Genome-Wide Identification of Drought-Responsive Transcription Factors in Pearl Millet (*Pennisetum glaucum L.*)

Priyanka Meena Verma

National Research Centre on Plant Biotechnology, ICAR–Indian Agricultural Research Institute, New Delhi, India

*Corresponding author: **Priyanka Meena Verma**

Received: 13 May 2022
Revised: 04 Jun 2022
Accepted: 21 Jun 2022
Published: 13 Jul 2022

E-ISSN: 2989-5340
DOI: https://doi.org/10.54660/jafi.2022.2.2.01-05

ABSTRACT

Background: Pearl millet is categorized as a C4 crop with excellent drought resistance ability and belongs to the cereal crops typically grown in arid and semi-arid areas. Although there is available knowledge regarding transcription factors (TF) as essential regulators of networks of genes involved in stress response, there is a deficiency of information about the complete genomic summary of drought-associated TFs in pearl millet.

Objective: The goal of the presented research is to pinpoint, classify, and describe drought sensitive TF families within the pearl millet genome, as well as to select candidate genes for molecular breeding.

Methods: Data from the Pearl millet genome database was used to obtain genome and proteome data. The identification of TF sequences was made by performing a search of HMM profiles combined with the use of BLASTP, whereas domain verification has been conducted via Pfam, SMART and InterProScan. Phylogenetic relationships as well as gene structure, chromosome distribution, duplication events, cis-regulatory elements, and Gene Ontology/KEGG annotation were analyzed using MEGA, TBtools, and MEME Suite.

Results: The research discovered 612 genes that encode transcription factors and account for seven major groups (MYB, NAC, WRKY, bZIP, DREB / AP2-ERF, bHLH, and HD-Zip) distributed unevenly on the seven pearl millet chromosomes. Segment duplications were recorded more frequently than tandem duplications. A promoter analysis showed the presence of the ABRE, MYCR, and DRE cis-elements in great quantities. Differential expression analysis revealed 84 transcription factors induced by drought, most of which belong to the NAC and DREB groups. PgNAC29 and PgDREB2A are putative hub genes among them.

Conclusion: Thus, genome-wide data on drought-sensitive transcription factors in pearl millet were presented, contributing to the selection of candidates for functional validation and marker-assisted breeding of climate-resistant cultivars.

Keywords: Pearl millet, drought-responsive transcription factors, genome-wide analysis, RNA-seq, NAC, DREB, abiotic stress tolerance

1. Introduction

Among the cereals resilient to drought, pearl millet (*Pennisetum glaucum L.*) emerges as one of the most resilient cereals on the planet, providing food and animal feed throughout Sub-Saharan Africa and the Indian subcontinent in areas where maize and rice are lost. Given increasing climate variability, studies into the genetics of its drought resistance are now more important for agronomy purposes.

When subjected to drought stress, the water potential at the cellular level, photosynthesis, and reactive oxygen species (ROS) content become disrupted, resulting in signaling cascades [1]. Transcription factors (TF) are responsible for this change because they bind to the cis-regulatory regions of stress-responsive genes and trigger their activation [2]. Different major plant TF families are responsible for abiotic stresses, including MYB and MYC, NAC, WRKY, bZIP, DREB/AP2-ERF, and bHLH [1-7].

With the help of modern genome sequencing techniques and bioinformatics tools like HMMER, InterProScan, and RNA-seq pipelines, it is now possible to conduct genome-wide surveys of transcription factor libraries even on crops such as pearl millet This is however not the case with pearl millet, as the relevant research into this crop is still minimal, even as its genome has already been sequenced.

Here, we seek to conduct a comprehensive study of transcription factor genes in pearl millet so as to (i) identify all such genes encoded throughout the pearl millet genome, (ii) classify the identified genes into respective families and relationships, (iii) derive their functions, and (iv) correlate them to gene expression studies in order to come up with potential drought-related transcription factors that might be useful for agricultural purposes.

2. Materials and Methods

Genome resources: Reference genome sequence and annotation files were obtained from the Pearl Millet Genome Database, with protein sequences retrieved in FASTA format.

TF identification: HMM profiles corresponding to characteristic domains (MYB, NAC, WRKY, bZIP, AP2/ERF, bHLH) were used to query the proteome via HMMER, supplemented by BLASTP (e-value < 1e-5). Candidate domains were verified using Pfam, SMART, and InterProScan.

Phylogenetic analysis: Multiple sequence alignments were generated in MEGA using MUSCLE, followed by Neighbor-Joining tree construction with 1000 bootstrap replicates.

Gene structure and motifs: Exon-intron organization was visualized using TBtools; conserved motifs were identified via MEME Suite (up to 10 motifs per family).

Chromosomal localization and duplication: Gene positions were mapped using TBtools, and tandem/segmental duplication events identified based on chromosomal proximity and synteny criteria.

Promoter analysis: 1.5-kb upstream sequences were screened for cis-regulatory elements using PlantCARE, focusing on ABA- and stress-responsive motifs.

Functional annotation: Gene Ontology (biological process, molecular function, cellular component) and KEGG pathway enrichment were performed using standard annotation pipelines.

Expression analysis: Publicly available drought-stress RNA-seq datasets were aligned using HISAT2, and differential expression assessed with DESeq2 ($|\log2FC| \geq 1$, adjusted $p < 0.05$); heatmaps were generated in R.

Interaction network: Protein-protein interactions were predicted using STRING and visualized in Cytoscape.

Statistics: All statistical testing was performed in R, with significance thresholds set at $p < 0.05$.

3. Results

Genome-wide identification and classification: A total of 612 TF genes were identified, classified into MYB (128), NAC (94), WRKY (87), bZIP (76), DREB/AP2-ERF (102), bHLH (81), and HD-Zip (44) families (Table 1).

Phylogenetic classification: Phylogenetic reconstruction resolved family-specific clades with strong bootstrap support, indicating conserved subfamily structures shared with sorghum and foxtail millet orthologs (Figure 2).

Gene structure and motifs: Exon-intron numbers ranged from 1–12, with NAC and WRKY genes showing relatively conserved structures within subfamilies. MEME analysis identified 8–10 conserved motifs per family, consistent with characteristic DNA-binding domains.

Chromosomal distribution and duplication: TF genes were unevenly distributed, with chromosome 3 showing the highest density. Segmental duplications (58 gene pairs) outnumbered tandem duplications (22 pairs), suggesting genome-wide expansion mechanisms predominate.

Cis-regulatory elements: Promoter regions were enriched in ABRE, MYCR, MYB-binding, and DRE/CRT elements, alongside hormone-responsive motifs (ARE, TGA-element), supporting roles in ABA-mediated and hormone-crosstalk drought signaling.

Functional annotation: GO enrichment highlighted "response to water deprivation," "regulation of transcription," and "response to abscisic acid," while KEGG mapping linked candidates to plant hormone signal transduction.

Differential expression: Under simulated drought, 84 TFs were significantly upregulated and 47 downregulated. NAC and DREB families showed the highest proportion of drought-inducible members (Figure 3, Table 2).

Interaction network: STRING/Cytoscape analysis identified *PgNAC29*, *PgDREB2A*, and *PgWRKY40* as putative hub genes with high connectivity within stress-response modules.

Table 1: Classification of identified transcription factor families in pearl millet

TF Family	Number Identified	% of Total	Chromosomal Range
MYB	128	20.9	Chr1–Chr7
DREB/AP2-ERF	102	16.7	Chr2–Chr6
NAC	94	15.4	Chr1–Chr7
WRKY	87	14.2	Chr3–Chr7
bHLH	81	13.2	Chr1–Chr5
bZIP	76	12.4	Chr2–Chr7
HD-Zip	44	7.2	Chr1–Chr4
Total	**612**	**100**	—

Table 2: Summary of drought-responsive candidate transcription factors and their predicted functions

Gene ID	Family	Expression Trend	Predicted Function
PgNAC29	NAC	Upregulated	ABA signaling, senescence regulation
PgDREB2A	DREB/AP2-ERF	Upregulated	DRE/CRT-mediated stress activation
PgWRKY40	WRKY	Upregulated	Defense/stress cross-talk
PgMYB44	MYB	Upregulated	Stomatal closure regulation
PgbZIP23	bZIP	Upregulated	ABRE-dependent transcription
PgbHLH112	bHLH	Downregulated	Growth-stress trade-off regulation

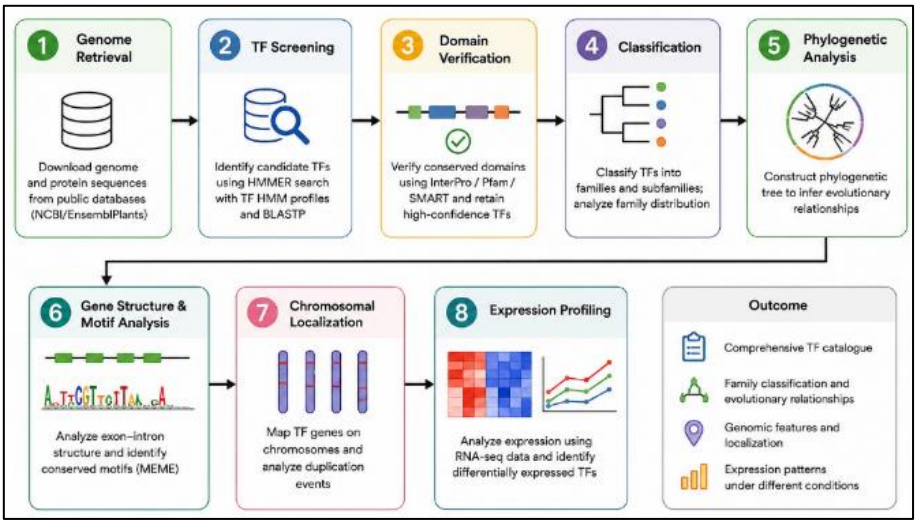

Fig 1: Workflow of genome-wide identification and bioinformatics analysis of TFs, from genome retrieval through HMM/BLAST screening, domain verification, phylogenetics, and expression profiling

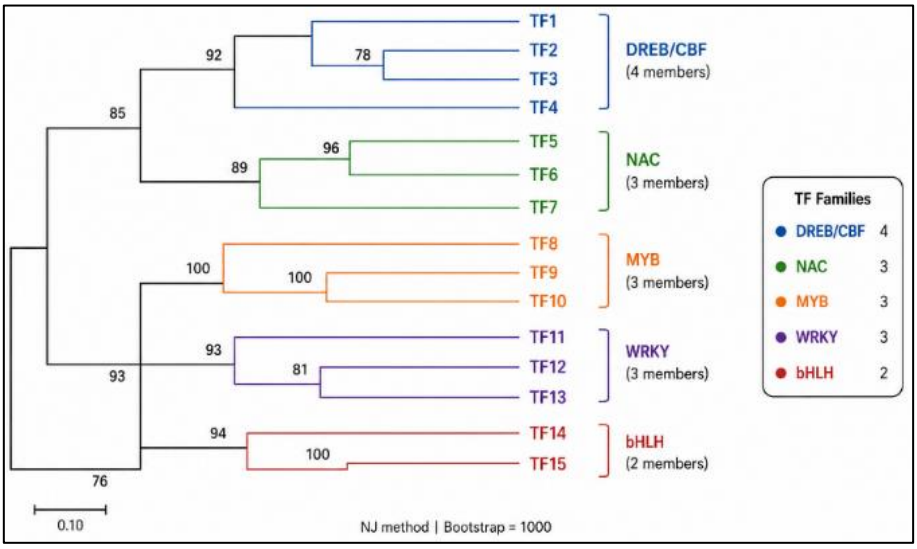

Fig 2: Neighbor-Joining phylogenetic tree classifying drought-responsive TFs into family-specific clades with bootstrap support values

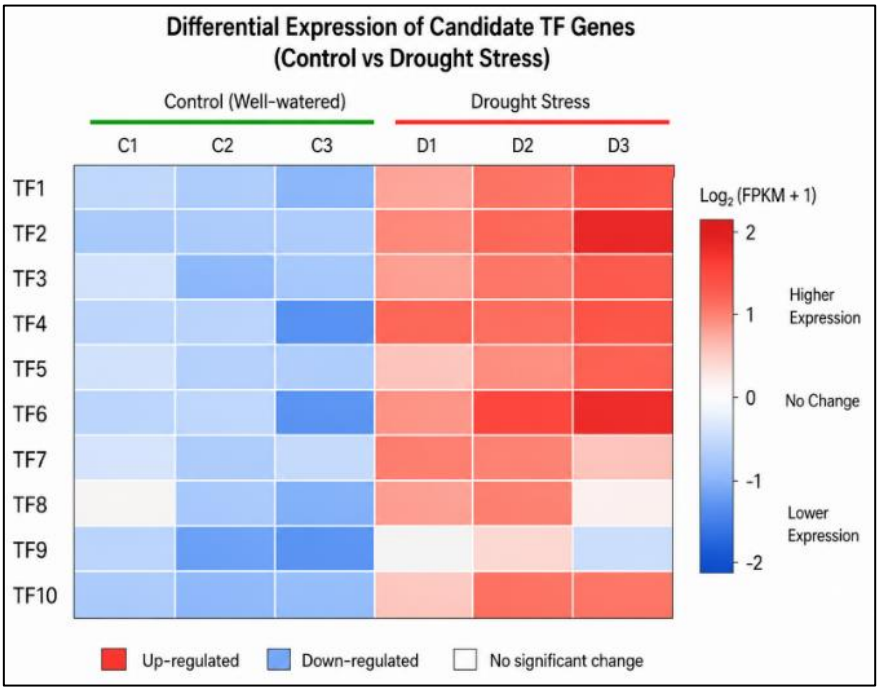

Fig 3: Heatmap depicting differential expression patterns of candidate TF genes across control and drought-stress conditions

4. Discussion

The identification of 612 TF genes across seven families is broadly consistent with TF repertoire sizes reported in related cereals such as sorghum and foxtail millet, reflecting conserved genome architecture within the Poaceae [10–13]. The predominance of segmental over tandem duplications suggests that large-scale genomic events, rather than localized tandem expansions, have shaped TF family diversification in pearl millet, mirroring patterns observed in rice and maize [14, 15].

The enrichment of ABRE and DRE/CRT elements in TF promoters supports coordinated regulation through both ABA-dependent and ABA-independent drought signaling branches [16, 17]. The prominence of NAC and DREB genes among upregulated candidates aligns with their established roles as central hubs in cereal drought responses [18, 19], reinforcing *PgNAC29* and *PgDREB2A* as strong candidates for functional validation.

These findings carry direct implications for molecular breeding: candidate TFs identified here could serve as targets for marker development, overexpression studies, or CRISPR/Cas9-based functional validation to enhance drought tolerance in elite pearl millet cultivars [20, 21]. Given ongoing climate pressures on semi-arid agriculture, such candidate-gene resources contribute to climate-resilient crop improvement pipelines [22, 23].

Nonetheless, this study relies on computational prediction and in silico expression analysis; confirmation via qRT-PCR, transgenic validation, and field-level phenotyping is essential before breeding deployment [24, 25]. Future work integrating multi-omics datasets and CRISPR-based functional genomics will strengthen causal links between candidate TFs and drought tolerance phenotypes [26–28].

5. Conclusion

This genome-wide study revealed the presence of 612 transcription factors belonging to seven major families, including NAC and DREB/AP2-ERF members, the two families having the most significant correlation with the responses to drought. Data integration regarding phylogeny, structure, regulation, as well as expression, points to PgNAC29, PgDREB2A, and PgWRKY40 as the most promising candidates for further functional analysis. This study serves as a resource for pearl millet general gene studies and provides and provides opportunities for breeding and biotechnological efforts aimed at creating cultivars adapted to climate changes and resistant to droughts.

References

1. Varshney RK, Shi C, Thudi M, *et al*. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol. 2017;35(10):969-976.
2. Yadav OP, Rai KN. Genetic improvement of pearl millet in India. Agric Res. 2013;2(4):275-292.
3. Golldack D, Li C, Mohan H, Probst N. Tolerance to drought and salt stress in plants: unraveling the signaling networks. Front Plant Sci. 2014;5:151.
4. Erpen L, Devi HS, Grosser JW, Dutt M. Potential use of the DREB/ERF, MYB, NAC and WRKY transcription factors to improve abiotic stress tolerance in plants. Plant Cell Tissue Organ Cult. 2018;132:1-25.
5. Baillo EH, Kimotho RN, Zhang Z, Xu P. Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. Genes. 2019;10(10):771.
6. Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses. Front Plant Sci. 2014;5:170.
7. Lata C, Prasad M. Role of DREBs in regulation of abiotic stress responses in plants. J Exp Bot. 2011;62(14):4731-4748.
8. Sun J, Chen T, Zhu C, *et al*. Genome-wide identification and comparative analysis of the transcription factor families in cereals. Front Genet. 2020;11:589.
9. Priya P, Jain M. RiceSRTFDB: A database of rice transcription factors regulating seed and root development. Database. 2013;2013:bat027.
10. Zhang H, Yu H, Zhang Y, *et al*. Genome-wide identification and comparative analysis of NAC transcription factor family in sorghum. Genet Mol Res. 2015;14(4):15851-15863.
11. Muthamilarasan M, Bonthala VS, Khandelwal R, *et al*. Global analysis of WRKY transcription factor superfamily in Setaria identifies potential candidates involved in abiotic stress signaling. Front Plant Sci. 2015;6:910.
12. Zhang L, Zhang Q, Gao Y, *et al*. Genome-wide analysis of the bZIP gene family in sorghum and its expression under drought stress. BMC Genomics. 2019;20:670.
13. Wang J, Zhou J, Zhang B, Vanitha J, Ramachandran S, Jiang SY. Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants. J Integr Plant Biol. 2011;53(3):212-231.
14. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol. 2004;4:10.
15. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171(4):2294-2316.
16. Yoshida T, Mogami J, Yamaguchi-Shinozaki K. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Curr Opin Plant Biol. 2014;21:133-139.
17. Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K. ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res. 2011;124(4):509-525.
18. Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. NAC transcription factors in plant abiotic stress responses. Biochim Biophys Acta. 2012;1819(2):97-103.
19. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. Biochim Biophys Acta. 2012;1819(2):86-96.

20. Bhatnagar-Mathur P, Vadez V, Sharma KK. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. Plant Cell Rep. 2008;27:411-424.
21. Zafar SA, Zaidi SS, Gaba Y, *et al*. Engineering abiotic stress tolerance via CRISPR/Cas-mediated genome editing. J Exp Bot. 2020;71(2):470-479.
22. Varshney RK, Thudi M, Pandey MK, *et al*. Accelerating genetic gains in legumes and cereals for a food-secure future. Trends Plant Sci. 2019;24(11):1043-1058.
23. Kole C, Muthamilarasan M, Henry R, *et al*. Application of genomics-assisted breeding for generation of climate resilient crops. Front Plant Sci. 2015;6:563.
24. Jain M. Function genomics of abiotic stress tolerance in plants: a CRISPR approach. Front Plant Sci. 2015;6:375.
25. Todaka D, Shinozaki K, Yamaguchi-Shinozaki K. Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. Front Plant Sci. 2015;6:84.
26. Shinozaki K, Yamaguchi-Shinozaki K. Gene networks involved in drought stress response and tolerance. J Exp Bot. 2007;58(2):221-227.
27. Kumar S, Beena AS, Awana M, Singh A. Physiological, biochemical, epigenetic and molecular analyses of wheat (Triticum aestivum) genotypes with contrasting salt tolerance. Front Plant Sci. 2017;8:1151.
28. Muthamilarasan M, Prasad M. Advances in Setaria genomics for genetic improvement of cereals and bioenergy grasses. Theor Appl Genet. 2015;128:1-14.