database, utilizing an e-value cutoff of 1e-5. The resulting sequences containing NAC genes were subjected to translation using

the Transdecoder package to obtain the CDS sequence. To reaffirm the accuracy of our search, we conducted a BLASTp analysis with a threshold e value of 1e-5, incorporating the NAC domain's hidden Markov model (HMM) profile, specifically the Pfam-A.hmm.gz file, as the query (Mistry et al. 2021). This additional step ensured a meticulous search with a stringent E value threshold of 1e-5. Sequence identifiers with exact matches to PF02365(NAM) and PF01849(NAC) are selected as candidates for further analysis. All protein sequences that contained the conserved NAC domain were examined, and genes encoding these proteins were identified as potential NAC gene candidates. To analyze the physicochemical properties of the DiNAC genes, we employed the ExPASy online program (https://web.expasy.org/protparam/). Additionally, the BUSCA web-based program (http://busca.biocomp.unibo.it/) was used for predicting the subcellular localization of the proteins (Gasteiger et al. 2005; Savojardo et al. 2018).

2.2 Phylogenetic Analysis and Classification of the DiNAC Gene Family

The identified NAC genes were subjected to phylogenetic analysis using MEGA 11.0.13 software (Kumar et al. 2018). A phylogenetic tree was constructed to explore the relationships between gene sequences of *Arabidopsis* and *Drimia indica* species (Lamesch et al. 2012). Sequence alignment was performed using the ClustalW algorithm with default parameters. To assess the robustness of the tree topology, the maximum likelihood (ML) method was utilized with 1000 repetitions for the bootstrap test. The resulting phylogenetic tree was visualized and analyzed using the iTOL (Tree Of Life) (Letunic and Bork 2021). This subsequent analysis provides insights into the evolutionary relationships of the DiNAC gens and its potential functional roles.

2.3 Sequence Analysis and Structural Characterization

The AUGUSTUS (version 3.3.3) programme was used to examine the exon/intron organisation of individual NAC genes in *Drimia indica*. This web-based programme aided in the investigation of the arrangement of exonic and intronic regions within the DiNAC gene sequence (Stanke et al. 2006). To examine the motifs in the DiNAC proteins the MEME tool (Version 5.5.2; https://meme-suite.org/meme/tools/meme) was used with a defined parameters; with a minimum motif width set to six, a maximum motif width of 50, and a target number of motifs to 10. (Bailey et al. 2015). TBtools (TBtools_v0.53) was used for data interpretation and visualisation. (Chen et al. 2020).

2.4 Predictions of Cis-Elements, TFBS, and miRNA Targets, in the DiNAC Gene Promoter

The gene coding sequence (CDS) from *Drimia indica* was retrieved and subjected to analysis. The sequence was then submitted to PlantCARE for further analysis of putative cis-elements (Rombauts et al. 1999). TFBS predictions were performed using Binding Site Prediction tool in the PlantTFDB database (http://planttfdb.cbi.pku.edu.cn/) with p-value < $1 \times 10 - 6$ (Riaño-Pachón et al. 2007). For miRNA target prediction, NAC gene sequences were submitted to the psRNATarget server to check the possible target

sites by miRNAs with default parameters (Dai et al. 2018). This analysis provided valuable information about the presence and distribution of specific cis-acting elements within the promoter regions of the gene. The analysis found different cis-acting elements in the CDS sequence, such as MeJA LR, ABRE, and MYB-LR (Zhang et al. 2020). These elements are involved in various biological processes, including response to methyl jasmonate (MeJA), light responsiveness, and MYB transcription factor binding. The results were visualized using the TBtools software.

3. Results

3.1 Identification of NAC Family Genes in Drimia indica

We utilized the NAC domain's hidden Markov model (HMM) profile obtained from the Pfam-A.hmm.gz file to retrieve the NAC transcription factor sequences in *Drimia indica*. 61 non-redundant putative NAC genes were initially identified, which were further analyzed to determine their physicochemical properties using the ExPASy online program. Supplementary File 1 provides detailed information on the NAC family genes in *Drimia indica*, including their names, identifiers, amino acid lengths, isoelectric points, molecular weights, and subcellular localizations (Supplementary Files 2 and 3). The composition of amino acids and physicochemical properties varied among the NAC family proteins, with varying numbers of amino acids. The protein sequences of the DiNAC genes ranged from 103 to 739 amino acids, with an average length of 287 amino acids. The predicted molecular weights ranged from 11,635.62 Da to 81,784.65 Da, with an average molecular weight of 32,296.12 Da. The isoelectric points ranged from 4.33 to 10.46, with an average pI of 7.15. The approx. half of the DiNAC family proteins (49.18%) had a pI below 7, indicating an acidic range and a higher proportion of acidic amino acids. The subcellular localizations of the *Drimia* NAC transcription factors were predicted using the BUSCA online software, revealing that one protein was localized in chloroplast, one in the endomembrane system, four in extracellular space, two in the plasma membrane, and 53 proteins were localized in the nucleus. Moreover, of the 61 DiNACs, only two contained a signal peptide, and three were transmembrane Alpha Helix, indicating that it has an important role in protein subcellular localization and signal transduction in the downstream process.

3.2 Classification and Phylogenetic Analysis of the DiNAC Gene Family

We conducted a phylogenetic study to explore the potential functions of the NAC gene family. It is believed that genes with similar evolutionary relationships often have similar structures. To clarify the phylogenetic relationships of NAC family proteins in *Drimia indica* and *Arabidopsis*, an unrooted tree was established based on the aligned NAC domain sequences. The results indicated that the phylogenetic tree clustered all NAC members into 17 subgroups, as shown in Fig. 1. DiNAC proteins were diverse as ANAC proteins and 61 DiNACs were unevenly distributed in 17 subgroups. In contrast, none of the members of the DiNAC gene family were detected in NAC-III, NAC-V, NAC-VI, NAC-XII, and NAC-XIII subgroups. These subgroups exclusively consist of *Arabidopsis* members, suggesting that they might have either been lost in *Drimia* plants or acquired in *Arabidopsis* after the divergence of their last common ancestor.

Additionally, these NACs in *Arabidopsis* may possess specialized roles. Therefore, studying the characteristics of this subgroup can be crucial for understanding the genetic relationships among different plant species. Genes that share a close evolutionary relationship often exhibit similar functions, and phylogenetic analysis is a valuable tool for predicting gene function, which is essential for subsequent functional studies. Predicting stress-related functional genes and secondary metabolite genes based on phylogenetic relationships has proven to be effective. The genes within the NAC-X subgroup were the published stress-related NAC family members. In *Arabidopsis*, there are 18 members (ANAC055, ANAC019, and ANAC072/RD26 etc.) of the NAC-X subgroup, all involved in responding to multiple stresses (Ling et al. 2023). However, we discovered that AT3G04070.1 (SHYG) is closely associated with leaf senescence in *Arabidopsis*. Senescence is a common phenomenon observed in plants when they are exposed to various severe stresses. These findings suggest that the NAC-X subgroup is associated with stress responses, and we propose that at least 11 NAC genes from this subgroup are present in *Drimia indica*, potentially induced by stress response as stress-related proteins. The members of subgroup II, including SOG1, NAC044, NAC085, NAC073, NAC010, SND5, and NAC075, have diverse functions. SOG1 regulates cell death and stress responses, while NAC044 is involved in gene expression. NAC085, NAC073, and NAC010 are regulators of leaf senescence, and SND5 contributes to secondary cell wall biosynthesis. Another member, NAC075, is a regulator of leaf senescence. Likely, in subgroup III, NAC084 is a transcription factor impacting various cellular processes. NAC095 regulates lateral root development, NAC065 is involved in programmed cell death. NAC093 and NAC064 contribute to lateral root development and nutrient uptake. In contrast, other NAC groups (VII, VIII, IX, X, XI, XIV, XV, XVI, and XVII) also contain various members in *Drimia indica*. Detailed information about the putative functions of these NAC members provided in the Supplementary File 8. Overall, these findings shed light on the diverse roles of NAC transcription factors in *Drimia indica*. To further support the reliability of the subgroup classification, we also examined other evidence, such as gene structure and patterns of motif compositions.

Figure 1: Illustrates the phylogenetic tree depicting the relationship between NAC proteins from *A. thaliana* and *D. indica*. The phylogenetic tree was generated using MEGA 11.0 through the maximum likelihood (ML) method with 1000 bootstrap replicates. The NAC proteins were classified into 17 distinct subgroups. "DiNACs (DIL & DIR)" refers to the NAC proteins derived from *D. indica*, while "ANACs" denotes the NAC proteins from *A. thaliana*.

3.3 Sequence Analysis and Structural Characterization

Plants have a wide range of protein sequences and structures, developed through evolution, leading to diverse multigene families and functions. This diversity helps plants efficiently utilize natural resources and adapt to challenging environments. To understand the structural diversity of NAC proteins in the *Drimia* plant, we aligned their sequences, predicting conserved motifs, gene structures, and phylogenetic relationships (Fig. 2).

To examine the structure of NAC genes, we analyzed their RNA sequences and determined the arrangement of their introns and exons. AGUSTUS software was employed to map the intron-exon structure of the NAC gene family. The findings revealed that none of the NAC genes contained introns. Among these genes, six were found to possess coding sequences (CDS) with a 3' untranslated region (UTR), and one gene occupied the 5' UTR region. Analysis of the gene structure demonstrated significant variations in length among the NAC gene members. The shortest NAC gene measured only 778 base pairs (bp) in length (DIR_DN515387_c0_g1_i2), while the longest gene (DIL_DN4028_c0_g1_i11) spanned 2462 bp. The gene structure of the NAC gene family exhibited moderate conservation across different subfamilies, with similar numbers and locations of exons among the NAC genes within each subfamily, indicating potential functional similarities (Fig. 1).

Additionally, we employed the MEME program to identify conserved motifs within the NAC domain, leading to the definition of ten conserved motifs (Supplementary File 4). Members with close evolutionary relationships displayed consistent or similar compositions of motifs. The analysis uncovered several significant motifs that were conserved across the sequences. The anticipated motif pattern for the *Drimia* members of subgroup NAC-X exhibited a similar distribution in a comparative study, which aligned with the phylogenetic analysis within the group. Furthermore, in a subsequent analysis, Motif 5 emerged as the most prevalent motif, present in the majority of the sequences. Motifs 8 and 10 were also identified, although they occurred less frequently. Almost all conserved motifs were found in the N-terminal conserved domain region of the DiNAC proteins, with only a limited occurrence in the diverse C-terminal region of the NAC protein. This distribution of conserved motifs suggests that the functional characteristics of NAC proteins are primarily determined by the N-terminal conserved domain, while the C-terminal region may also play a role in determining the functions of these proteins. The consensus sequences and their corresponding positions within the NAC domain were determined. The identified motifs provide valuable information about the conserved regions and potential functional elements within the NAC domain in our dataset. These results contribute to a better understanding of the regulatory mechanisms and functional roles of NAC transcription factors in the studied system.

Figure 2: Depicts the structural Diversity of NAC Proteins in the *Drimia* Plant. (A) A phylogenetic tree was constructed using MEGA 11.0, illustrating the relationships among DiNAC proteins. (B) Visualization of the motif composition and relative motif lengths within the NAC protein, with distinct coloured boxes to represent each motif. (C) and the organization of exons and introns within the NAC gene.

3.4 Predicted Regulatory Mechanism of the 61 NAC Genes

Gene expression responses are largely related to their promoters; therefore, we investigated the putative stimulus-responsive cis-elements in the promoter regions of all of the DiNAC genes (Supplementary File 5). In this study, we analyzed the presence of four responsive elements, namely LR (Light Responsive), ABRE (Abscisic Acid Responsive), MeJA (Methyl Jasmonate Responsive), and MYB-LR (MYB Transcription Factor Responsive), within the promoters of target genes. These responsive elements are

known to play important roles in regulating gene expression in response to various stimuli and are associated with multi-stress responses and secondary metabolite production (Fig. 3) (Nicolas et al. 2014; Wasternack and Strnad 2019).

Our analysis revealed that a significant number of gene promoters contained one or more of these responsive elements. Specifically, we observed that LR-responsive elements were present in 38.99% of the gene promoters, indicating their potential involvement in light-mediated stress responses and secondary metabolite biosynthesis. Furthermore, ABRE-responsive elements were identified in 22.20% of the gene promoters, suggesting their role in abscisic acid-mediated stress responses and the regulation of secondary metabolite synthesis. The presence of MeJA-responsive elements in 37.68% of the gene promoters indicates their involvement in the jasmonate signalling pathway, associated with various stress responses and the production of secondary metabolites. Interestingly, MYB-LR-responsive elements were found in 1.11% of the gene promoters. MYB transcription factors are known to regulate diverse biological processes, including stress responses and secondary metabolite biosynthesis. The presence of MYB-LR-responsive elements suggests their potential role in coordinating multi-stress responses and the regulation of secondary metabolite production.

Overall, our analysis of Light responsive elements (LR), ABA-responsive elements (ABRE), Methyl Jasmonate (MeJA), and MYB Light Responsive (MYB-LR) responsive elements within the gene promoters provides valuable insights into the regulatory mechanisms underlying multi-stress responses and secondary metabolite synthesis. These findings highlight the potential involvement of these responsive elements in orchestrating complex signalling networks that contribute to plant adaptation and the production of bioactive compounds. Further functional characterization of these responsive elements and their associated transcription factors will enhance our understanding of plant stress responses and facilitate the manipulation of secondary metabolite pathways for agricultural and pharmaceutical applications. TFBS analysis showed that there were 612 transcription factors of 44 families which were predicted to bind to the promoter regions of 61 NAC genes. (Supplementary File 6). The transcription factors (TFs) encompassed a range of families, including GATA, Dof, HD-ZIP, TCP, MIKC_MADS, bHLH, ERF, MYB, and C2H2, which were associated with diverse aspects of developmental regulation and responses to environmental stimuli. Many evidences showed that miRNAs play important roles in plant development regulation (Zhu et al. 2011; Lima et al. 2012; Sunkar et al. 2012). To check whether miRNA also participated in stress-responsive regulation, these 61 NAC gene sequences were used to scan potential miRNA targets in the online software psRNATarget with all published *Arabidopsis* 427 miRNA sequences. The analysis of miRNAs revealed that almost every single one of the 61 NAC genes examined has potential target sites for 203 miRNAs (Supplementary File 7). Further investigation into the mode of miRNA-mediated inhibition identified two main types of inhibitory mechanisms: cleavage and translation inhibition. Cleavage inhibition triggers subsequent degradation of the target mRNA and prevents the translation of the mRNA into a functional protein, and is predominantly observed in plants. In contrast, translation inhibition miRNAs bind to target mRNAs without inducing cleavage. Instead, the miRNA binding forms inhibitory complexes that impede the translation process at various stages, such as

blocking ribosome binding, disrupting translation initiation, or promoting ribosome drop-off during elongation.

The identification of these inhibitory mechanisms provides valuable insights into the regulatory roles of miRNAs in gene expression. By targeting specific mRNAs, miRNAs contribute to the fine-tuning of gene expression levels and participate in a wide range of biological processes and regulatory networks.

Figure 3: Identification of cis-responsive elements for DiNAC genes

4. Discussion

Transcription factors are essential in regulating gene expression and influencing physiological processes. There is a growing focus on identifying and characterizing these factors in plants, especially in response to drought, salt, low temperatures, hormones, and pathogens. These factors can enhance plant resistance and adaptability to various stress conditions. Recently, NAC genes (NAM, ATAF1/2, and CUC2) have gained significant attention for their roles in medicinal plants to understand potential functions and molecular mechanisms behind their therapeutic properties. The availability of the *Drimia indica* transcriptome sequence, generated by our research group, has allowed us to conduct extensive bioinformatic analysis and shed light on the structural and functional aspects of the discovered NAC genes.

In this study, we carried out a thorough analysis of the NAC gene family in the *Drimia* plant, focusing on several different factors, including physicochemical properties, subcellular localizations of proteins, phylogenetic relationships, identification of conservative motifs, intron-exon organization, promoter cis-acting elements, Transcription Factor Binding Site identification, and miRNA Targets, in the DiNAC Gene Promoter. We identified 61 DiNAC genes and found that the NAC gene family contains many acidic amino acids. Additionally, it was found that the NAC proteins' predicted isoelectric points (pI) were lower than 7, indicating an acidic pH range. Suggesting their acidic nature and may have particular functional properties, activities, or conditions within an acidic cellular environment. The presence of acidic amino acids in the protein sequences, such as aspartic acid and glutamic acid, contributes to the low pI values observed. Furthermore, Only two of the detected 61 DiNACs (*Drimia indica* NAC genes) have a signal peptide, and three belong to the transmembrane alpha helix. The presence of a signal peptide suggests that they may be involved in targeting the protein to specific cellular compartments or organelles, while the presence of a transmembrane alpha helix suggests that they may be associated with cell membranes and participate in membrane-related signaling pathways. The discovery of these traits adds to our understanding of the DiNAC proteins' functional variety, cellular localization, and their roles in *Drimia indica*. In addition, the comparison of the phylogenetic trees of *Drimia* and *Arabidopsis* revealed that the NAC gene family members were unevenly distributed among subfamilies in both species. This finding suggests that the NAC gene family existed before the two species diverged, suggesting a shared ancestor for the genes within the subfamilies and pointing to probable functional similarities. Due to their exposure to distinct environments throughout evolution, *Drimia* and *Arabidopsis* underwent gene

differentiation, resulting in variations in the number of NAC genes within their respective subfamilies. Hence, the differences in the number of NAC genes among subfamilies reflect the evolutionary adaptations and environmental influences experienced by each species. The identification of these cis-acting elements elucidates the precise roles of these cis-acting elements and their interactions with transcription factors as well as miRNA in the regulation of NAC gene expression and their implications for the growth, development, and stress responses of the *Drimia* plant.

The analysis of the NAC gene family in the *Drimia* plant is currently in its preliminary stage, and further investigations are necessary to gain deeper insights into its functionality. Validating the findings will require essential techniques such as qPCR validation, gene cloning, and expression analyses. Uncovering the biological functions of NAC transcription factors in the *Drimia* plant will be a major focus of future research. This study of the gene family holds great potential for providing valuable theoretical resources to guide future gene cloning endeavours. Moreover, it will generate critical reference data regarding the regulation, structure, and function of these genes, thereby advancing our understanding of their roles in *Drimia* plant biology.

Conclusion

In this in-silico study, sixty-one NAC genes were identified. The study of different physicochemical parameters, including The isoelectric points, protein length, subcellular localization, and evolution, aligned with families in *Arabidopsis* with some observed variability. Phylogenetic analysis of identified 61 NAC genes from the *Drimia* genome, categorized into 17 subgroups (I–XVII). Genes within the same subgroup share similar gene structures and motifs, thus attributed to the same functions. However, such variation in the architecture of the genes may be attributed to the events of deletion, duplication, and splicing, eventually leading to diverse gene functions. Gene structure analysis found that they were likely composed of only exonic regions. The presence of cis-acting elements in promoter regions of NAC genes might be associated with a high expression upon exposure to heat, drought, and salinity stresses in response to ABRE and LR, MeJA. Moreover, identifying microRNA target sites for all 61 DiNAC genes provides targeted strategies to enhance plant defense and improve crop resilience against biotic stress. In conclusion, this study lays the groundwork for future research on the NAC gene family in medicinal plants, particularly *Drimia indica*. The findings provide valuable insights into the structural, functional, and evolutionary aspects of NAC genes, paving the way for further exploration of their molecular mechanisms, metabolic pathways, comparative genomics, and translational applications will significantly contribute to our understanding and utilization of these genes for the advancement of medicinal plant research and applications.

Abbreviation

NAC	NAM, ATAF1/2, and CUC2	BLAST	Basic Local Alignment Search Tool
CDS	Coding Sequence	ML	Maximum Likelihood
UTR	Untranslated Region	ANAC	*Arabidopsis* NAC
LR	Light Responsive	GATA	GATA Transcription Factor Family
ABRE	Abscisic Acid Responsive	Dof	DNA Binding with One Finger
MeJA	Methyl Jasmonate Responsive	HD-ZIP	Homeodomain-Leucine Zipper
MYB-LR	MYB Transcription Factor Responsive	TCP	Teosinte Branched/Cycloidea/PCF
TFBS	Transcription Factor Binding Site	MIKC_MADS	MIKC Type MADS-box
miRNA	MicroRNA	bHLH	Basic Helix-Loop-Helix
HMM	Hidden Markov Model	ERF	Ethylene-Responsive Element Binding Factor
RNA-seq	RNA Sequencing	RNA	Ribonucleic Acid
pI	Isoelectric Point	qPCR	Quantitative Polymerase Chain Reaction
NCBI	National Center for Biotechnology Information		

Declarations

Acknowledgment

The authors acknowledge the University Grants Commission (UGC) of India for generously financing this research through the fellowship (No. F. 82-44/2020 (SA-III)). We also acknowledge the assistance and resources provided by the Central University of Jharkhand, Ranchi.

Funding

This research was funded by the University Grants Commission (UGC), India, under fellowship No. F. 82-44/2020 (SA-III).

Competing Interests

The authors declare no competing interests.

Author Contributions

Vivek Shit and Manoj Kumar conceptualized the study and designed the experiments. Vivek Shit performed the laboratory experiments and collected the data and contributed to the data analysis and interpretation. Mahesh Kumar Dhakar provided critical feedback and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

References

1. Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. Nucleic acids research 43:W39–W49
2. Bar-Lev Y, Senden E, Pasmanik-Chor M, Sapir Y (2021) De novo transcriptome characterization of Iris atropurpurea (the Royal Iris) and phylogenetic analysis of MADS-box and R2R3-MYB gene families. Scientific Reports 11:16246. https://doi.org/10.1038/s41598-021-95085-5
3. Binny R, Saheb Cm, Saheb Chittoor M, et al (2012) Anthelmintic and antimicrobial studies of *Drimia indica* (Roxb.) Jessop. bulb aqueous extracts. Journal of Pharmacy Research 5:3677–3686
4. Chen C, Chen H, Zhang Y, et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular plant 13:1194–1202
5. Dai X, Zhuang Z, Zhao PX (2018) psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic acids research 46:W49–W54
6. De Bodt S, Raes J, Florquin K, et al (2003) Genomewide Structural Annotation and Evolutionary Analysis of the Type I MADS-Box Genes in Plants. Journal of Molecular Evolution 56:573–586. https://doi.org/10.1007/s00239-002-2426-x
7. De Siqueira Pinto M, Oliveira P, Riaño-Pachón DM, et al (2019) Analysis of NAC Domain Transcription Factor Genes of Tectona grandis L.f. Involved in Secondary Cell Wall Deposition. Genes 11:20. https://doi.org/10.3390/genes11010020
8. Gasteiger E, Hoogland C, Gattiker A, et al (2005) Protein identification and analysis tools on the ExPASy server. Springer
9. Gujjar RS, Akhtar M, Singh M (2014) Transcription factors in abiotic stress tolerance. Indian Journal of Plant Physiology 19:306–316
10. Hong JC (2016) General aspects of plant transcription factor families. In: Plant transcription factors. Elsevier, pp 35–56
11. Hu R, Qi G, Kong Y, et al (2010) Comprehensive Analysis of NAC Domain Transcription Factor Gene Family in Populus trichocarpa. BMC Plant Biology 10:145. https://doi.org/10.1186/1471-2229-10-145
12. Kumar R, Das S, Mishra M, et al (2021) Emerging roles of NAC transcription factor in medicinal plants: progress and prospects. 3 Biotech 11:425. https://doi.org/10.1007/s13205-021-02970-x
13. Kumar S, Stecher G, Li M, et al (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:. https://doi.org/10.1093/molbev/msy096

14. Lamesch P, Berardini TZ, Li D, et al (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research 40:D1202–D1210
15. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research 49:W293–W296
16. Lima JC de, Loss-Morais G, Margis R (2012) MicroRNAs play critical roles during plant development and in response to abiotic stresses. Genetics and molecular biology 35:1069–1077
17. Ling L, Li M, Chen N, et al (2023) Genome-Wide Identification of NAC Gene Family and Expression Analysis under Abiotic Stresses in Avena sativa. Genes 14:. https://doi.org/10.3390/genes14061186
18. Lv X, Lan S, Guy KM, et al (2016) Global Expressions Landscape of NAC Transcription Factor Family and Their Responses to Abiotic Stresses in Citrullus lanatus. Scientific Reports 6:30574. https://doi.org/10.1038/srep30574
19. Mistry J, Chuguransky S, Williams L, et al (2021) Pfam: The protein families database in 2021. Nucleic acids research 49:D412–D419
20. Mohanta TK, Yadav D, Khan A, et al (2020) Genomics, molecular and evolutionary perspective of NAC transcription factors. PLOS ONE 15:e0231425. https://doi.org/10.1371/journal.pone.0231425
21. Nicolas P, Lecourieux D, Kappel C, et al (2014) The Basic Leucine Zipper Transcription Factor ABSCISIC ACID RESPONSE ELEMENT-BINDING FACTOR2 Is an Important Transcriptional Regulator of Abscisic Acid-Dependent Grape Berry Ripening Processes. Plant Physiology 164:365–383. https://doi.org/10.1104/pp.113.231977
22. Puranik S, Sahu PP, Mandal SN, et al (2013) Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (Setaria italica L.). PloS one 8:e64594
23. Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC bioinformatics 8:1–10
24. Rombauts S, Déhais P, Van Montagu M, Rouzé P (1999) PlantCARE, a plant cis-acting regulatory element database. Nucleic acids research 27:295–296
25. Savojardo C, Martelli PL, Fariselli P, et al (2018) BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic acids research 46:W459–W466
26. Shen S, Zhang Q, Shi Y, et al (2020) Genome-Wide Analysis of the NAC Domain Transcription Factor Gene Family in Theobroma cacao. Genes 11:. https://doi.org/10.3390/genes11010035
27. Stanke M, Keller O, Gunduz I, et al (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids research 34:W435–W439
28. Sunkar R, Li Y-F, Jagadeeswaran G (2012) Functions of microRNAs in plant stress responses. Trends in plant science 17:196–203
29. Wang Y-X, Liu Z-W, Wu Z-J, et al (2016) Transcriptome-Wide Identification and Expression Analysis of the NAC Gene Family in Tea Plant [Camellia sinensis (L.) O. Kuntze]. PLOS ONE 11:e0166727. https://doi.org/10.1371/journal.pone.0166727

30. Wasternack C, Strnad M (2019) Jasmonates are signals in the biosynthesis of secondary metabolites—Pathways, transcription factors and applied aspects—A brief review. New biotechnology 48:1–11
31. Yan H, Ma G, Teixeira da Silva JA, et al (2021) Genome-Wide Identification and Analysis of NAC Transcription Factor Family in Two Diploid Wild Relatives of Cultivated Sweet Potato Uncovers Potential NAC Genes Related to Drought Tolerance. Frontiers in Genetics 12:. https://doi.org/10.3389/fgene.2021.744220
32. Zhang H, Ma F, Wang X, et al (2020) Molecular and Functional Characterization of CaNAC035, an NAC Transcription Factor From Pepper (Capsicum annuum L.). Frontiers in Plant Science 11:. https://doi.org/10.3389/fpls.2020.00014
33. Zhu C, Ding Y, Liu H (2011) MiR398 and plant stress responses. Physiologia plantarum 143:1–9

Figures

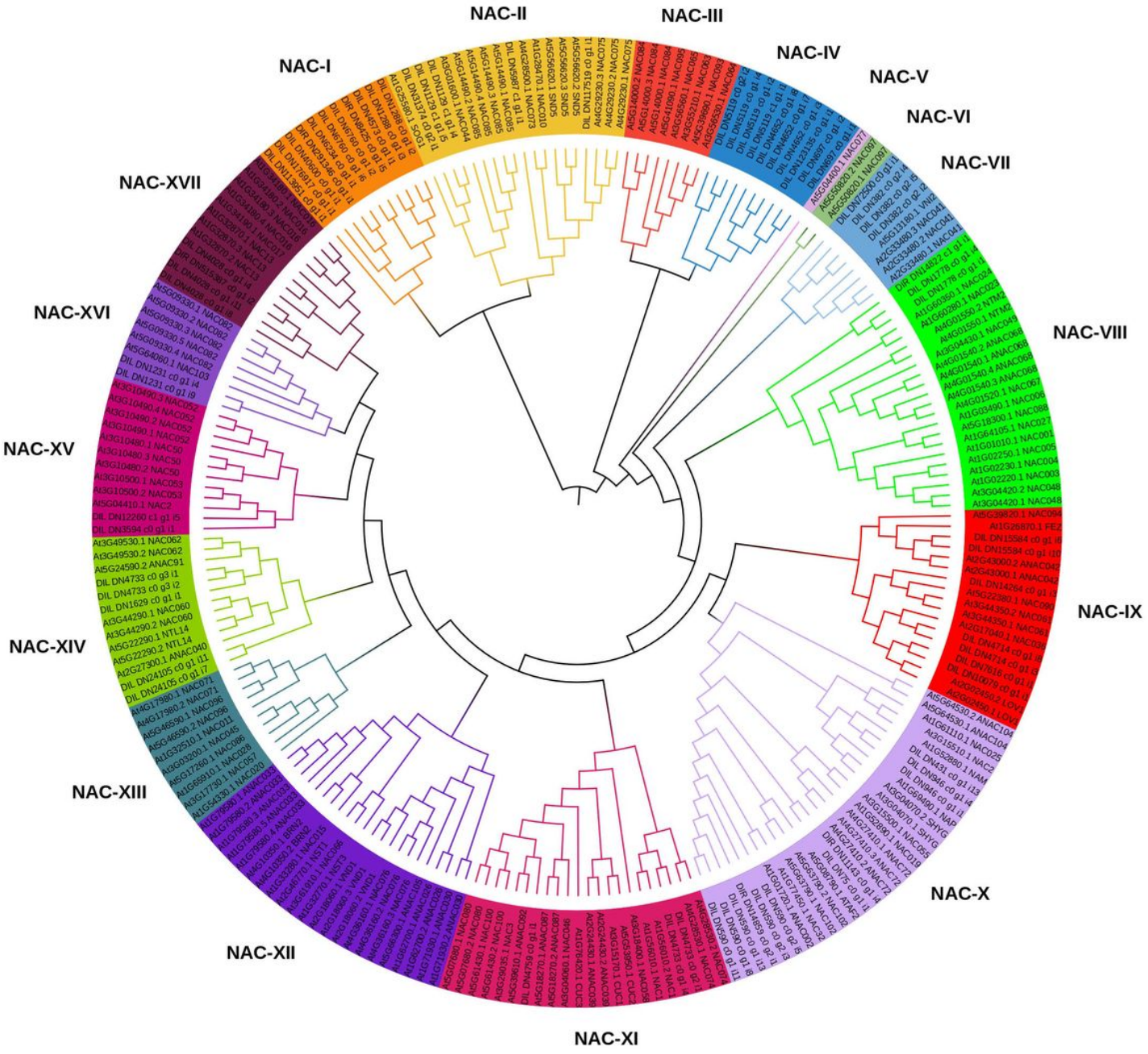

Figure 1

Illustrates the phylogenetic tree depicting the relationship between NAC proteins from *A. thaliana* and *D. indica*. The phylogenetic tree was generated using MEGA 11.0 through the maximum likelihood (ML) method with 1000 bootstrap replicates. The NAC proteins were classified into 17 distinct subgroups. "DiNACs (DIL & DIR)" refers to the NAC proteins derived from *D. indica*, while "ANACs" denotes the NAC proteins from *A. thaliana*.

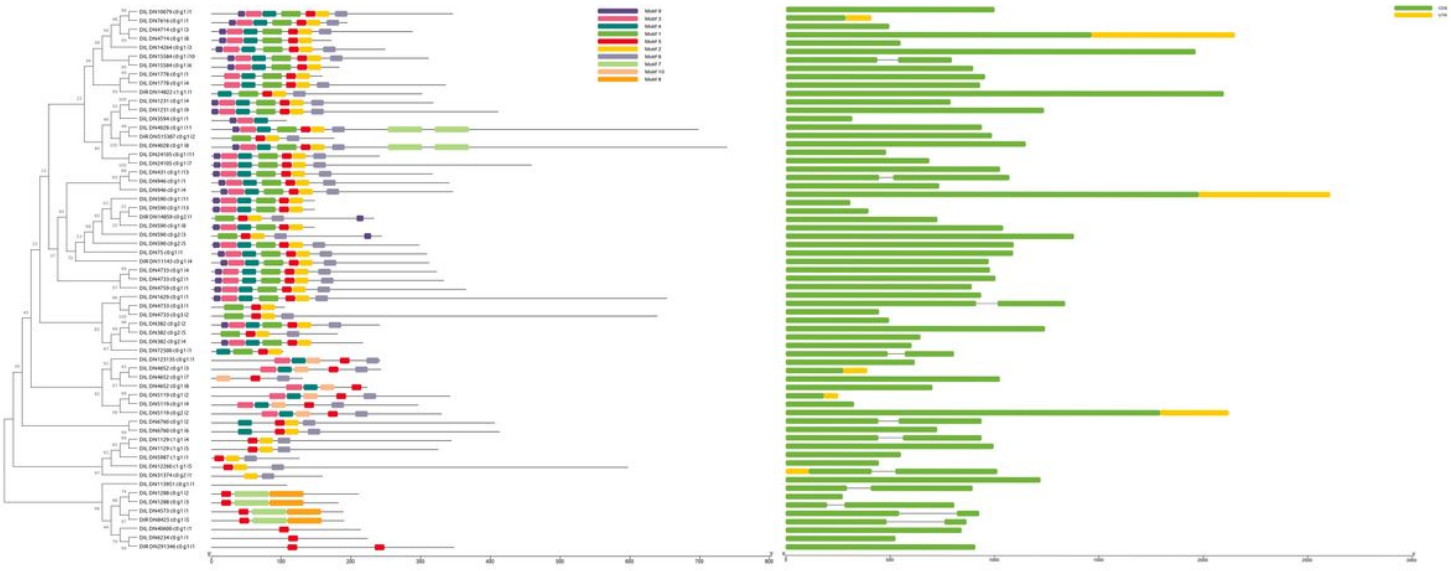

Figure 2

Depicts the structural Diversity of NAC Proteins in the *Drimia* Plant. (A) A phylogenetic tree was constructed using MEGA 11.0, illustrating the relationships among DiNAC proteins. (B) Visualization of the motif composition and relative motif lengths within the NAC protein, with distinct coloured boxes to represent each motif. (C) and the organization of exons and introns within the NAC gene.

Figure 3

Identification of cis-responsive elements for DiNAC genes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1.txt
- SupplementaryFile2.xlsx
- SupplementaryFile3.xlsx

- SupplementaryFile4.xlsx
- SupplementaryFile5.xlsx
- SupplementaryFile6.txt
- SupplementaryFile7.xlsx
- SupplementaryFile8.xlsx