DOI : 10.5958/0974-0112.2022.00002.0

Indian J. Hortic. 79(1), March 2022: 9-14

Molecular cloning and *in-silico* characterization of *NAC86* of *kalmegh* (*Andrographis paniculata*)

Ramesh Kumar[1], Chavlesh Kumar[2], Rishu Jain, Avantika Maurya, Ashok Kumar, Abha Kumari[1] and Rakesh Singh[*]

ICAR-National Bureau of Plant Genetic Resources, New Delhi - 110012, Delhi, India.

ABSTRACT

The Kalmegh (*Andrographis paniculata*) is a medicinal plant belonging to the family *Acanthaceae* and is widely cultivated in India for several medicinal uses. The NAC transcription factors are protein molecules that regulate the expression of targeted genes involved in controlling several biological mechanisms in plants, like growth and development and biotic/abiotic stresses. Therefore, NAC transcription factors are one of the emerging tools for crop improvement. The present study involves identifying and isolating the ApNAC86 gene, and its In-silico analysis, viz., tertiary structure prediction and its validation, binding prediction with the DNA molecule, and phylogenetic analysis. The isolated ApNAC86 has a molecular weight of 27.95 kD, and a theoretical isoelectric point (pI) is 6.76. The protein structure prediction and their validation through the Ramachandran plot indicated 87.27% of residues in the most favoured region, indicating good quality. Further, phylogenetic analysis revealed ApNAC86 is closely related to NAC86 from *Phtheirospermum japonicum*. Thus, the identified gene, ApNAC86, could be an essential genomic resource for the Kalmegh improvement programme.

Keywords: *Andrographis paniculata, Kalmegh,* Transcription factor, Ramachandran plot, *ApNAC86*, Phylogenetic analysis

INTRODUCTION

The *Kalmegh* (*Andrographis paniculata*) one of the important medicinal plants belonging to the Acanthaceae family, often referred as the 'King of Bitters'. The leaves and stem portion of the plant contain biologically active compound like andrographolide (Gajbhiye et al., 4). The active compound of *Kalmegh* have several pharmacological effects: immunostimulatory, anti-typhoid, anticancer, cardiovascular, anti-microbial, anti-inflammatory, anti-HIV, anti-parasitic, antioxidant, and antidiarrheal (Kurian et al., 8). Furthermore, The importance of the plant against the SARS-CoV-2 has come into the limelight, as suggested by recent reports (Hiremath et al., 7; Rajagopal, 10).

The NAC (NAM-No apical meristem, ATAF-*Arabidopsis thaliana* activation factor and CUC- cup shaped cotyledon) family is one of the important plant-specific transcription factor (TF) families. The NAC protein comprises two domains; one is a highly conserved domain (NAC) located at N- terminus and has DNA and protein binding ability. Another is a highly variable domain (transcriptional regulatory) located at C-terminus and can act as activator or repressor. Some of the family members inhibit or activate certain stress-related genes, which aid plants in coping with various environmental stresses such as salinity, cold, and drought. Besides, other family members regulate genes that are involved in plant growth and developmental processes (Bu et al., 2; Hao et al., 6; Trapnell et al., 12)pathogen infection and mechanical wounding. In this report, we provided biochemical and genetic evidence to show that the Arabidopsis thaliana NAC family proteins ANAC019 and ANAC055 might function as transcription activators to regulate JA-induced expression of defense genes. The role of the two NAC genes in JA signaling was examined with the anac019 anac055 double mutant and with transgenic plants overexpressing ANAC019 or ANAC055. The anac019 anac055 double mutant plants showed attenuated JA-induced VEGETATIVE STORAGE PROTEIN1 (VSP1. The NAC TF was first reported in *Arabidopsis* (Aida et al., 1). Since its discovery, the research work on the NAC transcription factor has progressed over time in different plant species. Genome-based analysis of NAC proteins sequence reveals its existence in more than 150 plant species. To date, only one member of the NAC family is reported in *A. paniculata.* Being one of the most important medicinally plants, other members of the NAC family need to be explored in *A. paniculata*.

*Corresponding author: rakesh.singh2@icar.gov.in

[1]Amity Institute of Biotechnology, Amity University, Noida- 201313, Uttar Pradesh, India

[2]Div.of Fruits and Horticultural Technology, ICAR-IARI, New Delhi -110012, Delhi, India.

Therefore, the present effort was attempted to isolate the *ApNAC86* from *A. paniculata.* The isolated *ApNAC86* was subjected to ab initio modelling and its *In-silico* quality assessment was done. In addition, conserved domain prediction, conserved motifs prediction, genic SSRs prediction and phylogenetic analysis were also performed.

MATERIALS AND METHODS

The sequence data of *NAC86* was retrieved from Medicinal Plants Genomic Resource Database, developed, and maintained by ICAR-NBPGR, New Delhi, India. The coding region of NAC 86 was identified using expasy translator as well as pBLAST tool and gene-specific primer (GSP) pair for the isolation of CDS region of NAC86 was designed. The sequence information of the primers used in the present study is given in Table 1.

The *A. paniculata* accession seeds (IC342136) were collected from the national gene bank at ICAR-NBPGR, New Delhi. The seeds were sown in pot at the IARI-National Phytotron facility. The leaves from the young plant (60 days after sowing) were collected in liquid nitrogen and subjected to total RNA isolation using RNA isolation kit (Thermo Fisher Scientific, USA) (Fig.1).

Fig. 1. *Andrographis paniculata* plant

The DNA impurities were removed using *DNase* (Thermo Fisher Scientific, USA) treatment. The quality of purified RNA was assessed using Nanodrop and concentration was determined.

Further, cDNA was synthesized with the help of cDNA Synthesis kit (Invitrogen, USA) and confirmed by actin PCR amplification. The 20 µL PCR reaction containing 1 µL diluted cDNA (1:10) as a template, 2 µL of 10X PCR reaction buffer, 1.8 µL of 2.5 mM dNTPs, 1.5 µL of 2.5 mM $MgCl_2$, 0.6 µL of each gene specific primer (10 nmol), and 0.3 µL of Ex *Taq DNA polymerase* (TaKaRa, Japan), with 12.2 µL distilled water (autoclaved) added to make the final reaction volume of 20 µL. Then CDS region of *ApNAC86* was amplified in thermocycler using following PCR program: initial denaturation at 95 ºC for 4 min, followed by 35 cycles of denaturation at 95 ºC for 40 s, annealing temperature at 60 ºC for 50 s, and extension at 72 ºC for 50 s, and a final extension at 72 °C for 7 min. Further, PCR amplicon was observed in 2% agarose gel, gel image was captured, and desired single band was purified. The purified PCR product was ligated in pCR2.1 vector (Thermo Fisher Scientific, USA) and then transformed. The positive clones were screened using XGAL/IPTG selection method and subjected to colony PCR. The plasmids were isolated from five positive clones and sequenced (Macrogen Inc., Seoul, Korea). The sequenced plasmids were subjected to multiple sequence alignment to determine the overlapping region. The overlapped region was used to perform pBLAST to designate the name of gene. Further, the amino acid sequence and the gene's physical properties was calculated using Expasy server. The secondary, tertiary structure, and DNA binding pocket of *ApNAC86* protein were predicted with the help of I-TASSER webserver (https://zhanglab.dcmb.med.umich.edu). Furthermore, the predicted structure model was verified using the MolProbity server (http://molprobity.biochem.edu/), and the Ramachandran plot. Additionally, EST (Expressed sequence tag) SSRs (Simple sequence repeats) were also identified using SSR locator (http://microsatellite.org/ssr.php).

RESULTS AND DISCUSSION

The Expasy tool and pBLAST revealed the identity of *NAC86* from *A. paniculata*. To isolate

Table 1. The details of the primers used.

Primer ID	Forward (5'-3')	Reverse (5'-3')	Ta (ºC)	Application
N6	ATGACGATGACGATGAAATTG	TTAGTCTTGTGCATGAGTGACT	60	Coding region amplification of *ApNAC86*
Actin	AGGGACATCAAAGAGAAACTG	GCTGTGATCTCCTTGCTCATT	55	Actin amplification

Where. Ta- Annealing temperature

the gene, the nucleotide sequence was used to design gene-specific primers and subjected to PCR amplification. The total RNA was isolated from leaves of *A. paniculata* and RNA samples having 260/280 ratio of ~2.0 were further utilized for cDNA synthesis. To check the quality of cDNA, amplification of actin (house-keeping gene) was performed (Fig.2A). The NAC gene was amplified using appropriate primers and gave an amplicon of about 741bp (Fig.2B). The PCR amplicon was further purified from 2% agarose gel and ligated to pCR2.1 vector and further cloned in DH5α competent cells. The plasmids from 5 different clones were isolated and sequenced. Analysis of sequences gave the accurate nucleotide sequence of the CDS region.

The isolated sequence of NAC protein was designated as *ApNAC86* as its pBLAST result revealed the highest homology (75.72%) with NAC domain-containing protein 86 from *Phtheirospermum japonicum* (GFP92543.1). The isolated nucleotide sequence of *ApNAC86* was analysed and submitted to NCBI GenBank with accession number- MZ458368 (Fig. 3).

The *ApNAC86* gene is being translated into a protein containing 246 amino acids. The molecular weight, and iso-electric point (pI) of the protein was found to be 27.95 kD and 6.76, respectively. Diao et al. (Diao *et al*., 3) ATAF1/2, and CUC2 (NAC also reported NAC 86 from *Capsicum annuum* with an amino acid of 299, molecular weight of 34.40 kD, and pI of 6.68, which is congruent with the present study. The instability index (II) is a parameter that determines the stability of a protein normally. The proteins having II value below 40 are considered as stable protein. The Aliphatic index (AI) of a protein molecule is refers to the relative volume of aliphatic sidechains (alanine, leucine, isoleucine, and valine). The proteins whose AI value is high are considered as thermostable. The stability index and aliphatic index values, i.e., 32.19 and 59.57 respectively, of the isolated NAC indicated that the protein is thermostable. The total number of negatively charged residues (Asp + Glu), the total number of positively charged residues (Arg + Lys) and grand average of hydropathicity (GRAVY) found to be 37, 36 and -0.791, respectively. The details of physiochemical properties of *ApNAC86* are given in Table 2. ScanProsite (http:// www.expasy.ch/tools/scanprosite/) analysis indicated that *ApNAC86* contained a NAC domain from 10 to 169 amino acid residues (Fig.4). The NCBI conserved domain search revealed the presence of NAM domain superfamily (11-145) in the *ApNAC86*, suggesting that *ApNAC86* is NAM super-family protein. The NAM family proteins are known for their role in plant development.

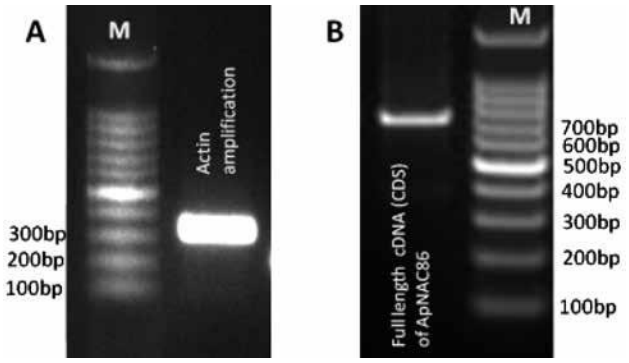

Fig. 2. The gel image of PCR amplification: A. *Actin* PCR amplification B. *ApNAC86* PCR amplification

```
1   ATGACGATGACGATGAAATTGGAGGCTTTGCCTAAGGGTTTCAGGTTCCGGCCGACGGACGAGGAGCTCATC
2   AGCCATTATCTCAGGCTCAAAATTAACGGCCACAATTCTTCCGTTCAGGCTATCCCTGAGGTTGATATCAGCAA
3   AAGAGAGCCCTGGGATCTTCCCGCACTCTCCGTGATAAAGTCTGATGACCCTGAATGGTTCTTCTTCTGTCCTC
4   GCGATAGAAAATATCCAAATGGTCATCGTTCTAATAGAGCTACAGAAGCAGGTTATTGGAAAGCTACTGGAA
5   AGGATCGCATGATCAAATCCCGCAAATTGTTGTCTTCAGGTCACTCCAGTTCCCACATAATTGGGATGAAGAA
6   GACTCTAGTTTTTCACAGGGGCCGAGCTCCCAAGGGTGAACGGACAAATTGGATAATGCGTGAGTATCGTGC
7   TGCTGAACCTGATCTTGATGGAACTGGACCTGGACAGGGAGATTATGTCCTGTGTCGCTTTTTCCACAAGGTT
8   GATGAGAAGGTCGAGAAATGTGATGATGCTGACCATACAGGATTATCTCCATCTACTAACAAGTCATCACCGG
9   ATGAGGTATCATCAGATTTGTTTCGAGAACCTGCAATGTTAGATGTGAAAAGTATAAAGGAACCTGAGGACAT
10  GAGTACATGGTGGACAGATCAGGCTGATTGTATGACTCCTATTAACCAAGTGTCATCAGGTGGAGTCACTCAT
11  GCACAAGACTAA
```

Fig. 3. The nucleotide sequence of isolated *ApNAC86*. Start codon (ATG) and Stop codon (TAA) are green highlighted.

Table 2. Physiochemical specifications of *ApNAC86* protein.

S. No.	Physicochemical parameters	*ApNAC86*
1	Length (total no. of amino acids)	246
2	Molecular weight (MW)	27.95 kDa
3	Iso-electric point (PI)	6.76
4	Instability index (II)	32.19
5	Total number of negatively charged residues (Asp + Glu)	37
6	Total number of positively charged residues (Arg + Lys)	36
7	Aliphatic index (AI)	59.47
8	Grand average of hydropathicity (GRAVY)	-0.791

Fig. 4. ScanProsite Results of *ApNAC86*.

The sub-cellular localization predicted that the protein is mainly found in the nucleus. Several reports indicated that NAC TFs are mainly present in the nucleus, which is also congruent with the present investigation. The MEME suit discovered five conserved motifs in the amino acid sequence of *ApNAC86*, indicating 1-5 conserved sub-domain of NAC (Fig. 5). The experimentally approved NACs at protein level from *Oryza sativa* (*OsNAC6*, *OsNAC5*) and *A. thaliana* (*ANAC037*, *ANAC076*, *ANAC016*, *ANAC096*, *CUC3)* were used to discover conserved motifs in *ApNAC86.* The motif locations, symbols and consensus are given in Fig. 5. Additionally, one dimer $(AG)_3$ and one trimer $(TTC)_3$ SSR were found in CDS region of *ApNAC86* (Table3).

The predicted secondary structure of *ApNAC86* contained α-helix (6.5%), β-sheet (18.29%), and coil regions (75.20%), indicating that *ApNAC86* is mainly formed by coil regions (Fig.6), which are reported to be involved in abiotic stress regulation, indicating *ApNAC86* could be involved in abiotic stress management (Garg et al., 5).

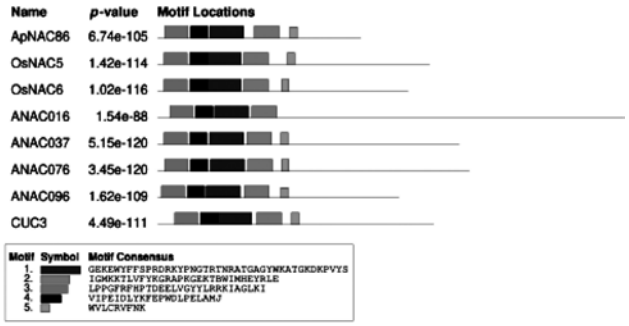

Fig. 5. Predicted motifs in *ApNAC86*

Table 3. Details of SSRs detected in *ApNAC86*

S. No.	Start SSR	End SSR	Type	repeats
1	152	158	2	(AG)3
2	209	218	3	(TTC)3

The tertiary structure of *ApNAC86* was also modelled using I-TASSER server (Fig.7A). The structure was predicted identical to crystal structure of the conserved domain of *OsNAC1* (PDB id- 3ulxA) with a TM score of 0.576 that support our findings (Fig.7B). The details of PDB structure hit are given in Table 4.

The structural assessment of the predicted 3D model of *ApNAC86* protein was carried out using the Ramachandran plot, which indicated that 87.27% residues are in the most favoured region, 2.42 Ramachandran outlier, 4.48% rotamer Outliers (Fig.8). The parameters of the Ramachandran plot have suggested that protein structure model of *ApNAC86* is acceptable and considered a good quality protein. The DNA binding pocket of the protein predicted that amino acid residues (Thr 96, Gly 97, Arg 136, Thr 137, Phe 167, and Lys 169) were form channels for DNA binding (Fig.9).

Earlier, Nawaz et al. (Nawaz et al., 9)including dehydration-responsive element binding (DREB documented 79.5% of residues favoured region from

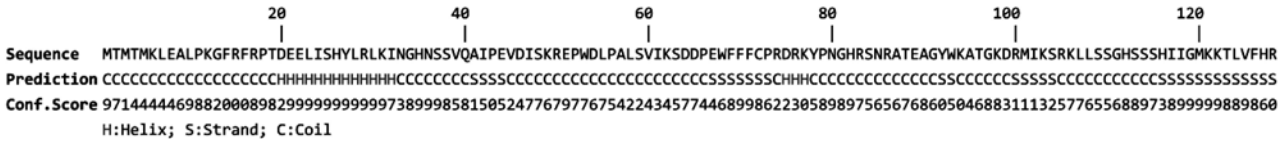

Fig. 6. Secondary structure of *ApNAC86* protein.

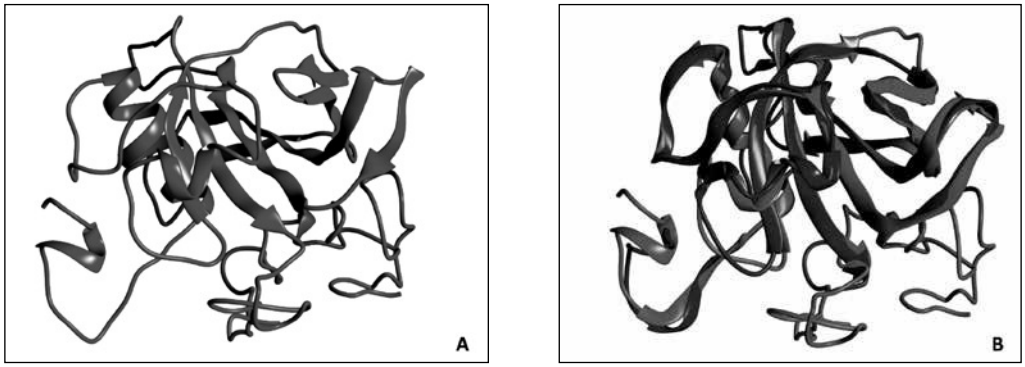

Fig. 7. Predicted protein structure of *ApNAC86* protein: (A). Tertiary structure of *ApNAC86* represented by magenta colour (B). Comparison of predicted structure with the conserved domain of *OsNAC1* represented by blue colour (PDB id- 3ulxA)

Table 4. The details of PDB structure hits

Rank	PDB Hit	TM score	RMSD[a]	IDEN[a]	Cov
1	3ulxA	0.576	1.07	0.486	0.594
2	1ut7B	0.529	2.28	0.503	0.581
3	2x2hA1	0.412	6.42	0.035	0.748
4	6w2jA	0.403	6.24	0.048	0.728
5	2vdcA	0.399	6.29	0.042	0.707
6	6u8yX	0.399	6.19	0.057	0.703
7	3nkmA	0.398	5.66	0.032	0.642
8	2w3yA	0.397	6.08	0.032	0.711
9	7d3uA	0.394	6.07	0.025	0.687
10	6d4hA	0.394	5.85	0.037	0.654

Where RMSD[a]- Root mean square deviation (RMSD) between residues that are structurally aligned.
IDEN[a]- the sequence identity (%) in the structurally aligned region.
Cov- the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein.

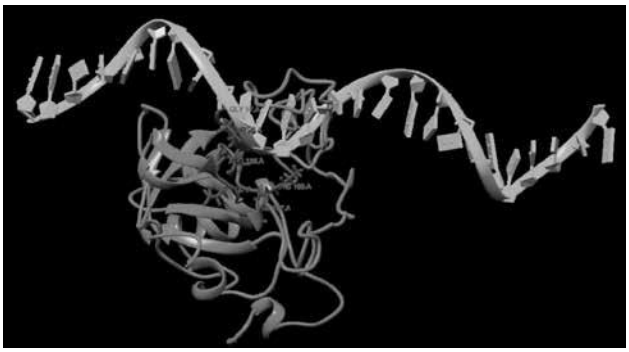

Fig. 9. Interaction of *ApNAC86* with DNA molecule. The protein structure of *ApNAC86* represented by magenta colour; DNA molecule is represented by green colour; and Residues interacting with DNA are represented by blue colour.

DREB1A from *O. sativa* var. IR6, which is congruent to present findings. The phylogenetic tree (N-J) of *ApNAC86* with NACs from other plants revealed that NAC86 from *A. paniculata* (GenBank accession-MZ458368) was closely related with NAC86 from *Phtheirospermum japonicum* (GenBank accession-GFP92543.1) (Fig.10). The green highlighted branches also suggesting that NAC reported from *Sesamum indicum* (GenBank accessions- XP_011081339.1 and XP_011070419), *Erythranthe guttata* (GenBank accession- XP_012839656.1), *Olea europaea* var. *sylvestris* (GenBank accession-XP_022878541.1), and *Olea europaea* subsp. *europaea* (GenBank accession-CAA3006313.1) is a close relative of *A. paniculata.* Interestingly, the grouping pattern of *ApNAC86* showed that *P. japonicum*, *E. guttata*,

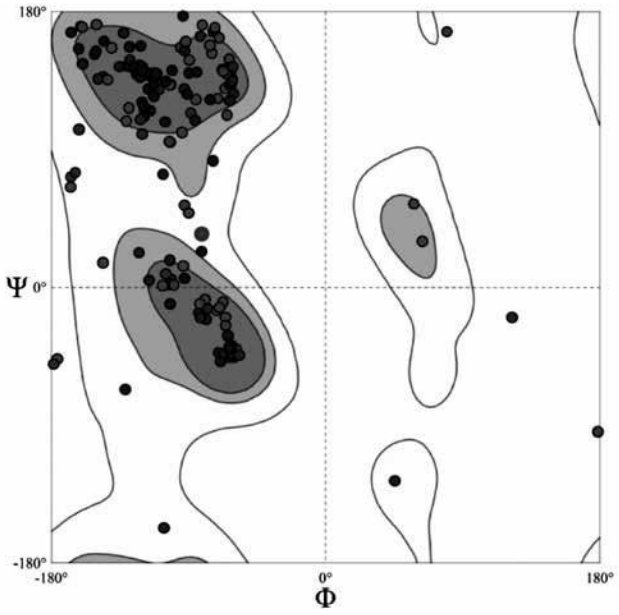

Fig. 8. Ramachandran plot of *ApNAC86* protein model.

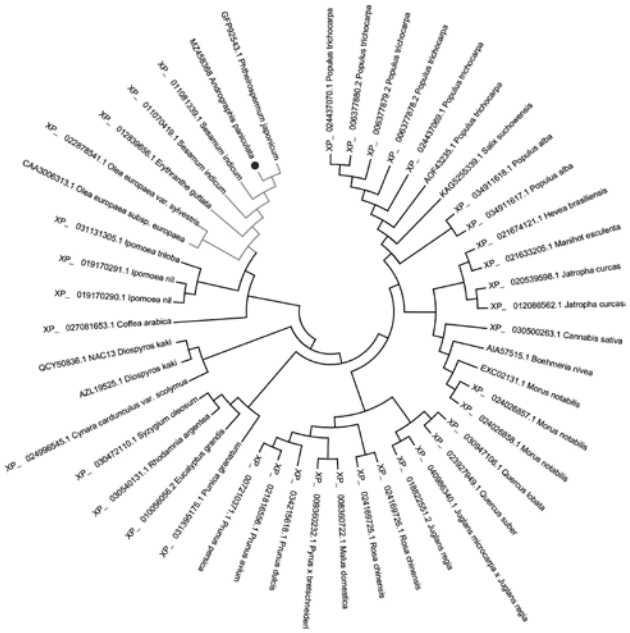

Fig. 10. Phylogenetic tree of *ApNAC86* with NACs reported in other plants. Pink colour circle represents NAC86 from *A. paniculata* (NCBI GenBank accession- MZ458368); Green branches of the tree represent the close relatives of *A. paniculata*.

O. europaea and *S. indicum* might be related plant species to *A. paniculata* because they come from same botanical classification order (Lamiales). Recent report also indicated that *A. paniculata* contained several gene families similar to *S. indicum*, which is congruent with present investigation (Sun *et al.*, 11)India and

other Southeast Asian countries. A. paniculata was used as a crucial therapeutic treatment during the influenza epidemic of 1919 in India, and is still used for the treatment of infectious disease in China. A. paniculata produces large quantities of the anti-inflammatory diterpenoid lactones andrographolide and neoandrographolide, and their analogs, which are touted to be the next generation of natural anti-inflammatory medicines for lung diseases, hepatitis, neurodegenerative disorders, autoimmune disorders and inflammatory skin diseases. Here, we report a chromosome-scale A. paniculata genome sequence of 269 Mb that was assembled by Illumina short reads, PacBio long reads and high-confidence (Hi-C).

AUTHORS' CONTRIBUTION

Conceptualization of research (RS); Designing of the experiments (RS); Contribution of experimental materials (RS, AK[1]); Execution of field/lab experiments and data collection (RK); Analysis of data and interpretation (RK, CK, AM); Preparation of the manuscript (RK, RJ, CK, AK[2]).

DECLARATION

All authors declare that they do not have any conflict of interest.

ACKNOWLEDGEMENT

Authors duly acknowledge Director, ICAR-NBPGR, New Delhi, for financial support to conduct the present study. The research work conducted in the manuscript is the part of PhD objectives of RK and Therefore, RK is thankful to UGC for providing fellowship. The authors also acknowledge the In-charge National Phytotron Facility, ICAR-IARI, New Delhi, for proving the space.

REFERENCES

1. Aida, M., Ishida, T., Fukaki, H., Fujisawa, H. and Tasaka, M. 1997. Genes involved in organ separation in Arabidopsis: An analysis of the cup-shaped cotyledon mutant. *Plant Cell* **9**: 841-57.

2. Bu, Q., Jiang, H., Li, C.B., Zhai, Q., Zhang, J., Wu, X., Sun, J., Xie, Q. and Li, C. 2008. Role of the *Arabidopsis thaliana* NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. *Cell Res*. **18**: 756-67.

3. Diao, W., Snyder, J.C., Wang, S., Liu, J., Pan, B., Guo, G., Ge, W. and Dawood, M.H.S.A. 2018. Genome-wide analyses of the NAC transcription factor gene family in *pepper* (*Capsicum annuum L.*): Chromosome location, phylogeny, structure, expression patterns, Cis-elements in the promoter, and interaction network. *Int. J. Mol. Sci.* **19**: 1028.

4. Gajbhiye, N.A. and Khristi, S. 2010. Distribution pattern of andrographolide and total lactones in different parts of Kalmegh plant. *Indian J. Hortic*. **67**: 591-93.

5. Garg, B., Puranik, S., Tuteja, N. and Prasad, M. 2012. and wali5 genes from wheat *Plant Signal. Behav.* **7**: 1393-96.

6. Hao, Y.J., Wei, W., Song, Q.X., Chen, H.W., Zhang, Y.Q., Wang, F., Zou, H.F., Lei, G., Tian, A.G., Zhang, W.K. and Ma, B. 2011. *Soybean* NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. *Plant J.* **68**: 302-13.

7. Hiremath, S., Kumar, H.D.V., Nandan, M., Mantesh, M., Shankarappa, K.S., Venkataravanappa, V., Basha, C.R.J., Reddy, C.N.L., 2021. In silico docking analysis revealed the potential of phytochemicals present in *Phyllanthus amarus* and *Andrographis paniculata*, used in Ayurveda medicine in inhibiting SARS-CoV-2. *3 Biotech* **11**: 1-18.

8. Kurian, A. and Sankar, M.A. 2007. Medicinal Plants: Vol. 02. Horticulture Science Series., Kerala Agriculture University, Thrissur, 374 p.

9. Nawaz, M., Iqbal, N., Idrees, S. and Ullah, I. 2014. DREB1A from *Oryza sativa* var. IR6: Homology modelling and molecular docking. *Turk. J. Bot.* **38**: 1095–1102.

10. Rajagopal, K. 2020. Activity of Phytochemical Constituents of *Curcuma longa* (Turmeric) Against SARS-CoV-2 Main Protease (Covid19): Anin-Silico Approach. *Int. J. Pharm.* **6**: 1-10.

11. Sun, W., Leng, L., Yin, Q., Xu, M.M., Huang, M., Xu, Z., Zhang, Y., Yao, H., Wang, C., Xiong, C., Chen, Sha, Jiang, C., Xie, N., Zheng, X., Wang, Y., Song, C., Peters, R.J. and Chen, Shilin 2019. The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J.* **97**: 841-57.

12. Trapnell, C., Roberts, A., Goff, L., Pertea, G. and protocols, D.K.-N., 2012, undefined, n.d. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**(3): 562-78.

Received : September, 2021; Revised : February, 2022; Accepted : February, 2022