International Journal of Agriculture Sciences
ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 6, 2022, pp.-11391-11396.
Available online at https://bioinfopublication.org/pages/jouarchive.php?id=BPJ0000217

Research Article

IN SILICO ANALYSIS FOR THE GENOME-WIDE IDENTIFICATION OF AP2 SUPERFAMILY TRANSCRIPTION FACTORS IN PEARL MILLET

ANJANI A.[1], RAMANA J.V.[2], SHANKER A.K.[3], LAL A.M.[2] AND SATISH Y.[4]

[1]Department of Molecular Biology and Biotechnology, SVU Agricultural College, Tirupathi, 517502, Acharya N. G. Ranga Agricultural University, Lam, 522034, India
[2]Department of Genetics and Plant Breeding, Agricultural College, Bapatla, Acharya N. G. Ranga Agricultural University, Lam, 522034, Guntur, Andhra Pradesh, India
[3]Principal Scientist, Department of Plant Physiology, ICAR-Central Research Institute for Dryland Agriculture (CRIDA), Hyderabad, 500059, Telangana, India
[4]Senior Scientist, Department of Genetics and Plant Breeding, RARS, Maruteru, Acharya N. G. Ranga Agricultural University, Lam, 522034, Andhra Pradesh, India
*Corresponding Author: Email - arunshank@gmail.com

Received: June 01, 2022; Revised: June 26, 2022; Accepted: June 27, 2022; Published: June 30, 2022

Abstract: The APETALA2/ethylene-responsive element binding factor (AP2/EREB) superfamily is one of the largest transcription factor families in plant kingdom which play a predominant role in plant growth processes and are involved in different biotic and abiotic stress responses. A detailed and comprehensive *in silico* analysis was performed to identify the AP2/ERF transcription factors in pearlmillet which identified 99 AP2/ ERF TFs classified into 19 AP2 family, 62 ERF subfamily, 15 DREB subfamily, 2 RAV family TFs and a soloist. Phylogenetic analysis was performed with the predicted protein sequences and conserved motifs were analysed. Seventy percent AP2 /ERF superfamily genes were found to be localised in nucleus and all of them were mapped onto the seven chromosomes. Physico-chemical parameters were computed and found that thirteen transcription factor proteins were stable. Fold Index was used and found that six transcription factors had no disordered regions and were folded completely. PMERF57 has good stability index and no disordered regions. These identified putative genes could be explored for further analysis.

Keywords: *AP2 superfamily, Transcription factors, Pearlmillet, In silico mapping, Protein folding*

Citation: Anjani A., *et al*., (2022) *In Silico* Analysis for the Genome-wide Identification of AP2 Superfamily Transcription Factors in Pearl Millet. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 14, Issue 6, pp.- 11391-11396.

Academic Editor / Reviewer: Anjani Kumari

Introduction

Pearl millet (*Pennisetum americanum*) Syn. Cenchrus americanus, a coarse cereal, one of the major millets widely grown in Southeast Asia and African countries. Being a staple food for millions of poor people, it is also used for fodder and fuel purposes [1]. Pearl millet can survive under mean annual rainfall 250-300 mm where major crops like maize, sorghum, rice and wheat fail to survive [2]. To many nutritional advantages like high protein and fibre content, low starch and presence of essential micro nutrients like iron and zinc has been added as advantages for pearl millet production to cope up with climate changes and food insecurity [3].

Plants are affected by different abiotic stresses like drought, salinity, low and high temperatures affecting their productivity and yields [4] which have to be monitored and diverse and unique strategies have to be developed [5]. Different physiological and biochemical regulatory mechanisms are involved as responses to these environmental stresses [6]. Many genes are induced in response to these stresses [7] and among these transcription factors play a crucial role in signal transduction pathways and activate many downstream stress responsive genes. These transcription factors bind to the cis-regulatory elements in the promoter region and regulate gene activity.

The AP2/EREBP (APETALA2/ethylene response element-binding protein) superfamily is one of the largest transcription factor families in plants. AP2 super family genes are involved in stress responses and also in the growth and development of plants. Analysis using genetic and molecular pathways has revealed that AP2 super family genes are involved in response to high salinity, drought, temperature changes, disease control and flowering control pathways [8]. First AP2/ERF domain was reported from Arabidopsis which was involved in flower development [9].

The AP2/ERF superfamily gene contain a highly conserved DNA binding domain with 60-70 amino acid residues which is known to interact with cis-acting elements directly with GCC box and/or dehydration responsive element (DRE)/C-repeat element (CRT) present in the promoter regions of downstream target genes [10,11]. Based on the number of DNA binding domains and sequence similarities the AP2/ERF was divided into four families AP2 family, ERF family, RAV family and a soloist [12]. ERF family is divided into two subfamilies; ERF subfamily and DREB subfamily both of them having single AP2 domain and they are mainly involved in environmental stress responses. ERF subfamily possesses conserved alanine and aspartic acid respectively at 14th and 19th position of DNA binding domain whereas DREB subfamily has valine and glutamine at respective positions [13]. The transcription factors of ERF subfamily bind to GCC boxes and are involved in hormone signalling pathways and also contribute to biotic stresses like disease stimuli [14] and abiotic stresses like drought, salinity and freezing tolerance.

DREB subfamily TFs bind to DRE/CRT elements and they are also involved in abiotic stress tolerance. AP2 family proteins possess two AP2/ERF binding domains involving plant regulating process [15] like flower development and seed growth. RAV proteins possess one AP2 domain and a B3 domain which are regulated by brassinosteroids or ethylene hormones which are directly involved in biotic and abiotic stress responses [16,17].

AP2 TFs were identified in different crops like maize [18], sorghum [19], rice [20], wheat [21], foxtail millet [22], soybean [23], castor [24], cauliflower [25], *Populus trichocarpha* [26], *Vitis vinifera* [27], zoysia grass [28]. The present study was conducted as the AP2/ERF superfamily was unexplored in pearlmillet and putative genes were identified.

Table-1 *Summary of AP2 superfamily genes identified in Pearlmillet*

Gene name	Gene Id	Family	Gene length (bp)	Amino acid residues (aa)
g4292-t1_chr1	PMERF1	ERF	296	98
g175-t1_chr2	PMERF12	ERF	719	224
g553-t1_chr3	PMERF25	ERF	3089	369
g4969-t1_chr4	PMERF38	ERF	635	211
g3012-t1_chr5	PMERF44	ERF	383	127
g30958-t1_chr6	PMERF49	ERF	757	193
g2755-t1_chr7	PMERF55	ERF	755	251
g29086-t1_chr1	PMDREB1	DREB	251	83
g14999-t1_chr2	PMDREB4	DREB	9383	622
g287-t1_chr3	PMDREB8	DREB	624	102
g19580-t1_chr4	PMDREB12	DREB	515	136
g6051-t1_chr5	PMDREB13	DREB	842	194
g2756-t1_chr7	PMDREB14	DREB	929	309
g11052-t1_chr1	PMAP2.1	AP2	7678	378
g118-t1_chr2	PMAP2.4	AP2	3203	346
g49654-t1_chr3	PMAP2.7	AP2	3633	420
g1923-t1_chr4	PMAP2.8	AP2	1148	342
g7153-t1_chr5	PMAP2.10	AP2	2018	348
g3768-t1_chr6	PMAP2.16	AP2	3915	646
g12698-t1_chr7	PMAP2.18	AP2	7012	263
g2789-t1_chr6	PMRAV1	RAV	862	241
g24465-t1_chr5	PMERF.SOLOIST	SOLOIST	362	119

Table-2 *Summary of number of AP2 superfamily genes in different crops*

	Pearlmillet	Arabidopsis	Maize	Foxtailmillet	Rice	*Vitis vinifera*	*Poplus trichocarpa*	cauliflower	Sugarcane
ERF subfamily	62	65	153	90	79	73	91	91	101
DREB subfamily	15	57	89	48	52	36	71	55	59
AP2 family	19	18	46	28	26	18	26	15	43
RAV family	2	6	4	5	7	4	5	9	11
Soloist	1	1	0	0	0	1	1	4	4
Total	99	147	292	171	164	132	194	174	218

Table-3 *Positions of AP2 superfamily genes on chromosomes in pearl millet*

Gene name	Gene Id	start position (Mbp)	stop position (Mbp)
Pm_AP2/ERF_g4292-t1_chr1	PMERF1	30.529166	30.52946
Pm_AP2/ERF_g118-t1_chr2	PMAP2.4	0.819482	0.822685
Pm_AP2/ERF_g287-t1_chr3	PMDREB8	2.018437	2.019061
Pm_AP2/ERF_g1923-t1_chr4	PMAP2.8	14.249439	14.25059
Pm_AP2/ERF_g3012-t1_chr5	PMERF44	16.806244	16.80663
Pm_AP2/ERF_g2789-t1_chr6	PM RAV 1	15.669253	15.67012
Pm_AP2/ERF_g2755-t1_chr7	PMERF55	17.565863	17.56662

Table-4 *Characterization of amino acid sequences of AP2 superfamily transcription factors*

Gene Id	Amino acid residues	Instability index	Aliphatic Index	Grand Average of hydropathicity (GRAVY)
PMERF1	98	42.33	101.43	0.389
PMERF2	216	66.05	47.6	-0.832
PMERF3	191	55.49	53.35	-0.274
PMERF4	277	57.51	70.87	-0.337
PMERF5	339	58.27	52.77	-0.714
PMERF6	281	55.13	61.25	-0.455
PMERF7	285	50.25	71.86	-0.685
PMERF8	131	30.09	36.16	-0.364
PMERF9	636	62.48	81.68	-0.483
PMERF10	305	61.12	54.85	-0.703

Materials and methods

Sequence retrieval and identification of AP2/ERF superfamily genes in pearlmillet

The genome information of pearlmillet was downloaded from NCBI [29]. The protein sequences of AP2/ERF superfamily genes in Arabidopsis, maize, sorghum and foxtail millet were retrieved from Plant Transcription Factor Database (PlantTFDB) [30]. These proteins were used as query to create local blast program to find AP2/ERF super family genes in pearlmillet with expected value ≤1E-5.

All the blast hits were searched for the AP2 domain in the Pfam database. Augustus, gene finding software was used to predict the genes and their protein sequences were used as query with Pfam AP2 accession number PF00847. All the non-redundant blast hits were verified using NCBI-CDD [31] with default parameters and cut off set to 0.01.

MSA and phylogenetic analysis of pearl millet AP2/ERF proteins

Multiple Sequence analysis was performed for all the protein sequences using ClustalOmega [32] with default parameters. All the results were visualised using MEGA 7 software and a neighbour-joining phylogenetic tree was constructed [33]. The evolutionary relationship between the genes were found out using phylogenetic tree with the parameters pair-wise deletion and 1000 bootstrap replications.

Conserved motif analyses

Multiple Em for Motif Elicitation (MEME) tool version 5.0.5 [34] was used to identify conserved motifs with the parameter: (1) the maximum number of motifs found was set as 25; (2) the number of occurrences of a single motif distributed among the sequences within the model was set to zero or one per sequence.

Chromosomal distribution, subcellular localisation, Gene Ontology (GO) annotation of AP2 superfamily genes

Augustus gene prediction results has revealed the exact positions of genes on the chromosomes and these are plotted used MapChart software [35]. The localisation of proteins was analysed using CELLO v.2.5: subCELlular LOcalization predictor [36]. The GO annotation of protein sequences was predicted using CELLO2GO [37].

Characterization of amino acid sequences and protein folding of AP2/ERF superfamily TFs

ExPASy-ScanProSite [38] was used to calculate the amino acid composition of predicted proteins. Physical and chemical parameters of proteins like Aliphatic index, instability index, Grand Average of Hydropathicity were calculated using ExPASy-ProtParam [39]. The folding states of AP2 superfamily transcription factor proteins were predicted using FoldIndex programme [40].

Secondary structure prediction and Homology modelling of proteins

CFSSP, an online secondary structure prediction server was used to predict the regions like alpha helix, beta sheet and turns from the amino acid sequence [41]. Tertiary structure of protein was predicted using ExPASy-SWISSMODEL [42] based on homology modelling. Templates were selected based on GMQE and QSQE and models were built.

Results

Identification of AP2 genes in pearl millet

A total of 99 AP2/ERF superfamily genes were identified as putative genes in pearl millet and all of them were renamed and were presented in the [Table-1]. According to the [12] classification 19 genes belonged to AP2 family (PMAP2.1 to PMAP2.19) were predicted to encode proteins with two repeated AP2 domains and two RAV family genes (PMRAV1 and PMRAV2) were predicted to encode one AP2 DNA binding domain and a B3 domain. Seventy-seven genes were predicted to encode ERF family with one AP2 domain and was subdivided into ERF and DREB subfamilies. ERF subfamily (PMERF1 to PMERF62) has sixty-two genes encoding proteins which had Alanine and Aspartic acid at 14th and 19th positions and the DREB subfamily (PMDREB1 to PMDREB15) has fifteen genes encoding proteins with valine and glutamic acid at respective positions. Remaining gene was assigned as soloist having single AP2 domain as it has similarity with At4g13040 [20]. [Table-2] showed the number of AP2 superfamily genes identified in other crops like Arabidopsis, maize, sorghum etc. The gene length varied from 251bp (PMDREB1) to 12116 bp (PMERF36) and the amino acid residues ranges from 75 aa (PMERF32) to 1251 aa (PMAP2.3). At the whole genome level AP2 super family genes account for about 0.2566% (99/38579) smaller than rice (0.4315), poplar (0.4390), Arabidopsis (0.5481) and maize (0.6952).

Phylogenetic analysis

A neighbour-joining phylogenetic tree was constructed for the identified putative AP2 gene sequences encoded by protein sequences [Fig-1]. These sequences were classified into three larger clades with mixed families. Group I have ERF subfamily along with RAV family sequences. Group II has shared relationship between ERF subfamily and AP2 family. Group III has relationship between ERF and DREB subfamilies.

Chromosomal distribution and subcellular localization of AP2 superfamily genes in pearlmillet

All the genes were mapped on the seven chromosomes which revealed that there was an uneven distribution of genes in the genome. Among all the 99 genes, chromosomes about 19% (19) of genes were located on chromosome 2 followed by chromosomes 1and 3 with about 17% of identified genes. Chromosome five has 13% of genes and chromosome 7 had 12% of genes of AP2 family. Only 9% of genes were located on chromosome six. The precise position of genes (in Mbp) was given in the [Table-3]. The genes were spread all over the third and fourth chromosomes whereas most of the genes appear as clusters on the lower end of the arm of the all-other chromosomes.

ERF subfamily genes were distributed all over the seven chromosomes and the DREB subfamily genes were found over all chromosomes except chromosome six. RAV family contain only two genes and both of them were located on chromosome six. The subcellular localization of all the 99 AP2 superfamily genes was predicted and the analysis has shown that 79% of predicted genes were found in nucleus followed by chloroplast with 7.7% of genes. Cytoplasm and mitochondria have equal number of genes and plasma membrane has 3.3% of genes. Only one gene is present in extracellular region. Location of seven genes was not found.

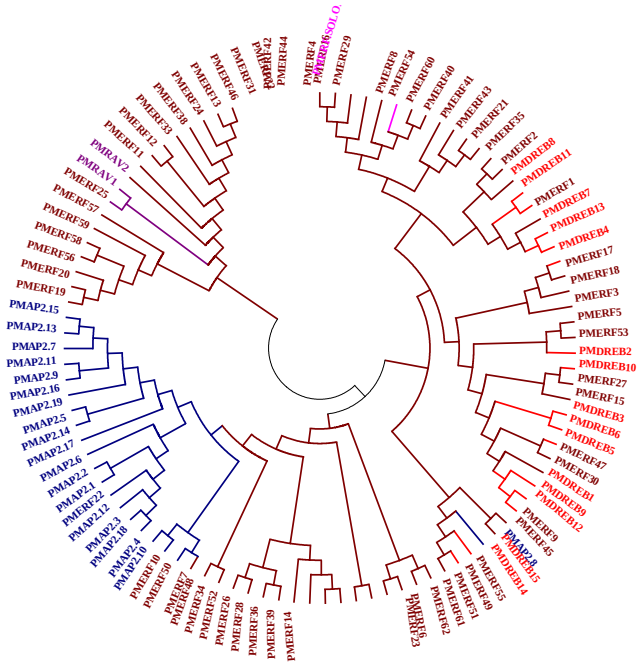

Fig-1 Phylogenetic tree of AP2 super family transcription factors constructed using MEGA 7 software

Gene ontology annotation of AP2 superfamily transcription factors in pearlmillet

Gene ontology studies were conducted using the predicted protein sequences of AP2 superfamily transcription factors revealed that all the proteins were shown up in molecular, biological and cellular processes. The molecular process analysis has revealed that all the identified proteins possess DNA-binding transcription factor activity. The analysis of biological processes has revealed that the pearlmillet super family proteins were signal transduction pathways, stress responses. These proteins participate in different metabolic processes of the plant like cellular nitrogen compound metabolic process, lipid metabolic process, carbohydrate metabolic process, secondary metabolic process and small molecule metabolic process. It is found that these transcription factors involve in diverse plant development processes like cell differentiation, biosynthetic processes, embryo development, reproduction, anatomic structure development. Cellular component analysis has shown that the localization of pearlmillet AP2 superfamily proteins in nucleus, chloroplast, mitochondria, plasma membrane and extracellular region. The detailed analysis was tabulated in [Table-4].

Characterization of aminoacid sequences of pearlmillet AP2 superfamily proteins

A comprehensive analysis was performed using ExPASy-Protparam to calculate the amino acid composition of predicted proteins and also physical and chemical parameters of proteins were calculated [Table-4]. A large amount of variation was observed between the subfamilies of AP2 superfamily. The length of amino acid residues varied from 75aa (PMERF32) to 1251aa (PMAP2.3). Instability index was less than 40 for PMERF14, PMERF28, PMERF29, PMERF30, PMERF54, PMERF57, PMDREB6, PMDREB8, PMDREB12, PMAP2.9, PMAP2.16, PMAP2.19, PMRAV1 and PMRAV2 indicating the good stability of proteins. Good stability index of proteins was observed in all the subfamilies except soloist.

The range of aliphatic index values varied from 34.25 (PMERF30) to 101.43 (PMERF1) and majority of predicted proteins have higher aliphatic index values indicating that the proteins are stable over wide temperature ranges. Grand Average of hydropathicity (GRAVY) values of proteins were calculated and all the proteins had negative values except PMERF1 which interpreted that all the proteins were hydrophilic except PMERF1 which is hydrophobic.

Conserved motifs of AP2 superfamily transcription factors

The conserved motifs were discovered using MEME tool. A total of 25 conserved motifs were identified given in [Table-5] and [Fig-2]. Motif 2 contained a WLG sequence and this motif was present in all the AP2 superfamily proteins. Motif 1 and Motif 6 were present in most of the proteins. The DREB subfamily transcription factors have motif 1, motif 2, motif 6, motif 11 and motif 18. ERF subfamily proteins have minimum two to maximum 6 motifs. Among 25 motifs detected ERF subfamily has about 18 motifs in total. RAV subfamily proteins have motif 2, motif 6, motif 11 and motif 24. Soloist has motif 1, motif 2 and motif 15. AP2 family proteins have minimum four to maximum ten motifs.

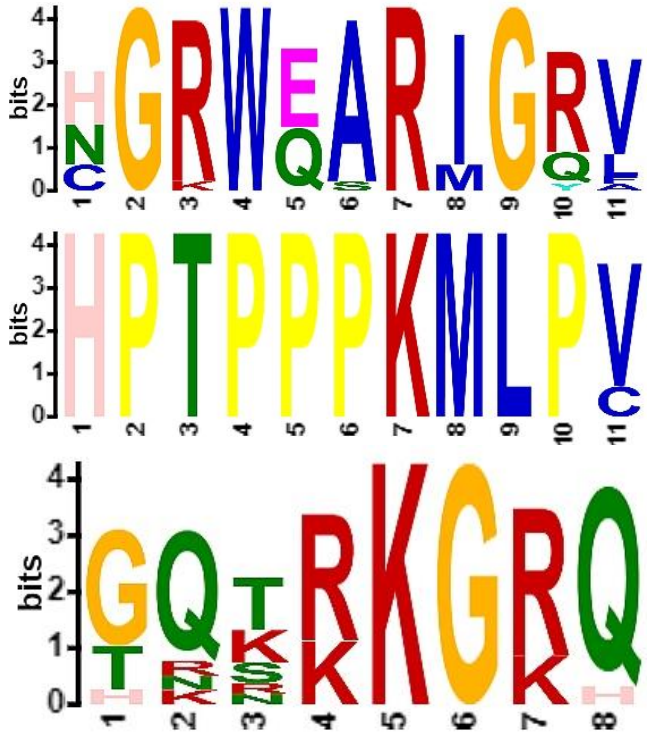

Fig-2 Conserved motifs analysed using MEME suite tools.

Table-5 Conserved motifs identified through MEME suite tool in pearl millet AP2 superfamily

GENE ID	START	STOP	MOTIF NUMBER
PMDREB1	59	79	MOTIF 1
PMDREB2	62	82	MOTIF 1
PMDREB2	83	110	MOTIF 2
PMDREB2	112	122	MOTIF 6
PMDREB3	38	53	MOTIF 2
PMDREB4	135	155	MOTIF 1
PMDREB4	156	183	MOTIF 2
PMDREB5	25	45	MOTIF 1
PMDREB5	46	73	MOTIF 2
PMDREB6	49	69	MOTIF 1
PMDREB6	70	97	MOTIF 2

Prediction of folding states of AP2 superfamily transcription factors of pearl millet

The folding states of proteins were predicted using FoldIndex program. The prediction results stated the number of disordered regions, longest disordered region and the number of disordered residues present in protein were presented in the [Table-6]. The number of disordered regions varied from one to twelve (PMAP2.3).

The protein PMAP2.3 has longest aminoacid residues among all the transcription factors predicted but the longest disordered region is present in PMERF25. The length of disordered regions varied from 7 to 206 but a greater number of proteins possess medium range of disordered regions. The number of disordered residues present in the disordered regions were also calculated. Highest number of disordered residues were found in PMDREB4 with nine disordered regions. Lowest number of disordered residues were observed in PMERF28. Six proteins PMERF1, PMERF21, PMERF32, PMERF34, PMERF46, PMERF57 all belonging to ERF subfamily were found to have no disordered regions and are expected to be completely folded and these are visualised in [Fig-3].

Fig-3 Prediction of the folding state of PMERF1, PMERF21, PMERF32, PMERF34, PMERF46, PMERF57 and PMERF2.3. Positive and negative numbers represent ordered and nonordered protein, respectively. Amino acids being ordered and nonordered regions are shown in red and green characters, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Homology modelling of pearl millet AP2/ERF proteins

Three-dimensional protein models were constructed using homology modelling for all the proteins. Most of proteins were found with similar templates like ATERF1, Ethylene Responsive Element Binding factor 1 and few proteins were found with no templates. No potential active binding sites were observed in the constructed models. The 3-D models of selected proteins were shown in the [Fig-4].

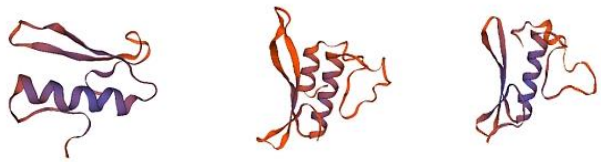

Fig-4 Tertiary structure of Pearl millet AP2/ERF proteins. a. PMERF1 b. PMAP2.1 c. PMRAV1

Discussion

Structural and functional developments observed in model plants like Arabidopsis and Rice has revealed that several hundreds to thousands of genes have involved in the stress responses by activating in the downstream signalling pathways. AP2/ERF family genes encode transcriptional regulators involving in different cellular, biological and molecular pathways. Genome-wide analysis is required for the identification of these regulatory genes which is essential to understand the molecular pathways involved and also to create the transcriptional network frames.

Genome-wide analysis was performed for identification of AP2 super family genes encoding proteins in foxtail millet and observed that these genes were upregulated in drought and salinity stress [22]. Sorghum was also studied for these transcription factors which revealed that 52 genes encode for DREB elements and the remaining 106 code for ERF subfamily proteins [43]. About forty ERF genes among them exhibited differential accumulation to drought and heat stress. Zhang *et al* (2022) [44] has demonstrated that the AP2 super family genes were involved in abiotic stress tolerance.

Rashid *et al* (2012) [45] has conducted genome-wide analysis in rice and also studied on syntenic relations between monocots and eudicots. The entire analysis has revealed that 170 AP2/ERF genes were identified and maximum synteny was observed on chromosome 3. Macrosynteny between rice and Arabidopsis has showed that eleven homologs/orthologs loci were found in both the genomes.

Table-6 *Protein folding characteristics of AP2 superfamily transcription factors in pearl millet*

Gene name	Gene Id	Family	Residues	Unfoldability	Charge	Phobic	Number Disordered Regions	Longest Disordered Region	Number Disordered Residues
Pm_AP2/ERF_g29086-t1_chr1	PMDREB1	ERF	83	0.057	0.036	0.447	1	32	32
Pm_AP2/ERF_g32600-t1_chr1	PMDREB2	ERF	236	0.072	0.03	0.45	3	91	129
Pm_AP2/ERF_g32904-t1_chr1	PMDREB3	ERF	159	-0.003	0.013	0.417	4	48	108
Pm_AP2/ERF_g14999-t1_chr2	PMDREB4	ERF	622	-0.109	0.072	0.4	9	167	504
Pm_AP2/ERF_g15932-t1_chr2	PMDREB5	ERF	158	0.092	0.063	0.469	3	49	89
Pm_AP2/ERF_g31531-t1_chr2	PMDREB6	ERF	163	0.026	0.037	0.436	1	73	73
Pm_AP2/ERF_g34062-t1_chr2	PMDREB7	ERF	213	0.039	0.038	0.441	5	44	72
Pm_AP2/ERF_g287-t1_chr3	PMDREB8	ERF	102	0.04	0.029	0.438	3	33	45
Pm_AP2/ERF_g37902-t1_chr3	PMDREB9	ERF	136	0.089	0.044	0.461	2	44	50
Pm_AP2/ERF_g43489-t1_chr3	PMDREB10	ERF	287	0.009	0.021	0.424	6	63	215

Conclusion

The AP2 superfamily transcription factors are the regulators involving in the plant growth, flowering, development and different stress responses. These transcription factors are involved in signal transduction pathways and activate many downstream genes. Importance of these transcription factors has made them study and explore extensively in different crops which was not explored in pearl millet. After a comprehensive study, a total of 99 AP2 superfamily transcription factors were identified which were classified based on the number of AP2 domains. Presence of AP2 domain was confirmed using Pfam database and NCBI-CDD. Gene ontology studies were performed and the biological, molecular and cellular roles of proteins in the plants were analysed and found that most of the proteins were involved in different stress response pathways. The entire *in silico* prediction analysis might provide the basic understanding of AP2 super family in pearl millet. The identified putative candidate genes can be explored further to understand their biological roles.

Application of research: This entire research acts as a basic platform for the analysis on transcription factors directly or indirectly associated to drought tolerance. The identified AP2 superfamily genes may be used for developing elite varieties that could be grown in dryland areas.

Research Category: Molecular Biology and Biotechnology

Abbreviations:
AP2/EREBP: APETALA2/ethylene response element-binding protein
DRE: Dehydration responsive element, CRT: C-repeat element
TF: Transcription factor, NCBI: National Centre for Biotechnology Information

Acknowledgement / Funding: Authors are thankful to Department of Molecular Biology and Biotechnology, SVU Agricultural College, Tirupathi, 517502, Acharya N. G. Ranga Agricultural University, Lam, 522034, Guntur, Andhra Pradesh, India

****Principal Investigator or Chairperson of research: Dr. J V Ramana**
University: Acharya N. G. Ranga Agricultural University, Lam, 522034, India
Research project name or number: Research station study

Author Contributions: All authors equally contributed

Author statement: All authors read, reviewed, agreed and approved the final manuscript. Note-All authors agreed that- Written informed consent was obtained from all participants prior to publish / enrolment

Study area / Sample Collection: SVU Agricultural College, Tirupathi, 51750

Cultivar / Variety / Breed name: Pearl Millet

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.
Ethical Committee Approval Number: Nil

References

[1] Sehgal D., Rajaram V., Armstead I.P., Vadez V., Yadav Y.P., Hash C.T. and Yadav R.S. (2012) *BMC Plant Biology*, 12(1), 9.
[2] Vadez V., Hash T., Bidinger F. and Kholova J. (2012) *Frontiers in Physiology*, 3(386).
[3] Singh P., Boote K.J., Kadiyala M.D.M., Nedumaran S., Gupta S.K., Srinivas K. and Bantilan M.C.S. (2017) *Science of the Total Environment*, 601-602,1226-37.
[4] Wang W., Vinocur B. and Altman A. (2003) Planta, 218, 1-14.
[5] Wang H., Wang H., Shao H. and Tang X. (2016) *Frontiers in Plant Sciences*, 7, 67.
[6] Gong Z., Xiong L., Shi H., Yang S., Herrera-Estrella L.R., Xu G., Chao D.Y., Li J., Wang P.Y. and Qin F. (2020) *Science China Life Sciences*, 63(5), 635-74.
[7] Kizis D., Lumbreras V. and Pages M. (2001) FEBS Letters, 498, 187-189.
[8] Yamaguchi-Shinozaki K. and Shinozaki K. (2006) Annual Review of Plant Biology, 57, 781-803.
[9] Jofuku K.D., Den Boer B.G., Van Montagu M. and Okamuro, J.K. (1994) *The Plant Cell*, 6(9), 1211-1225.
[10] Lata C., Yadav A. and Prasad M. (2011) *In: Shanker A. and Venkateshwarulu B. (eds) Abiotic Stress Response in Plants, INTECH Open Access Publishers*, 269-296.
[11] Mizoi J., Shinozaki K. and Yamaguchi-Shinozaki K. (2012) *Biochimica et Biophysica Acta*, 1819, 86-96.
[12] Sakuma Y., Liu Q., Dubouzet J.G., Abe H., Shinozaki K. and Yamaguchi-Shinozaki K. (2002) *Biochemical and biophysical research communications*, 290(3), 998-1009.
[13] Lata C. and Prasad M. (2011) *Journal of experimental botany*, 62, 4731-4748.
[14] Dong L., Cheng Y., Wu, J., Cheng Q., Li W., Fan S., Jiang L., Xu Z., Kong F., Zhang D. and Xu P. (2015) *Journal of experimental botany*, 66(9), 2635-2647.
[15] El Ouakfaoui S., Schnell J., Abdeen A., Colville A., Labbe H., Han S., Baum B., Laberge S. and Miki B. (2010) *Plant Molecular Biology*, 74(4-5), 313-26.
[16] Je B.I., Piao H.L., Park S.J., Park S.H., Kim C.M., Xuan Y.H., Park S.H., Huang J., Do Choi Y. and An G. (2010) *Plant Cell*, 2(6), 1777-91.
[17] Li C.W., Su R.C., Cheng C.P., Sanjaya Y.S.J., Hsieh T.H., Chao T.C. and Chan M.T. (2011) *Plant Physiology*, 156(1), 213-27.
[18] Zhou M.L., Tang Y.X. and Wu Y.M. (2012) *Current Bioinformatics*, 7(3), 324-332.
[19] Yan H.W., Hong L., Zhou Y.Q., Jiang H.Y., Zhu S.W., Fan, J. and Cheng B.J. (2013) *Genetics and Molecular Res.*, 12(2), 2038-2055.
[20] Nakano T., Suzuki K., Fujimura T. and Shinshi H. (2006) Plant physiology, 140(2), 411-432.
[21] Zhuang J., Chen J.M., Yao Q.H., Xiong F., Sun C.C., Zhou X.R., Zhang J. and Xiong A.S. (2011) *Mol. biology reports*, 38(2), 745-753.
[22] Lata C., Mishra A.K., Muthamilarasan M., Bonthala V.S., Khan Y. and Prasad M. (2014) *PLoS One*, 9(11), e113092.
[23] Zhang G., Chen M., Chen X., Xu Z., Guan S., Li L.C., Li A., Guo J., Mao L. and Ma Y. (2008) *Journal of experimental botany*, 59(15), 4095-4107.
[24] Xu W., Li F., Ling L. and Liu A. (2013) *BMC genomics*, 14(1), 1-15.
[25] Li H., Wang Y., Wu M., Li L., Li C., Han Z., Yuan J., Chen C., Song W. and Wang C. (2017) *Frontiers in plant science*, 8, 946.

[26] Zhuang J., Cai B., Peng R.H., Zhu B., Jin X.F., Xue Y., Gao F., Fu X.Y., Tian Y.S. Zhao W. and Qiao Y.S. (2008) *Biochemical and biophysical research communications*, 371(3), 468-474.
[27] Licausi F., Giorgi F.M., Zenoni S., Osti F., Pezzotti M. and Perata P. (2010) *BMC genomics*, 11(1), 1-16.
[28] Zhao J., Li W., Guo C. and Shu Y. (2018) *Biotechnology & Biotechnological Equipment*, 32(2), 303-308.
[29] Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Ostell J., Pruitt K.D., Sayers E.W. (20158) *Nucleic Acids Res*., 46(D1), D41-D47.
[30] Jin J.P., Tian F., Yang D.C., Meng Y.Q., Kong L., Luo J.C. and Gao G. (2017) *Nucleic Acids Research*, 45(D1), D1040-D1045.
[31] Shennan Lu., *et al*. (2020) *Nucleic Acids Research*, 48(D1)265-8.
[32] Madeira F., Pearce M. and Tivey A.R.N. (2022) *Nucleic Acids Research*, gkac240.
[33] Sudhir Kumar., Glen Stecher., Michael Li., Christina Knyaz. and Koichiro Tamura. (2018) *Molecular Biology and Evolution*, 35, 1547-1549.
[34] Timothy L., Bailey. and Charles Elkan. (1994) *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, 28-36, AAAI Press, Menlo Park, California*, 1994.
[35] Voorrips R.E. (2002) *The Journal of Heredity*, 93(1), 77-78.
[36] Yu C.S., Chen Y.C., Lu C.H. and Hwang J.K. (2006) *Proteins: Structure, Function and Bioinformatics*, 64, 643-651.
[37] Yu C.S., Cheng C.W., Su W.C., Chang K.C., Huang S.W., Hwang J.K. and Lu C.H. (2014) *PLoS ONE*, 9(6), e99368.
[38] De Castro E., Sigrist C.J.A., Gattiker A., Bulliard V., Langendijk-Genevaux P.S., Gasteiger E., Bairoch A. Hulo N. (2006) *Nucleic Acids Research*, W362-5.
[39] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch A. (2005) *(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press* (2005), 571-607.
[40] Jaime Prilusky., Clifford E. Felder., Tzviya Zeev-Ben-Mordehai., Edwin H. Rydberg., Orna Man., Jacques S. Beckmann., Israel Silman. and Joel L. (2005) *Bioinformatics*, 21(16), 3435-3438.
[41] Ashok Kumar T. (2013) *Wide Spectrum: Research Journal*, 1(9), 15-19.
[42] Waterhouse A., Bertoni M., Bienert S., Studer G., Tauriello G., Gumienny R., Heer F.T., de Beer T.A.P., Rempfer C., Bordoli L., Lepore R. and Schwede T. (2018) *Nucleic Acids Research*, 46, W296-W303.
[43] Mathur S., Priyadarshini S.S., Singh V., Vashisht I., Jung K.H., Sharma R., Sharma M.K. (2020) *3 Biotech*, 10(3), 139.
[44] Zhang J., Liao J. (2022) BMC Genomics, 23, 125.
[45] Rashid M., Guangyuan H., Guangxiao Y., Hussain J. and Xu Y. (2012) *Evolutionary Bioinformatics*, 8, EBO-S9369.