

***In Silico* Study of bZIP TFs Family Genes for Osmotic Stress Responsive in Genomic Elements Pigeon Pea (*Cajanus cajan* L.)**

Arti¹, Shobhit Kumar Srivastava² and Manoj Kumar Sharma³

¹Department of Botany, Asian International University, Imphal West, Manipur, INDIA.

²Associate Professor, Department of Botany, Asian International University, Imphal West, Manipur, INDIA.

³Department of Bioinformatics, Janta Vedic College, Baraut, Baghpat, Uttar Pradesh, INDIA.



www.jrasb.com || Vol. 3 No. 5 (2024): October Issue

Received: 13-10-2024

Revised: 16-10-2024

Accepted: 28-10-2024

ABSTRACT

The bZIP transcription factor (TF) gene family plays a pivotal role in regulating plant responses to abiotic stresses and developmental processes. In pigeon pea (*Cajanus cajan* L.), a legume adapted to semi-arid environments, 69 putative bZIP genes (CcbZIP1 to CcbZIP69) were identified through homology-based BLASTP searches and Pfam verification for the bZIP_1 domain (PF00170). These genes exhibit significant structural diversity, with coding sequence lengths ranging from 273 to 2,235 base pairs, encoding proteins of 90 to 744 amino acids. Functional domain analysis confirmed the bZIP_1 domain in 33 proteins, with additional domains suggesting functional diversification. Subcellular localization analysis revealed predominant nuclear localization, consistent with their role in transcriptional regulation. Motif analysis identified conserved patterns critical for DNA binding and dimerization, while physicochemical properties indicated diverse protein architectures suited for stress responses. Phylogenetic analysis grouped the 32 bZIP_1-containing proteins into three clades, reflecting evolutionary divergence. These findings highlight the CcbZIP family's role in pigeon pea's stress resilience and provide a foundation for molecular breeding to enhance crop tolerance.

Keywords- bZIP, Abiotic stress, bZIP_1 domain.

I. INTRODUCTION

Pigeon pea (*Cajanus cajan* L.), a diploid legume ($2n=2x=22$, genome size ~ 833 Mb), is a vital crop in semi-arid tropics, valued for its protein-rich seeds and resilience to abiotic stresses such as drought and salinity (Varshney et al., 2012). Basic leucine zipper (bZIP) transcription factors (TFs) are key regulators of plant stress responses and developmental processes, mediating abscisic acid (ABA) signaling and gene expression under environmental challenges (Jakoby et al., 2002). In model plants like *Arabidopsis thaliana* (78 bZIP genes) and rice (89 genes), bZIP TFs regulate processes ranging from stress tolerance to seed development (Dröge-Laser et al., 2018; Nijhawan et al., 2008). In pigeon pea, the identification of 69 putative

bZIP genes (CcbZIP1 to CcbZIP69) suggests a robust regulatory network, potentially driven by gene duplication events, as observed in other legumes like soybean and pea (Jia et al., 2016; Li et al., 2023).

This study characterizes the CcbZIP gene family through homology-based identification, functional domain analysis, subcellular localization, motif analysis, physicochemical properties, and phylogenetic relationships. The bZIP_1 domain, essential for DNA binding and dimerization, was confirmed in 33 proteins, with additional domains indicating functional diversity. Subcellular localization analysis highlights nuclear roles, while motif and phylogenetic analyses reveal conserved and divergent features. Physicochemical properties suggest adaptations for stress tolerance. These findings aim to elucidate the

molecular basis of pigeon pea's resilience and inform breeding strategies for improved stress tolerance and productivity.

II. METHODOLOGY

Gene Identification

A total of 69 putative bZIP TF genes (CcbZIP1 to CcbZIP69) were identified in the pigeon pea genome using a homology-based approach. BLASTP searches were conducted with an *Arabidopsis thaliana* bZIP protein (UniProtKB ID: A0A059VE58) as the query, applying an e-value threshold of $1e-10$. Candidate genes were verified for the bZIP_1 domain (PF00170) using the Pfam database and the Conserved Domain Database (CDD) (Marchler-Bauer et al., 2017). Gene annotations and visualizations were generated using TBtools (Chen et al., 2020).

Functional Domain Analysis

The 69 CcbZIP proteins were analyzed for conserved domains using CDD with default parameters, focusing on the bZIP_1 domain. Additional domains were identified to assess functional diversity. TBtools was used to visualize domain architecture alongside gene sequences, facilitating comparative analysis of structural organization.

Subcellular Localization Analysis

Subcellular localization of CcbZIP proteins was predicted using WoLF PSORT, which employs machine learning to assess signal peptides and amino acid composition (Horton et al., 2007). Localization scores were generated for compartments including the nucleus, cytoplasm, and chloroplasts. A heatmap of nuclear-localized proteins was created using TBtools.

Motif Analysis

Conserved motifs in 32 bZIP_1-containing CcbZIP proteins were identified using the MEME Suite, with parameters set to detect 10–20 motifs of 6–50 residues (Bailey et al., 2009). Motifs were aligned with the bZIP_1 domain to assess functional relevance. Phylogenetic clustering was performed using MEGA 6.0, and results were visualized with TBtools.

Physicochemical Properties

Physicochemical properties, including molecular weight (Mw), isoelectric point (pI), aliphatic index (AI), instability index, and grand average of hydropathicity (GRAVY), were computed for the 32 bZIP_1-containing proteins using ProtParam on the ExPASy server (Gasteiger et al., 2005).

Phylogenetic Analysis

Full-length amino acid sequences of the 32 bZIP_1-containing CcbZIP proteins were aligned using ClustalW ver. 2.1 (Thompson et al., 1994). A phylogenetic tree was constructed in MEGA 6.0 using

the neighbor-joining method, Poisson substitution model, uniform rates, and pairwise deletion, with 1,000 bootstrap iterations (Tamura et al., 2013). The tree was visualized using the Interactive Tree of Life (iTOL) tool (Letunic & Bork, 2021).

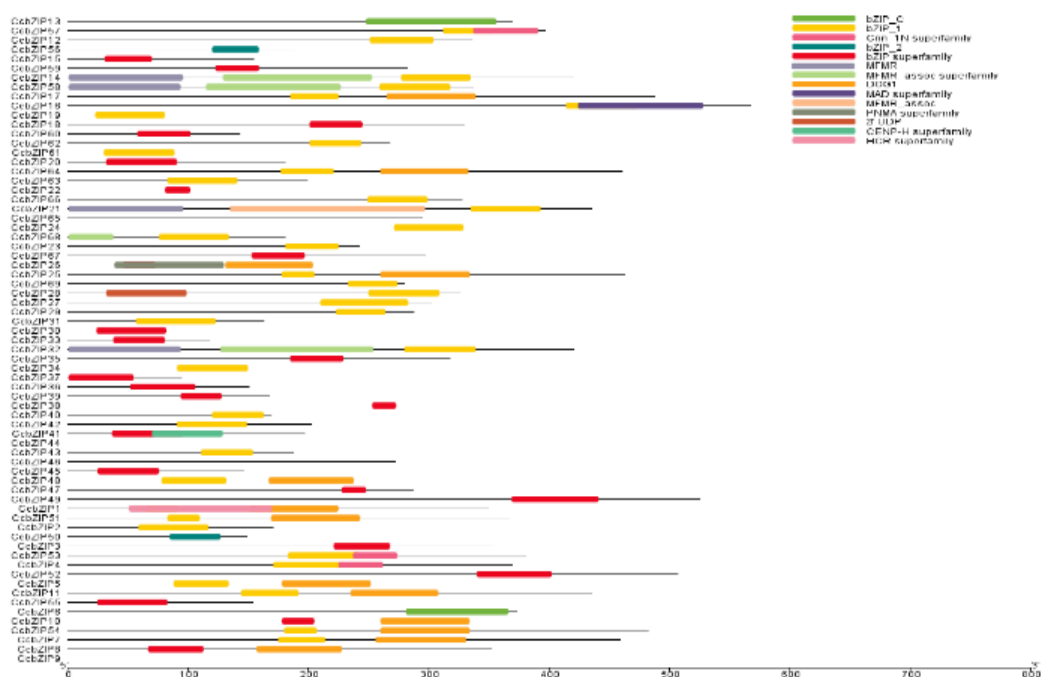
III. RESULTS AND DISCUSSION

Gene Identification

A total of 69 putative bZIP TF genes (CcbZIP1 to CcbZIP69) were identified in the pigeon pea genome, comparable to 77 in tobacco and 87 in pea (Zhou et al., 2022; Li et al., 2023). Coding sequence lengths ranged from 273 bp (CcbZIP44) to 2,235 bp (CcbZIP24), encoding proteins of 90 to 744 amino acids, with an average of ~890 bp and 296 amino acids. This structural diversity suggests functional specialization, with larger genes like CcbZIP24 potentially encoding complex proteins involved in intricate regulatory interactions, similar to *Arabidopsis* AtbZIP1 (Sun et al., 2012). Smaller genes like CcbZIP44 may have targeted roles, analogous to AtbZIP19 in zinc deficiency response (Assunção et al., 2010). The homology-based approach, using an *Arabidopsis* bZIP protein linked to salt and osmotic stress, likely enriched for stress-responsive genes, relevant for pigeon pea's semi-arid habitat. Comparative studies with soybean (GmbZIP44) suggest CcbZIP genes may enhance drought and salt tolerance (Jia et al., 2016). The variation in gene size may reflect alternative splicing or exon/intron differences, as seen in *Arabidopsis* (Dröge-Laser et al., 2018). Future phylogenetic and genomic mapping studies could reveal duplication patterns, as observed in wheat (Agarwal et al., 2019).

Functional Domain Analysis

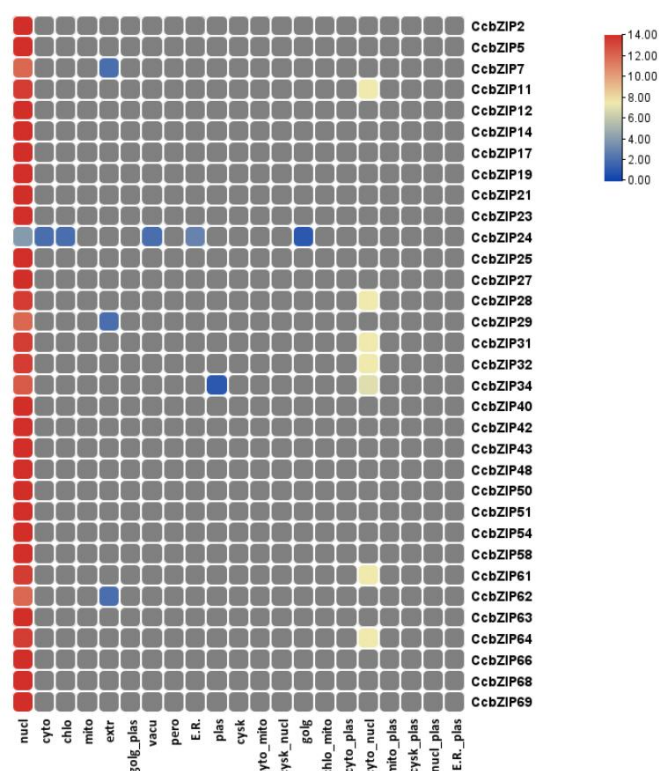
CDD analysis confirmed the bZIP_1 domain in 33 CcbZIP proteins (e.g., CcbZIP2, CcbZIP5, CcbZIP69), essential for DNA binding and dimerization (Jakoby et al., 2002). These proteins likely form a specialized subgroup optimized for transcriptional regulation under stress, similar to *Arabidopsis* AtbZIP17 (Liu et al., 2014). Additional domains in other CcbZIP proteins suggest roles in protein-protein interactions or post-translational modifications, enhancing regulatory complexity (Dröge-Laser et al., 2018). TBtools visualization revealed consistent bZIP_1 domain presence and variable auxiliary domains, indicating both conserved and divergent functions. This diversity aligns with rice bZIPs, where additional domains modulate stress responses (Nijhawan et al., 2008). The streamlined architecture of bZIP_1-only proteins suggests rapid transcriptional activation, while complex domain structures may support multifaceted stress responses.



Subcellular Localization Analysis

WoLF PSORT analysis revealed strong nuclear localization for most CebZIP proteins (e.g., CebZIP2, CebZIP5, score 14), consistent with their role as TFs (Jakoby et al., 2002). Proteins like CebZIP11 and CebZIP28 showed dual nuclear and cytoplasmic localization (scores 13.5 and 7.5), suggesting dynamic shuttling, as seen in Arabidopsis bZIPs involved in ABA signaling (Dröge-Laser et al., 2018). CebZIP24's diverse

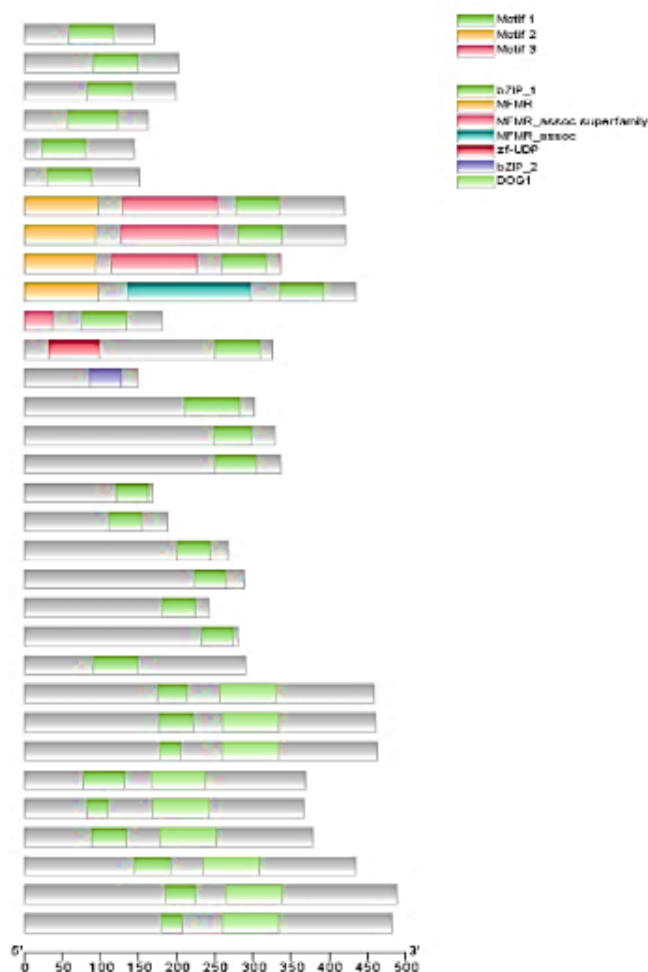
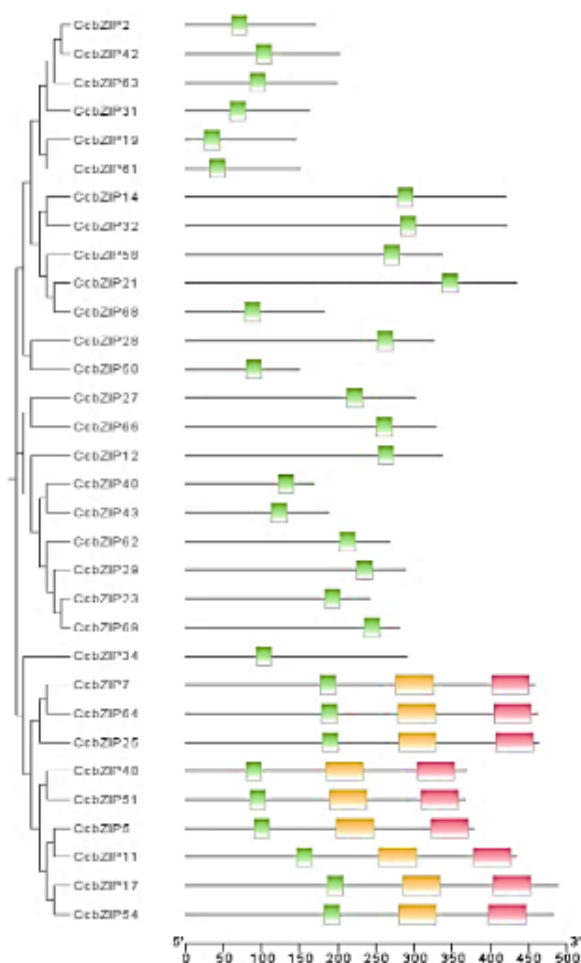
localization (nucleus: 4, cytoplasm: 2, chloroplast: 2) indicates potential multifunctionality, possibly akin to rice bZIPs with cytoplasmic roles (Nijhawan et al., 2008). These predictions align with Arabidopsis and rice studies, where nuclear bZIPs regulate stress-responsive genes (Liu et al., 2014). The TBtools heatmap highlighted nuclear localization patterns, supporting their transcriptional roles in stress responses.



Motif Analysis

MEME Suite analysis identified conserved motifs in the 32 bZIP_1-containing CcbZIP proteins, including basic region and leucine zipper motifs critical for DNA binding and dimerization (Bailey et al., 2009). Additional motifs suggest roles in protein-protein interactions or regulatory functions, as seen in Arabidopsis (Dröge-Laser et al., 2018). Phylogenetic

clustering via MEGA 6.0 grouped proteins into clades, with TBtools visualization mapping motifs to gene sequences. These motifs align with stress-responsive bZIPs in rice, indicating conserved functions (Nijhawan et al., 2008). The motif distribution suggests functional specialization, with some proteins potentially modulating specific stress pathways.



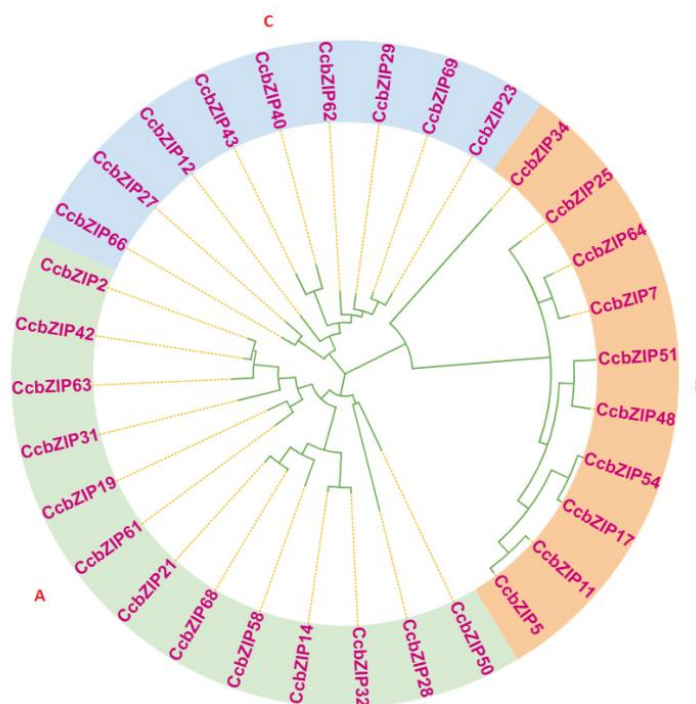
Physicochemical Properties

ProtParam analysis revealed molecular weights from 16,343.49 Da (CcbZIP19) to 53,580.15 Da (CcbZIP17), indicating diverse protein sizes. Isoelectric points ranged from 4.72 (CcbZIP40) to 10.02 (CcbZIP27), with acidic pI values favoring nuclear localization (Liu et al., 2014). Aliphatic indices (50.47–85.96) suggest thermostability, particularly for CcbZIP48 (85.96), suited for semi-arid conditions (Nijhawan et al., 2008). Negative GRAVY values (-1.074 to -0.38) indicate hydrophilicity, facilitating nuclear interactions. Most proteins were unstable (instability index >40), except CcbZIP64 (36.65), suggesting dynamic regulation (Gasteiger et al., 2005). These properties align with bZIP functions in stress

signaling, with outliers like CcbZIP24 warranting further study.

Phylogenetic Analysis

Phylogenetic analysis grouped the 32 bZIP_1-containing CcbZIP proteins into three clades: Clade A (CcbZIP2, CcbZIP42, etc.), Clade B (CcbZIP5, CcbZIP17, etc.), and Clade C (CcbZIP66, CcbZIP27, etc.). High bootstrap support (1,000 iterations) ensures robust clustering (Tamura et al., 2013). Clade A's diversity suggests broad stress roles, while Clade B's high aliphatic indices indicate stability (Nijhawan et al., 2008). Clade C's varied pI values suggest functional versatility (Dröge-Laser et al., 2018). The iTOL visualization highlighted motif conservation, supporting evolutionary divergence likely driven by gene duplication (Jakoby et al., 2002).



IV. CONCLUSION

The identification of 69 CcbZIP genes in pigeon pea highlights a robust regulatory network for stress responses and development. The bZIP_1 domain in 33 proteins, coupled with additional domains, suggests both specialized and multifaceted roles. Predominant nuclear localization aligns with transcriptional regulation, while motif and phylogenetic analyses reveal conserved and divergent features. Physicochemical properties indicate adaptations for stress tolerance, with diverse protein architectures supporting dynamic regulation. The phylogenetic division into three clades reflects evolutionary divergence, likely driven by duplication events. These findings underscore the CcbZIP family's role in pigeon pea's resilience and provide a framework for molecular breeding to enhance crop productivity in semi-arid environments.

FUTURE RECOMMENDATIONS

Future research should validate CcbZIP gene functions through expression profiling (e.g., qRT-PCR) under drought and salinity stress to identify key regulators, particularly in Clades A and B. Gene editing (e.g., CRISPR/Cas9) can confirm roles of proteins like CcbZIP2 and CcbZIP17. Mass spectrometry and isoelectric focusing should verify physicochemical properties. Comparative genomics with legumes like soybean could elucidate duplication patterns, while chromosomal mapping may reveal regulatory clusters. Investigating additional domains' roles in protein interactions or post-translational modifications could

uncover novel stress signaling mechanisms. Integrating these findings with agronomic data will support breeding of stress-resilient pigeon pea cultivars, enhancing food security in semi-arid regions.

REFERENCES

- [1] Agarwal, P., Baranwal, V. K., & Khurana, P. (2019). Genome-wide analysis of bZIP transcription factors in wheat and functional characterization of a TabZIP under abiotic stress. *Scientific Reports*, 9(1), 4608.
- [2] Alonso, R., Oñate-Sánchez, L., Weltmeier, F., Ehlert, A., Diaz, I., Dietrich, K., Vicente-Carbajosa, J., & Dröge-Laser, W. (2009). A pivotal role of the basic leucine zipper transcription factor bZIP53 in the regulation of Arabidopsis seed maturation gene expression based on heterodimerization and protein complex formation. *The Plant Cell*, 21(6), 1747–1761.
- [3] Assunção, A. G. L., Herrero, E., Lin, Y. F., Huettel, B., Talukdar, S., Smaczniak, C., ... & Aarts, M. G. M. (2010). Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proceedings of the National Academy of Sciences*, 107(22), 10296–10301.
- [4] Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., ... & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server issue), W202-W208.

- [5] Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13(8), 1194–1202.
- [6] Dröge-Laser, W., Snoek, B. L., Snel, B., & Weiste, C. (2018). The Arabidopsis bZIP transcription factor family—An update. *Current Opinion in Plant Biology*, 45, 36–49.
- [7] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (pp. 571-607). Humana Press.
- [8] Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, 35(Web Server issue), W585-W587.
- [9] Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., & Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends in Plant Science*, 7(3), 106–111.
- [10] Jia, B., Sun, M., Sun, X., Li, R., & Wang, X. (2016). Overexpression of *GmbZIP44* enhances drought and salt tolerance in transgenic Arabidopsis. *Plant Physiology and Biochemistry*, 108, 431–438.
- [11] Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296.
- [12] Li, M., Chen, Q., Dong, H., Zhang, S., & Huang, X. (2023). Genome-wide identification and expression analysis of the bZIP transcription factors, and functional analysis in response to drought and cold stresses in pear (*Pyrus bretschneideri*). *BMC Plant Biology*, 21(1), 583.
- [13] Li, X., Liu, C., Zhao, Z., Ma, D., & Zhang, J. (2024). Genome-wide identification, characterization and expression analysis of the bZIP transcription factors in garlic (*Allium sativum* L.). *Frontiers in Plant Science*, 15, 1365123.
- [14] Liu, J. X., Srivastava, R., & Howell, S. H. (2014). Stress-induced expression of bZIP transcription factors in Arabidopsis. *Plant Signaling & Behavior*, 9(10), e977426.
- [15] Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Bryant, S. H. (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(D1), D200–D203.
- [16] Nijhawan, A., Jain, M., Tyagi, A. K., & Khurana, J. P. (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiology*, 146(2), 333–350.
- [17] Sun, X., Xu, L., Wang, Y., Yu, R., Zhu, X., & Zhang, X. (2012). The Arabidopsis AtbZIP1 transcription factor is a positive regulator of plant tolerance to salt, osmotic, and drought stresses. *Journal of Plant Research*, 125(3), 429–438.
- [18] Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729.
- [19] Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680.
- [20] Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., Donoghue, M. T. A., Azam, S., Fan, G., Whaley, A. M., Farmer, A. D., Sheridan, J., Iwata, A., Tuteja, R., Penmetsa, R. V., Wu, W., Upadhyaya, H. D., Yang, S. P., Shah, T., ... Cook, D. R. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology*, 30(1), 83–89.
- [21] Wang, Y., Liu, Z., Wu, Z., Li, H., Wang, W., Cui, X., (2019). Genome-wide identification and expression analysis of GRAS family transcription factors in tea plant (*Camellia sinensis*). *Scientific Reports*, 9(1), 1–15.
- [22] Zhou, Y., Wang, X., Wang, Z., Yan, L., & Zhang, W. (2022). Genome-wide identification and expression analysis of the bZIP transcription factor family genes in response to abiotic stress in *Nicotiana tabacum* L. *BMC Genomics*, 23(1), 318.