Plant Archives Volume 20 No. 2, 2020 pp. 3551-3673 **e-ISSN:2581-6063 (online), ISSN:0972-5210**

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF NAC TRANSCRIPTION FACTOR GENES IN *ARTEMISIA ANNUA* USING COMPUTATIONAL APPROACHES

Nidhi Rai, Lakee Sharma, Bipin Maurya, Krishna Kumar Rai, Ram Prasad Meena and Shashi Pandey Rai*

Laboratory of Morphogenesis, Centre of Advance Study in Botany, Department of Botany, Faculty of Science, Banaras Hindu University (BHU), Varanasi (Uttar Pradesh), India.

Abstract

The Transcription factors act as key regulatory switches of transcription networks and gene expressivity level. The plant specific NAC transcription factor is derived from no apical meristem (NAM), Arabidopsis transcription activation factor (ATAF) and cup-shaped cotyledon (CUC) superfamily. It has crucial regulatory roles against various abiotic stresses and hormone signalling. In *Artemisia annua* 28 NAC genes were retrieved from PlnTFDB; Plant Transcription Factor Database and protein sequences were obtained by using ExPaSy server and analyzed by various *in- silico* applications. Homology search of these proteins was done by employing The NCBI BLAST tool which revealed the higher similarity percentage with homologous; mainly of members of the Asteraceae family. Sequences were further studied for phylogenetic analysis and motif identification. Results of the NCBI CDD server and clustal W suggested that in majority of *Aa*NAC proteins, presence of NAM domains was confirmed except in *Aa*NAC12 and *Aa*NAC27. In some proteins, harboring signature NAM domain, gap sequences were observed between N- to C- terminus which were also conserved in nature. Multiple sequence alignment of proteins was used for phylogenetic analysis. At 1000 bootstrapping the 28 members were grouped in 2 clusters and 7 groups. *Cis*-regulatory domains were analyzed by Plant Care and PLACE tools, from where the presence of TATA, CAAT, TCC boxes and binding sites for MYB, MYC, ARE and other transcriptional regulators were observed. A multilevel consensus sequence of conserved motifs from all the NAC protein members of *A. annua* was retrieved by employing MEME suit. Out of 10, 1st and 2nd conserved motifs were found to cover the NAM domain. For the analysis of physical characters, the NAC members were studied for molecular weight, acidic- basic, polar- nonpolar behaviour and localization. All members share some common features like non- polar behaviour, nuclear localization and absence of any transmembrane domain.

Key words: Transcription factor, NAC, NAM, *in- silico*, abiotic stresses, multiple sequence alignment

Introduction

The transcription factors are modulatory protein containing two domains, one is DBD (DNA binding domain) and the second one is TRD (transcription regulatory domain). TFs can facilitate plant growth, development and primary metabolism by regulating the expressions of downstream genes. Several TFs like WRKY, MYB, NAC and AP2/ERF act as major TFs that play indespensible role in plants to cope up abiotic and biotic stress. Among these, NAC is the largest plant-specific TF family reported in various plant species like 28 in *Artemisia annua*, 251 in switchgrass, 204 in Chinese cabbage, 152 in soybean, 151 in rice, 110 in S*olanum tuberosum*, 147 in foxtail millet, 117 in *Arabidopsis*, 104 in tomato, 82 in melon and 74 in grape (Hu *et al*., 2010; Nuruzzaman *et al*., 2010; Puranik *et al*., 2013; Singh *et al*., 2013; Wang *et al*., 2013a; Liu *et al*., 2014; Wei *et al*., 2016).

The NAC transcription factor was firstly identified in *Petunia for pattern formation in embryo and flower* (Souer *et al*., 1996). Aida *et al*., 1997, discovered regulatory role in *Arabidopsis thaliana* as a chief TF for and organ separation. Later it was found that these proteins are also involved in the regulation of various plant developmental processes, like formation of

***Author for correspondence :** E-mail : shashi.bhubotany@gmail.com

adventitious shoots (Hibara *et al*., 2003), progression of the shoot apical meristem (Souer *et al*., 1996; Aida *et al*., 1997), flower development (Sablowski *et al*., 1998), leaf senescence (Lee *et al*., 2012), lateral root formation (Xie *et al*., 2002), floral morphogenesis (Aida *et al*., 1997), cell cycle control (Kim *et al*., 2006; Wilensen *et al*., 2008), hormone signalling (Xie *et al*., 2002; Kim *et al*., 2006; Fujita *et al*., 2004), ubiquitin-dependent proteolysis and grain nutrient remobilization (Uauy *et al*., 2006).

The name NAC was originally taken from three proteins that harbor similar DNA binding domain, first one NAM (no apical meristem) second one ATAF1/2 (Arabidopsis transcription activation factor) and the last one is CUC2 (cup-shaped cotyledon). The structure of NAC proteins consists of two regions, in which N-terminus is highly conserved and involve in DNA binding while C- terminal being highly diverse and involved in regulation of other TF. The DNA- binding domain consists of 150-160 amino acid residues containing five sub-domains (A- E) that play important role in DNA binding, nuclear localization and formation of homo and heterodimers. Out of five, 3-subdomains A, C and D are highly conserved, whereas B and E are highly divergent and might be responsible for functional diversity of NAC transcription factor (Ooka *et al*., 2003; Puranik *et al*., 2012). The transcription regulatory regions (TRs), activate/repress transcription and responsible for diverse function of NAC proteins. The transcriptional activities are mainly conducted through quietly diverse C-terminal region. Mainly, NAC proteins are known to reside in nucleus, lacking any transmembrane (TM) domain. Contrary to this, due to variable C- terminus some of the NAC TF are known to contain TM motif shows subcellular localization in the plasma membrane and endoplasmic reticulum.

In several species such as *Arabidopsis, Solanum lycopersicum*, *Glycine max* and *Oriza sativa,* membrane bound NAC TFs are known to induce in response to abiotic stress. However, the exact regulatory mechanism of NAC genes is not much known yet. The activity of the NAC gene can be regulated through various processes: at the transcriptional level binding of specific TFs to NAC regulatory region at the promoter region and degradation of NAC protein mediated by ubiquitins, dimerization and interaction with other proteins, at post-transcriptional level cleavage of NAC gene mediated by miRNA 164. In recent years, the NAC gene family has gained much attention due to its involvement in numerous fundamental biological processes including, hormone signalling in response to biotic and abiotic stresses (Puranik *et al*., 2012), cell expansion, cell differentiation and tolerance to environment cues (Yang *et al*., 2011).

In *Arabidopsis,* multi-abiotic stress tolerance have been achieved by regulating the expression of stress-responsive genes *AtNAC2, AtNAC3, ANAC019 a*nd *RD26* (Tran *et al*., 2004; He *et al*., 2005; Balazadeh *et al*., 2010; Lockhart *et al*., 2013). Also, many of the NAC proteins have been linked with plant defense against various stress conditions, for example, drought (Jeong *et al*., 2010), salinity (Zheng *et al*., 2009) and heavy metals (Wang *et al*., 2013b). Evidently, NAC family genes are critical regulator implicated in the survival of plants under various stress conditions. In *Arabidopsis,* during various stress conditions like ABA, salt and drought the expression level of three NAC genes (*ANAC072, ANAC055* and *ANAC019*) were found to be elevated significantly. Furthermore, overexpression of these genes was advocated to increase abiotic stress tolerance as compared to the wild-type (Hu *et al*., 2006; Jensen *et al*., 2010). Similarly, in transgenic *Arabidopsis, ATAF1* showed enhanced tolerance to dehydration, ABA, salinity and oxidative stress. In plants, some major TFs like NAC, N-terminal WRKY domain, basic helix-loop-helix (bHLH), myeloblastosis viral oncogene homolog (MYB), basic leucine zipper (bZIP) and APETALA2/ethylene-responsive element binding factor (AP2/ERF) are well defined in biotic and abiotic responses. In addition to these, TFs like, Zinc finger protein, C_2H_2 and MADS box are also having well documented role in plant immunity.

A. annua L. is a member of family Asteraceae (Brown, 2010) and is commonly known as sweet wormwood, Chinese wormwood. This genus comes under small herbs and shrubs and can produce various natural products, including artemisinin and its derivatives. Artemisinin has been extracted from mature leaves or buds of *A. annua*. Artemisinin is an endoperoxide sesquiterpene lactone that is known to be efficient against multidrug-resistance malaria and is also known to act on *Plasmodium falciparum*, causal organism of cerebral malaria. Bbesides to malaria, artemisinin can cure other diseases including cancers, parasitic diseases like schistosomiasis and viral disease such as hepatitis B.

In *A. anuua* there are 28 NAC TFs have been reported (PlantTFDB v5.0). These NAC TF regulate the different stress induced signalling pathway in the plant. Among these expression of *AaNAC1* could be induced by dehydration, cold, salicylic acid (SA) and methyl jasmonate (MJ). *AaNAC1* was shown to be localized to the nuclei by transforming tobacco leaf epidermal cells. By the overexpression of *Aa*NAC1 in *A. annua*, the content of artemisinin and dihydroartemisinic acid

increased by 79 and 150%, respectively. The expression levels of artemisinin biosynyhetic pathway gene, i.e. amorpha-4, 11- diene synthase (ADS), artemisinic aldehyde delta11 (13) reductase (DBR2) and aldehyde dehydrogenase 1 (ALDH1), were increased. Therefore, NAC transcription factors can modulate artemisinin biosynthesis when *AaNAC1* was overexpressed in the *A. annua* (Lv *et al*., 2016) and its functional analysis can answer its proper role as a regulatory element in *A. annua* against abiotic stresses along with its association with the secondary metabolism of the plant.

Materials and Methods

Database search for identification of NAC family members in *Artemisia annua*-

By employing the Plant Transcription Factor Database http://planttfdb.cbi.pku.edu.cn/ for the identification of NAC genes (NAM, ATAF and CUC) in *Artemisia annua*, a total of 28 NAC genes were obtained. The complete domain of each sequence was retrieved from plant GDB (http://www.plantgdb.org/) which provided species based sequences from various sources like NCBI GenBank https://www.ncbi.nlm.nih.gov/genbank/ and UniProt https://www.uniprot.org/. Coding sequences (CDS)/ open reading frame of each member of NAC gene was obtained by using ORFfinder https://www.ncbi.nlm.nih.gov/orffinder/ of NCBI. Protein sequences of each NAC gene was translated and retrieved by applying ExPASy tool of NCBI https://web.expasy.org/translate/.

NAC sequence alignment and phylogenetic analyses-

After retrieving the protein sequences, NAC genes of *A. annua* were subjected to homology search and phylogenetic tree construction to obtain suitable homologues sequences from other members of the Asteraceae family. For this, first of all sequences were searched against homologous and orthologuos sequences of close relatedness with the help of the BLASTp tool https://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul *et al*., 1990). All the 28 protein sequences were aligned to each other by employing multiple sequence alignment (MSA) tool, T- Coffee (Tree-based Consistency Objective Function for Alignment Evaluation) https://www.ebi.ac.uk/Tools/msa/tcoffee/ by selecting "BLOSUM" and "input" parameters. This program combines the results of many multiple alignment programs like Muscle https://www.ebi.ac.uk/Tools/msa/muscle/, ClustalW https://www.genome.jp/tools-bin/clustalw, Mafft (multiple alignments using fast Fourier transform) https://www.ebi.ac.uk/Tools/msa/mafft/ and ProbCons (Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences) http://probcons.stanford.edu/ (Larkin *et. al*., 2007). Further, the phylogenetic tree was constructed by the Mega 6.0 tool https://www.megasoftware.net/, using UPGMA (unweighted pair group method with arithmetic mean) methods with 1,000 replication bootstrap values (Tamura *et al*., 2007). Similarities and differences based on sequential analyses of proteins were done by Circos visualization tool http://circos.ca/ which was based on the similarity percentage index, derived from Clustal Omega software (Krzywinski *et al*., 2009).

Identification of conserved domain, gene families and motifs-

Each NAC protein was scanned throughout its length for the analysis of gene families as well as domains present. This was performed with the help of ExPASy-Prosite scan http://prosite.expasy.org/scanprosite/ (DeCastro *et al*., 2006), InterPro tools https://www.ebi.ac.uk/interpro/search/sequence/, Pfam https://pfam.xfam.org/(Mistry *et al*., 2007). Functional and conserved motifs were identifying by submitting each sequence to motif finder tools such as CDD https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml and SMART http://smart.embl-heidelberg.de/ (Schultz *et. al*., 2000). *Cis*- acting regulating DNA elements in the all 28 NAC genes were detected by submitting the nucleotide sequences of each member to the online promoter predicting servers such as PlantCARE http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ and the PLACE (Plant *cis*- acting regulatory DNA elements) http://www.dna.affrc.go.jp/PLACE/ (Lescot, 2002 and Higo *et al*., 1999). All the *Aa*NAC protein sequences of *Artemisia annua* were further submitted together in MEME (Multiple EM for Motif Elicitation) software version 4.8.0 http://meme-suite.org/ to find out the consensus sequences of the conserved motifs (Bailey and Michael 1998; Bailey *et al*. 2006). For this, selection criteria were set as, maximum number of motifs to 10 with minimum and maximum width of 10 and 30 respectively, while the rest of the parameters were set to default.

Analyses of physical characters of NAC proteins-

Various chemical and physical properties of 28 members of NAC protein from *A. annua* were assessed by various online *in- silico* tools. Theoretical isoelectric point (pI) and molecular weight of the proteins were computed by Compute pI/Mw tool of ExPASy, https://web.expasy.org/compute_pi/ (Gasteiger *et al*., 2005). Hydropathicity values of all the proteins were calculated

by employing GRAVY (grand average of hydropathy) module of Sequence Manipulation Suite (SMS), https://www.bioinformatics.org/sms2/protein_gravy.html (Stothard, 2000). To depict the subcellular localization(s) ofNAC proteins, the protein sequences were submitted to Plant-mPloc server (Predicting subcellular localization of plant proteins including those with multiple sites), http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/ (Chou and Shen, 2010). The presence of transmembrane helices in NAC proteins, if any, was observed by employing TMHMM Server v. 2.0, http://www.cbs.dtu.dk/services/TMHMM-2.0/ (Krogh *et al*., 2001).

Results

NAC family members of *Artemisia annua*-

All the 28 NAC genes were extracted from the Plant Transcription Factor database (PlnTFDB), then translated by ExPASy server and further characterized as well as analyzed by various *in- silico* analyzing tools. Except a few, all the 28 NAC genes were found to contain full coding sequences that were started from methionine and ended to stop codon. We have analyzed the length of coding sequences of identified NAC genes in *A. annua* to be 186- 1403 nucleotides with 801.3 average lengths which encode proteins in a range of 61- 430 amino acid residues with an average 306 amino acids table 1. ORF in the plus direction, *i.e.* 5'-3'only was considered.

Table 1: The gene ID of the NAC transcription factor genes of *Artemisia annua*, retrieved from Plant Transcription Factor database, http://planttfdb.cbi.pku.edu.cn/.

S. No.	Accession no./gene ID	Gene name	Total length of nucleotide	No. of ORF (s)in 5'-3' direction	Length of longest ORF(nt)	Length of protein sequence
1	Aan000926	*Aa*NAC1	918	4	918	305
2	Aan001843	*Aa*NAC2	1185	3	1185	394
3	Aan002468	*Aa*NAC3	1008	3	1008	335
4	Aan006201	*Aa*NAC4	909	4	909	302
5	Aan006361	*Aa*NAC5	873	3	873	290
6	Aan008548	*Aa*NAC6	645	4	645	214
7	Aan009721	*Aa*NAC7	539	1	539	179
8	Aan010235	*Aa*NAC8	498	3	498	165
9	Aan010386	*Aa*NAC9	483	2	366	121
10	Aan010878	*Aa*NAC10	438	3	438	145
11	Aan012884	*Aa*NAC11	210	1	210	69
12	Aan012975	*Aa*NAC12	192	1	192	63
13	Aan013239	*Aa*NAC13	236	1	236	79
14	Aan013749	*Aa*NAC14	1289	1	1289	430
15	Aan014755	*Aa*NAC15	789	2	789	262
16	Aan015209	*Aa*NAC16	607	1	607	203
17	Aan015985	*Aa*NAC17	651	1	651	216
18	Aan016410	*Aa*NAC18	962	1	962	320
19	Aan017655	*Aa*NAC19	998	1	996	331
20	Aan018369	*Aa*NAC20	1005	5	186	61
21	Aan018435	*Aa*NAC21	1251	4	1251	416
22	Aan018568	*Aa*NAC22	459	2	459	152
23	Aan019479	*Aa*NAC23	753	3	753	250
24	Aan019671	*Aa*NAC24	943	1	853	284
25	Aan020678	*Aa*NAC25	816	3	792	263
26	Aan020942	*Aa*NAC26	1403	1	999	330
27	Aan021211	*Aa*NAC27	1267	3	670	226
28	Aan021242	*Aa*NAC28	1113	5	1113	370

Characterization of the *Aa*NAC protein by basic local alignment search tool (BLAST)

The protein sequences were further analyzed by the BLASTp program to observe the position and sequence similarity of NAM domains to other members of Asteraceae family. In this context, the only homologues were selected for further study, which show more than 70% similarity and less than 8% gap against *A. annua* NAC proteins table 2. 9 NAC protein sequences from *A. annua*, out of 28 showed highest resemblances with *Helianthus annuus* (Accessions No.- XM_022116662.1, XM_022137133.1, XM_022141545.1, XM_022117894.1, XM_022121205.1, XM_022184100.1, XM_022141545.1 and XM_022161322.1) while 5 with *Cynara cardunculus* (Accessions No.- XM_025122615.1, XM_025131726.1, XM_025114967.1, XM_025138538.1 and XM_025135904.1) and 4 with *Lactuca sativa* (Accessions No.- XM_023906937.1, XM_023894074.1, XM_023900271.1 and XM_023901052.1) and rest with other members of Asteraceae family (*Chrysanthemum morifolium a*nd *Achillea asiatica*. While orthologues from members of other families besides Asteraceae (*Cannabis sativa, Cicer arietinum, Cucumis sativus a*nd *Jatropha curcas*) have also shared sequence similarity with some members of *Aa*NAC proteins. Overall, the alignment was conserved, as the similarity of query and subject proteins were found to be maximum near the regions occupied by NAM domains of the NAC protein. On the other hand, two proteins (*Aa*NAC12 and *Aa*NAC21) showed no identity with any homologue which may be due to lack of full sequence of the NAC gene.

Multiple sequence alignment and

phylogenetic analysis of *Aa*NAC trancription factors

Mostly, all the sequences were found to harbour complete coding regions with presence of single NAM domain (except *Aa*NAC12 and *Aa*NAC27; as they don't contain any signature domain). With the help of motif finder tools like SMART, CDD server of NCBI, NAM domain of each NAC protein was extracted. In some protein sequences, partial NAM domain was also observed. An average alignment score was obtained Fig. 1. In our study, 26- 137 amino acid long NAM domains from 26 *A. annua* NAC proteins were attested. Further, in some sequences presence of conserved gap regions (between N- terminus to C- terminus of NAM domain) were observed which range from 35- 53 amino acid residues in length. In *Aa*NAC9, *Aa*NAC26 and *Aa*NAC27 only residual sequences on C-terminus end of the signature motif was present.

As shown above, to find out the phylogenetic relatedness amongst different NAC genes of *Artemisia annua* (*Aa*NAC1- *Aa*NAC28), all the deduced sequences were subjected to multiple sequence alignment (including *Aa*NAC12 and *Aa*NAC27 also). Phylogenetic tree constructed by employing Mega 6.0 with 1000 bootstraping revealed two main clusters which are designated as Cluster "A" and "B", containing 26 and 2 *Aa*NAC proteins, respectively Fig. 2. Major cluster "A" was again splited into 2 sub-clusters *viz*, A (i) and A (ii) harbouring 16 and 10 *Aa*NAC proteins, respectively. Overall, whole tree was divided into 7 subgroups, in which 1- 6th group was devoted to major cluster "A" and 7th to cluster "B". Further, major cluster "B" was not divided into any sub- cluster as it formed a distinct clad having only 2 *Aa*NAC proteins namely *Aa*NAC7 and *Aa*NAC25.

The relatedness and dissimilarities among different members of *Aa*NAC proteins was also assessed by employing circos visualization tool. In this study, significantly shorter sequences like *Aa*NAC9, *Aa*NAC10, *Aa*NAC11, *Aa*NAC12, *Aa*NAC13 and *Aa*NAC22 were automatically neglected by the default parameters of circos algorithm Fig. 3.

Table 2: Output of 28 NAC proteins of *Artemisia annua* by BLAST analysis.

S. No.	Gene name	Similarity with plant species	Accession No.	Similarity %	E-Value	GAP
1	*Aa*NAC1	*Helianthus annuus*	XM_022116662.1	77.12%	1e-125	6%
2	*Aa*NAC2	*Lactuca sativa*	XM_023906937.1	85.11%	0.0	0%
3	*Aa*NAC3	*Helianthus annuus*	XM_022137133.1	81.50%	0.0	2%
4	*Aa*NAC4	*Lactuca sativa*	XM_023894074.1	83.60%	1e-179	0%
5	*Aa*NAC5	*Lactuca sativa*	XM_023900271.1	82.07%	3e-111	2%
6	*Aa*NAC6	*Cannabis sativa*	XM_030641236.1	73.52%	5e-28	2%
7	*Aa*NAC7	*Cynara cardunculus*	XM_025122615.1	81.84%	8e-100	0%
8	*Aa*NAC8	*Helianthus annuus*	XM_022141545.1	85.92%	2e-121	0%
9	*Aa*NAC9	*Helianthus annuus*	XM_022117894.1	78.82%	5e-67	1%
10	*Aa*NAC10	*Lactuca sativa*	XM_023901052.1	83.40%	3e-59	0%
11	*Aa*NAC11	*Achillea asiatica*	KX722453.1	96.02%	8e-86	0%
12	*Aa*NAC12	-	-	-	-	-
13	*Aa*NAC13	*Achillea asiatica*	KX722453.1	96.20%	6e-78	0%
14	*Aa*NAC14	*Helianthus annuus*	XM_022121205.1	76.01%	5e-170	4%
15	*Aa*NAC15	*Cynaracardunculus*	XM_025131726.1	88.53%	2e-88	0%
16	*Aa*NAC16	*Cucumis sativus*	XM_004150332.2	71.19%	2e-21	0%
17	*Aa*NAC17	*Cynaracardunculus*	XM_025114967.1	82.92%	1e-148	0%
18	*Aa*NAC18	*Artemisia annua*	KX082975.1	98.84%	0.0	0%
19	*Aa*NAC19	*Helianthus annuus*	XM_022184100.1	83.93%	2e-114	0%
20	*Aa*NAC20	*Helianthus annuus*	XM_022184100.1	83.39%	2e-139	1%
21	*Aa*NAC21	-	-	-	-	-
22	*Aa*NAC22	*Jatrophacurcas*	XM_012223876.2	90.57%	1e-07	1%
23	*Aa*NAC23	*Helianthus annuus*	XM_022141545.1	81.42%	5e-168	2%
24	*Aa*NAC24	*Chrysanthemummorifolium*	HQ317452.1	97.22%	0.0	0%
25	*Aa*NAC25	*Cicerarietinum*	XM_004493081.3	78.46%	3e-36	2%
26	*Aa*NAC26	*Cynaracardunculus*	XM_025138538.1	83.40%	0.0	2%
27	*Aa*NAC27	*Cynaracardunculus*	XM_025135904.1	80.04%	2e-104	1%
28	*Aa*NAC28	*Helianthus annuus*	XM_022161322.1	75.96%	3e-52	8%

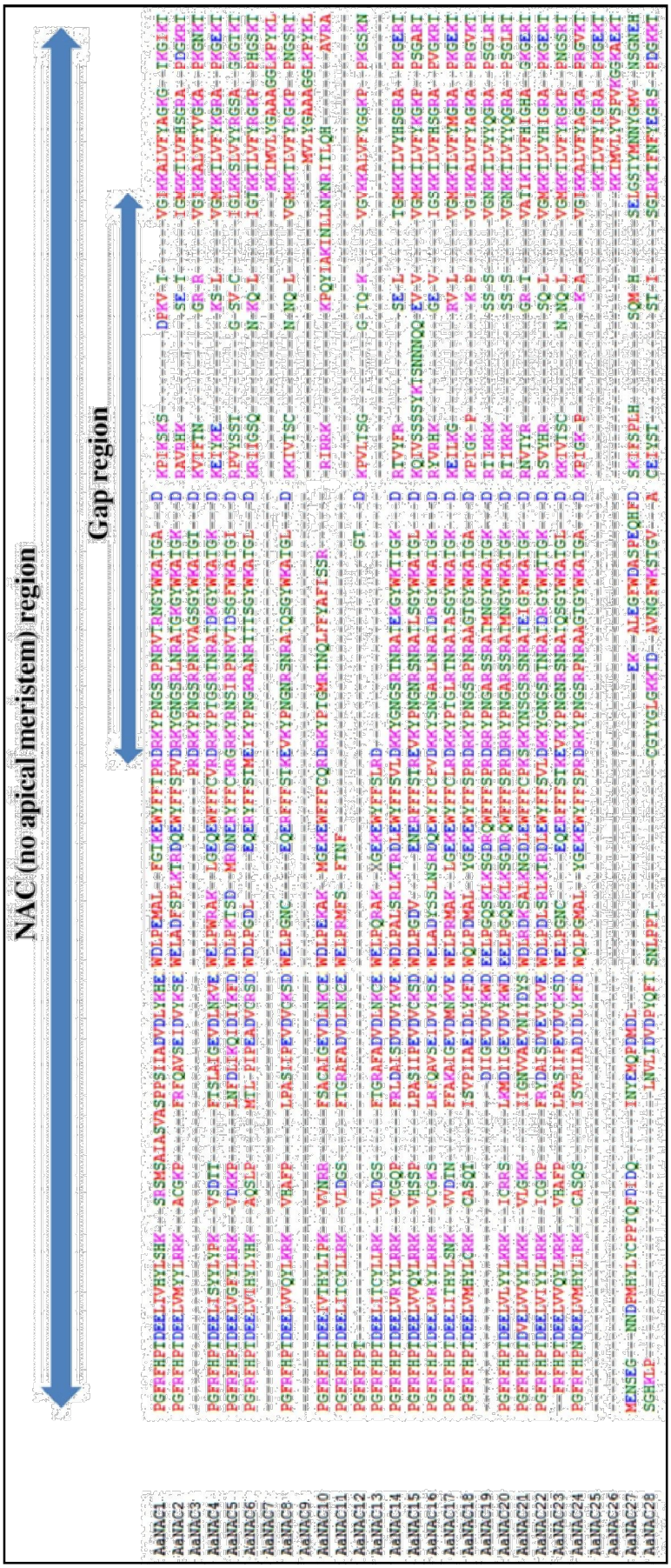

Fig. 1: Multiple sequence alignment (MSA) of different *Artemisia annua* NAC proteins.

Distribution analysis of regulatory *cis*-elements in members of *Aa*NAC genes in *Artemisia annua*

Analysis of distribution patterns for various *cis*- regulatory DNA binding elements in *Aa*NAC proteins was studied by employing various online applications like the PLACE and Plant Care. Majority of the *cis*- elements were found to be located on the C- terminus of NAC members with an average length of 6-8 nucleotides. Since pertaining to this study, only plus sequences were taken in consideration, thus the elements were not omnipresent in all 28 members. Only, CAAT regulatory box was present in all the 28 NAC genes (except *Aa*NAC2) Fig. 4.

Besides, CAAT other regulatory conserved box were found across the *A. annua* NAC genes such as CAT, TATA, TAC and TCA, etc. Proceeding in this way, presence of other transcription factor binding conserved motifs was attested such as G (CACGTG), W [(T)TGAC(C/T)], P- and H (DExD/ H) box which are known to involve in plant development as well as in metabolism and stress responses. In *Aa*NAC proteins, we have enlisted Transcription recognition and binding sites like MYB, MYC, ARE and others (not shown in table) at C- terminal end which reveals the interdisciplinary role of NAC protein in plants. The conserved motifs like WUN (wound responsive element), GATA and TCC were also present as regulatory proteins in some of the NAC proteins of *A. annua*.

Sequential analysis of conserved motis in *Aa*NAC protein of *Artemisia annua*

Motif distribution depicted by the MEME tool is shown in Fig. 5, which informs the location wise arrangement of some of the conserved domains found in NAC proteins. Out of 28 *Aa*NAC proteins, these domains were significantly found only in 19 members (absent in *Aa*NAC7, *Aa*NAC9, *Aa*NAC10, *Aa*NAC11, *Aa*NAC23, *Aa*NAC24, *Aa*NAC26, *Aa*NAC27 and *Aa*NAC28).

As we requested only 10 motifs, out of which motif 1st and 2nd were found to harbor the NAM domain and present only in 17

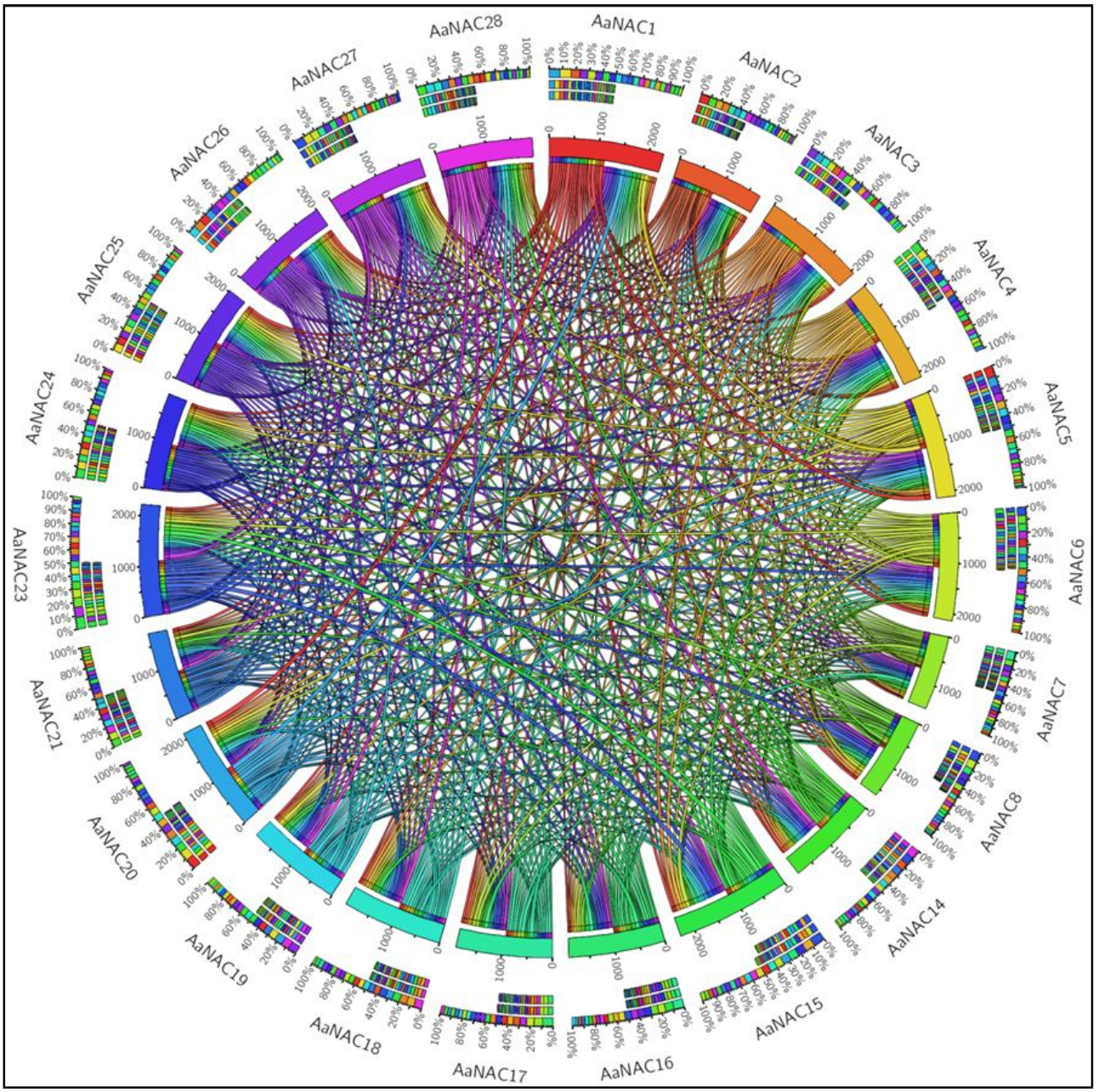

Fig. 3: Circos diagram to show Comparative analysis of similarities and differences in *Artemisia annua* NAC protein (AaNAC) among different members of *Artemisia annua*.

NAC proteins. In some *Aa*NAC proteins, such as *Aa*NAC5, *Aa*NAC6, *Aa*NAC7, *Aa*NAC13 and *Aa*NAC21 there is gap between 1st and 2nd motif which denotes the addition of extra amino acid sequences during the course of evolution. Significant variations were observed in the conserved motifs, located at the C-terminus of of the proteins.

Multilevel consensus sequence for theseconserved motifs were further analyzed to study some parameters like motif length and E- Value table 3. Width of all the 10 requested motifs were limited to 18 to 50 amino acid residue with an average of 33.3 residues. Among these, motif no 5 and 6 were found to conatin single level of consensus sequences i.e. there was no overlapping and differences in the motif, all over its length.

Characterization of physical properties of *Aa*NAC proteins

Next, physical parameters of 28 *Aa*NAC proteins of *A. annua* were studied so far to analyze the characters such as molecular weight, acidic and/or alkaline behaviour, hydrophobic and/or hydrophilic nature, presence of transmembrane domains (if any) and location

Table 3: Consensus sequence of conserved motifs identification in NAC proteins of *Artemisia annua* by multilevel motif based sequence analysis tool.

S. No.	Width	Multilevel consensus sequence	E-value
1.	37	[GR][ED][QK]E[WR][YF]FF[SC][PT]R[DE][RK]KYPNGSR[STP]NRATDSGYWKATG[KL]D[KR]	3.3e-347
2.	36	[IV][SY]XKSQL[VI]GMKKTLV[FY][YH]RG[RK][AP]PKGE[RK]T[DN]W[VI]MHEYRL	1.8e-304
3.	24	[SN][AP][GNP][QSV]A[QL][DE][DN][WY]V[LI]C[RK][IV][FY][KHL]K[KSN]G[SG]G[PK]KN	1.8e-064
4.	50	QKLQPIPVNDFKSAREEEAPLSNDDIEFLHRIIEPEVLPPLPQITEKAKM	5.1e-018
5.	50	MNDLLGPEPEKCSFDSLSEFDFCQEADAFGSQPFELQPAYMDNLEFGPAN	3.3e-017
6.	21	G[EA][QK]YGAP[FL][VIN]EEEW[SDE][DE][DE][DE][CIS][LAI][DFS][VC]	2.2e-015
7.	48	EHEREVQS[DE][PV]KK[DN][DE]FQFNY[IM]D[PS]F[AP]DDAFTPQ[NS]QYYND[FY]QLSPLQDMFM	5.2e-015
8.	30	LHPP[PS]S[IV]PHHVM[DN]D[GV]F[HN]F[DE][ST]SESVPTL[HQ]TD	1.2e-004
9.	19	FY[AV]GK[AG][PT]KG[IN]KTNWIMHEY	1.7e-003
10.	18	QISEME[IV]ETKPKIT[GP]Y[AT]P	1.8e-001

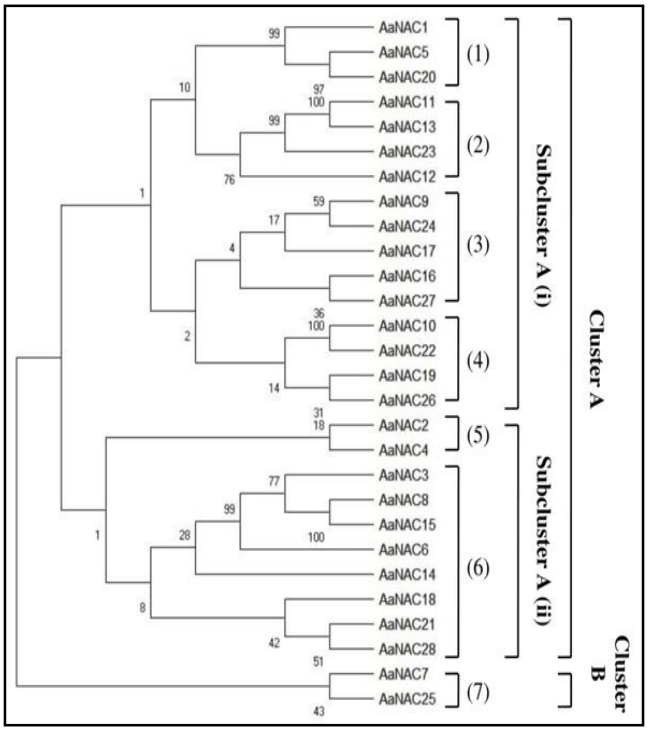

Fig. 2: Clustring of *Artemisia annua* NAC proteins by complete alignment and construction of Phylogenetic tree choosing 1000 bootstrap value.

of these proteins inside the cell table 4. In this context, molecular weight of proteins was attested to vary significantly between 28 members of NAC proteins and range from 6635.58 to 48659.84D with an average value of 28611. 43D. on the basis of pI values, the ratio of acidic and basic nature of NAC proteins were found to be 50- 50% each. Contrary to this, 2 members *i.e.* *Aa*NAC9 (pI value- 6.28) and *Aa*NAC24 (pI value- 6.72) were seems to be neutral as their Isoelectric point was near 7.0. Polarity of proteins was estimated by calculating the hydropathicity values of each member. Results have revealed that all the 28 members were non- polar in nature as they have negative values. Lastly, Subcellular localization of the proteins were done by *in- silico* analysis which confers the occurrences of proteins in nuclear regions, inside the cell.

Table 4: Comparative analysis of basic physical properties of NAC proteins in *Artemisia annua*.

S. No.	Gene name	Molecular weight	Isoelectric point (pI)	Hydropathicity value	Transmembrane domain	Subcellular localization
1.	*Aa*NAC1	34892.03	5.50	-0.812	No	Nucleus
2.	*Aa*NAC2	44304.76	5.40	-0.726	No	Nucleus
3.	*Aa*NAC3	37947.63	8.90	-0.726	No	Nucleus
4.	*Aa*NAC4	34858.05	5.15	-0.797	No	Nucleus
5.	*Aa*NAC5	33863.36	8.22	-0.766	No	Nucleus
6.	*Aa*NAC6	24456.95	8.77	-0.665	No	Nucleus
7.	*Aa*NAC7	20238.33	9.63	-0.733	No	Nucleus
8.	*Aa*NAC8	19253.08	9.37	-0.615	No	Nucleus
9.	*Aa*NAC9	13473.19	6.28	-0.668	No	Nucleus
10.	*Aa*NAC10	17170.75	9.41	-0.614	No	Nucleus
11.	*Aa*NAC11	8183.23	4.71	-0.397	No	Nucleus
12.	*Aa*NAC12	6635.58	9.85	-0.789	No	Nucleus
13.	*Aa*NAC13	9305.65	4.85	-0.665	No	Nucleus
14.	*Aa*NAC14	48659.84	4.53	-0.618	No	Nucleus
15.	*Aa*NAC15	29400.12	8.94	-0.687	No	Nucleus
16.	*Aa*NAC16	23015.05	9.43	-0.732	No	Nucleus
17.	*Aa*NAC17	24596.31	9.21	-0.541	No	Nucleus
18.	*Aa*NAC18	36988.00	8.62	-0.773	No	Nucleus
19.	*Aa*NAC19	38074.14	4.52	-0.804	No	Nucleus
20.	*Aa*NAC20	38277.17	4.98	-0.755	No	Nucleus
21.	*Aa*NAC21	45695.26	4.91	-0.447	No	Nucleus
22.	*Aa*NAC22	17982.40	9.40	-0.788	No	Nucleus
23.	*Aa*NAC23	28339.29	9.40	-0.688	No	Nucleus
24.	*Aa*NAC24	32710.09	6.72	-0.670	No	Nucleus
25.	*Aa*NAC25	29871.03	5.76	-0.830	No	Nucleus
26.	*Aa*NAC26	37189.15	4.83	-0.812	No	Nucleus
27.	*Aa*NAC27	25335.32	4.29	-0.852	No	Nucleus
28.	*Aa*NAC28	40404.19	4.74	-0.534	No	Nucleus

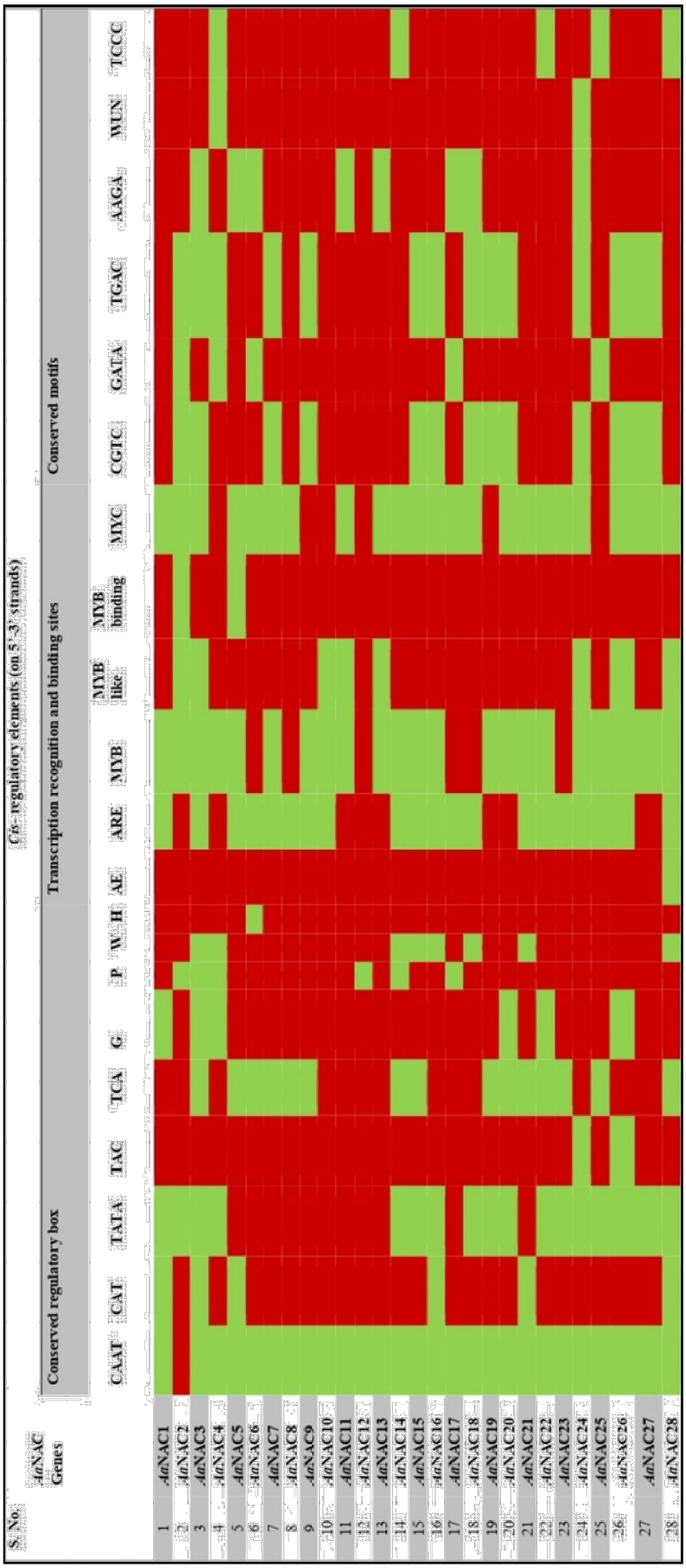

Fig. 4: Distribution of the regulatory *cis*-elements across 28 NAC proteins of *Artemisia annua*.

Discussion

In this report, relationship of *Artemisia* was advocated with the other members of Asteraceae family. By performing local alignment of *Aa*NAC proteins of *A. annua* we attested the close relationship with *Helianthus annus* and *Lectuca sativa* which reveal the monophyletic origin and same lineage (Liu *et al*., 2014). Besides Asteraceae, other members of different family have also showed sequence similarity especially near NAM domain which attested the evolution from nearby descendents. In context to relatedness of 28 *Aa*NAC members to each other, phylogenetic tree was constructed which clustered the whole query in 7 groups. This can clearly indicate the homologous nature of these proteins. During clustering, gap regions found between NAM domains of the NAC protein which can play crucial role in course of evolution. Gap sequences were conserved in nature thus members having gaps were seemed to cluster together in same nearby groups. Gaps between signature domains of NAC protein built the tree more conservative (Shen *et al*., 2020). In some member there was a complete absence of any conserved motif may be due to incomplete N- terminus or some evolutionary changes occurred during chasing the environmental criticalities.

Being multidisciplinary, all the transcription factors are known to contain multiple sites for DNA binding elements which are regulatory in nature (Phillips and Hoopes, 2008). *Cis*- regulatory motif recognition has suggested the presence of conserved boxes like TATA, CAAT and TCC in NAC genes of *A. annua*. This confirmed the regulatory role of NAC domain in various biological processes of plants like transcription and protein synthesis. Since, all the TFs work in a complex manner with other such entities by cross talking thus, presence of such motifs was also examined in *Aa*NAC proteins. This kind of synergistic approach to perform against environmental harshness by maintaining the primary

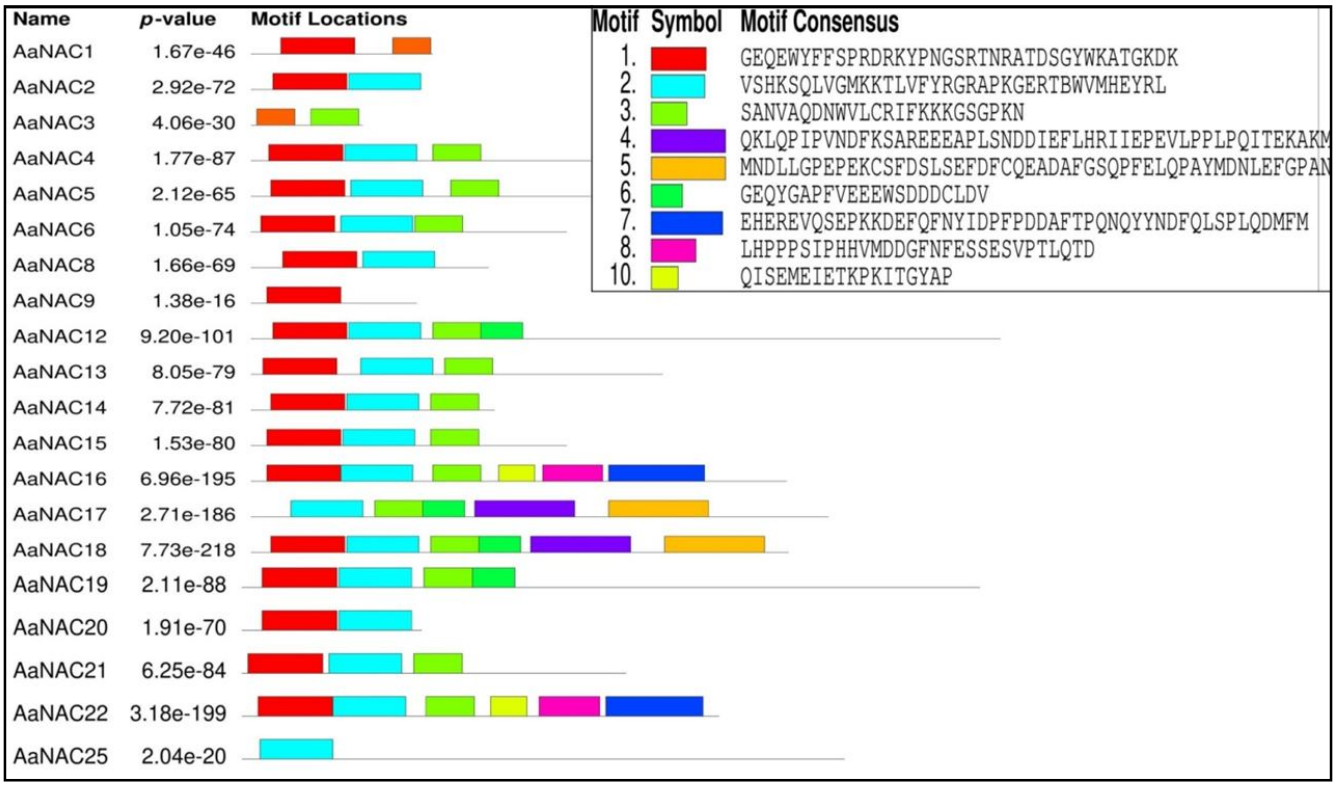

Fig. 5: Distribution patterns and schematic representation of 10 conserved motifs among *Artemisia annua* NAC proteins identified by the MEME programme.

metabolic pool is possible by stress responsive TFs like MYB, WRKY, MYC and DREB. Majority of the members of *Aa*NAC proteins were found to have such transcription recognition and binding sites. Next to this reference, NAC is also known to perform crucial role in abiotic stress tolerance by regulating the expression of other stress responsive transcription factors. Such results were advocated by the presence of WUN, GATA, TGAC, CGTC box. WUN (wound responsive elements) with its sequence TCATTACGAA was observed to be responsible for overexpression of stilbene synthase enzyme in *Vitis pseudoreticulata* (Xu *et al.*, 2011). Besides the role of NAC genes in developmental processes and biotic stress tolerances, some reports reveal its major role in abiotic stress like light induced alterations (Yong *et al.*, 2019). GATA box is also well reported in maintaining the light induced enhanced secondary metabolism, thus it is a matter of working behaviour of NAC and GATA together (Gupta *et al.*, 2017).

Since, vast variety of TFs is protein in nature thus their molecular weight is the important physical characteristic in way of their performance. We have found the large variations among weight of *Aa*NAC proteins perhaps due to presence of some partial sequences in which NAM domains were either totally absent or only trace region was observed. Same reason was found behind the values of Isoelectric point and Hydropathicity indexing. Contrary to this, all NAC members showed common pattern of results in context to non- polar behaviour, nuclear localization and absence of any transmembrane domain. However in some reports transmembrane domains are attested at both conserved N- terminus as well as at variable region of the C-terminus domain of NAC proteins (Mohanta *et al*., 2020).

Conclusion

Functional and physical analysis of *Aa*NAC proteins revealed the sequential conservation of these transcription factors at N- terminus, harbouring signature domain i.e. NAM while high variability was found near regulatory C- terminus. Significant similarity with other members (rather than Asteraceae family) argues for the evolutionary nature and common role of NAC proteins all over the plant kingdom. Presence of sites for *cis*-regulatory box like TATA, CAAT represented the fundamental role of NAC in plant primary metabolism and other developmental processes. On the other hand, binding affinity of same protein with other stress responsive TFs like MYB, MYC, ARE and WUN, GATA revealed the role in potentiate the abiotic stress tolerance.

Acknowledgement

We are thankful to DST-FIST programmes and CAS

for providing funds to Department of Botany, Institute of Science, Banaras Hindu University, along with BHU ISLS facility. Lakee Sharma is thankful to CSIR, New Delhi, India for providing support in the form of fellowship to carry out this research work.

Conflict of interest

There is no any conflict of interest between authors.

References

Aida, M., T. Ishida, H. Fukaki, H. Fujisawa and M. Tasaka (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The plant cell*, **9(6):** 841-857.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman (1990). Basic local alignment search tool. *Journal of molecular biology*, **215(3):** 403-410.

Bailey, T.L. and G. Michael (1998). Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics*, **14(1):** 48–54.

Bailey, T.L., N. Williams, C. Misleh and W.W. Li (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, **34:** W369–W3.

Balazadeh, S., H. Siddiqui, A.D. Allu, L.P. Matallana- Ramirez, C. Caldana, M. Mehrnia, Maria Inés Zanor, B. Köhler and B. Mueller-Roeber (2010). A gene regulatory network controlled by the NAC transcription factor ANAC092/ AtNAC2/ORE1 during salt-promoted senescence. *The Plant Journal*, **62(2):** 250-264.

Brown, G.D. (2010). The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of Artemisia annuaL.(Qinghao). *Molecules*, **15(11):** 7603-7698.

Chou, K.C. and H.B. Shen (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PloS one*, **5(6):** e11335.

De Castro, E., C.J. Sigrist, A. Gattiker, V. Bulliard, P.S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch and N. Hulo (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, **34(2):** W362-W365.

Fujita, M., Y. Fujita, K. Maruyama, M. Seki, K. Hiratsu, M. Ohme-Takagi, L.S.P. Tran, K. Yamaguchi Shinozaki and K. Shinozaki (2004). A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal*, **39(6):** 863-876.

Gasteiger, E., C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel and A. Bairoch (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, 571-607.

Gupta, P., K.K. Nutan, S.L. Singla-Pareek and A. Pareek (2017). Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice. *Frontiers in plant science*, **8:** 1944.

He, X.J., R.L. Mu, W.H. Cao, Z.G. Zhang, J.S. Zhang and S.Y. Chen (2005). AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *The Plant Journal*, **44(6):** 903-916.

Hibara, K.I., S. Takada and M. Tasaka (2003). CUC1 gene activates the expression of SAM-related genes to induce adventitious shoot formation. *The Plant Journal*, **36(5):** 687-696.

Higo K, Y. Ugawa, M. Iwamoto and T. Korenaga (1999). Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.*, **27:** 297–300.

Hu, H., M. Dai, J. Yao, B. Xiao, X. Li, Q. Zhang and L. Xiong (2006). Overexpressing a NAM, ATAF and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proceedings of the National Academy of Sciences*, **103(35):** 12987-12992.

Hu, R., G. Qi, Y. Kong, D. Kong, Q. Gao and G. Zhou (2010). Comprehensive analysis of NAC domain transcription factor gene family in Populus trichocarpa. *BMC plant biology*, **10(1):** 145.

Jensen, C.R., A. Battilani, F. Plauborg, G. Psarras, K. Chartzoulakis, F. Janowiak, R. Stikic, Z. Jovanovic, G. Li, X. Qi, F. Liu, Sven-Erik Jacobsen and M.N. Andersen (2010). Deficit irrigation based on drought tolerance and root signalling in potatoes and tomatoes. *Agricultural Water Management*, **98(3):** 403-413.

Jeong, J.S., Y.S. Kim, K.H. Baek, H. Jung, S.H. Ha, Y. Do Choi, M. Kim, C. Reuzeau and J.K. Kim (2010). Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. *Plant physiology*, **153(1):** 185-197.

Kim, Y.S., S.G. Kim, J.E. Park, H.Y. Park, M.H. Lim, N.H. Chua and C.M. Park (2006). A membrane-bound NAC transcription factor regulates cell division in Arabidopsis. *The Plant Cell*, **18(11):** 3132-3144.

Krogh, A., B. Larsson, G. Von Heijne and E.L. Sonnhammer (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, **305(3):** 567-580.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones and M.A. Marra (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, **19(9):** 1639-1645.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23(21):** 2947-2948.

Lee, S., P.J. Seo, H.J. Lee and C.M. Park (2012). A NAC transcription factor NTL4 promotes reactive oxygen species production during drought-induced leaf senescence in Arabidopsis. *The Plant Journal*, **70(5):** 831-844.

Lescot, M., P. Dehais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouze´ and S. Rombauts (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, **30:** 325–327.

Liu, T., X. Song, W. Duan, Z. Huang, G. Liu, Y. Li and X. Hou (2014). Genome-wide analysis and expression patterns of NAC transcription factor family under different developmental stages and abiotic stresses in Chinese cabbage. *Plant Molecular Biology Reporter*, **32(5):** 1041-1056.

Liu, Y., N. Huo, L. Dong, Y. Wang, S. Zhang, H.A. Youung, X. Feng and Y.Q. Gu (2013). Complete chloroplast genome sequences of Mongolia medicine Artemisia frigida and phylogenetic relationships with other plants. *PloS one,* **8(2):** e57533.

Lockhart, J. (2013). Frenemies: antagonistic bHLH/bZIP transcription factors integrate light and reactive oxygen species signaling in Arabidopsis.

Lv, Z., S. Wang, F. Zhang, L. Chen, X. Hao, Q. Pan, X. Fu, L. Li, X. Sun and K. Tang (2016). Overexpression of a novel NAC domain-containing transcription factor gene (AaNAC1) enhances the content of artemisinin and increases tolerance to drought and Botrytis cinerea in Artemisia annua. *Plant and Cell Physiology*, **57(9):** 1961-1971.

Mistry, J., A. Bateman and R.D. Finn (2007). Predicting active site residue annotations in the Pfam database. *BMC bioinformatics*, **8(1):** 298.

Mohanta, T.K., D. Yadav, A. Khan, A. Hashem, B. Tabassum, A.L. Khna, E.F. Abd-Allah and A. Al-Haraasi (2020). Genomics, molecular and evolutionary perspective of NAC transcription factors. *PloS one,* **15(4):** e0231425.

Nuruzzaman, M., R. Manimekalai, A.M. Sharoni, K. Satoh, H. Kondoh, H. Ooka and S. Kikuchi (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene*, **465(1-2):** 30-44.

Ooka, H., K. Satoh, K. Doi, T. Nagata, Y. Otomo, K. Murakami, K. Matsubara, Y. Hayashizaki *et al.,* (2003). Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. *DNA research*, **10(6):** 239-247.

Phillips, T. and L. Hoopes (2008). Transcription factors and transcriptional control in eukaryotic cells. *Nature Education*, **1(1):** 119.

Puranik, S., P.P. Sahu, S.N. Mandal, S.K. Parida and M. Prasad (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(5):** e64594.

Puranik, S., P.P. Sahu, P.S. Srivastava and M. Prasad (2012). NAC proteins: regulation and role in stress tolerance. *Trends in plant science*, **17(6):** 369-381.

Sablowski, R.W. and E.M. Meyerowitz (1998). A homolog of NO APICAL MERISTEM is an immediate target of the floral homeotic genes APETALA3/ PISTILLATA. *Cell*, **92(1):** 93-103.

Schultz, J., R.R. Copley, T. Doerks, C.P. Ponting and P. Bork (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic acids research*, **28(1):** 231-234.

Shen S., Z. Qianru, S. Yu, S. Zhenmei, Z. Qianqian, H. Sijia, W. Rongling, J. Libo, Z. Xiyang and G. Yunqian. (2020). Genome-wide analysis of the NAC domain transcription factor gene family in Theobroma cacao. *Genes*, **11(1):** 35.

Singh, A.K., V. Sharma, A.K. Pal, V. Acharya and P.S. Ahuja (2013). Genome-wide organization and expression profiling of the NAC transcription factor family in potato (Solanum tuberosum L.). *DNA research*, **20(4):** 403-423.

Souer, E., A. Van Houwelingen, D. Kloos, J. Mol and R. Koes (1996). The no apical meristem gene of Petunia is required for pattern formation in embryos and flowers and is expressed at meristem and primordia boundaries. *Cell*, **85(2):** 159-170.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, **28(6):** 1102-1104.

Tamura, K., J. Dudley, M. Nei and S. Kumar (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*, **24(8):** 1596-1599.

Tran, L.S.P., K. Nakashima, Y. Sakuma, S.D. Simpson, Y. Fujita, K. Maruyama, M. Fujita, M. Seki, K. Shinozaki and K. Yamaguchi-Shinozaki (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *The Plant Cell*, **16(9):** 2481-2498.

Uauy, C., A. Distelfeld, T. Fahima, A. Blechl and J. Dubcovsky (2006). A NAC gene regulating senescence improves grain protein, zinc and iron content in wheat. *Science*, **314(5803):** 1298-1301.

Wang, N., Y. Zheng, H. Xin, L. Fang and S. Li (2013). Comprehensive analysis of NAC domain transcription factor gene family in Vitis vinifera. *Plant cell reports*, **32(1):** 61-75.

Wang, S.L., X.R. Xu, Y.X. Sun, J.L. Liu and H.B. Li (2013). Heavy metal pollution in coastal areas of South China: a review. *Marine pollution bulletin*, **76(1-2):** 7-15.

Wei, S., L. Gao, Y. Zhang, F. Zhang, X. Yang and D. Huang (2016). Genome-wide investigation of the NAC transcription factor family in melon (*Cucumis melo* L.) and their expression analysis under salt stress. *Plant cell reports*, **35(9):** 1827-1839.

Xie, Q., H.S. Guo, G. Dallman, S. Fang, A.M. Weissman and N.H. Chua (2002). SINAT5 promotes ubiquitin-related degradation of NAC1 to attenuate auxin signals.

Nature, **419(6903):** 167-170.

Xu, W., Y. Yu, Q. Zhou, J. Ding, L. Dai, X. Xie, Y. Xu, C. Zhang and Y. Wang (2011). Expression pattern, genomic structure and promoter analysis of the gene encoding stilbene synthase from Chinese wild Vitis pseudoreticulata. *Journal of experimental botany,* **62(8):** 2745-2761.

Yang, X. and L. Li (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics*, **27(18):** 2614-2615.

Yong Y., Y. Zhang and Y. Liu (2019). A Stress-Responsive NAC transcription factor from tiger lily (LlNAC2) interacts with LlDREB1 and LlZHFD4 and enhances various abiotic stress tolerance in Arabidopsis." *International journal of molecular sciences*. **20(13):** 3225.

Zheng, C., D. Jiang, F. Liu, T. Dai, W. Liu, Q. Jing and W. Cao (2009). Exogenous nitric oxide improves seed germination in wheat against mitochondrial oxidative damage induced by high salinity. *Environmental and Experimental Botany*, **67(1):** 222-227.