The Journal of Animal & Plant Sciences, 31(4): 2021, Page: 1145-1159
ISSN (print): 1018-7081; ISSN (online): 2309-8694

COMPUTATIONAL ANALYSIS OF HOMEODOMAIN-LEUCINE ZIPPER (HD-ZIP) FAMILY OF TRANSCRIPTION FACTORS IN *ARABIDOPSIS THALIANA*

M. S. Javaid, U. A. Ashfaq, M. Ijaz, B. Naz, B. Saleem, A. Arshad and Mahmood-ur-Rahman*

Department of Bioinformatics & Biotechnology, GC University, Faisalabad-38000, Pakistan
*Corresponding Author's email: mahmoodansari@gcuf.edu.pk

ABSTRACT

Homeodomain-leucine Zipper (HD-Zip) proteins are a class of transcription factors having conserved homeodomain. Homeo box, conserved part of homeodomain consists of 56 residues and is involved in plant development. We retrieved protein sequences of all the 48 HD-Zip family members, performed genome-wide analysis, and grouped them into four subfamilies. A comprehensive genomic analysis including mapping of all the transcription factors on 5 chromosomes of *Arabidopsis thaliana*, multiple sequence alignment to find out the conserved domains, phylogenetic, and gene structure analysis by mapping introns and exons, promoter analysis by taking 1000bp upstream genomic sequence of all these transcription factors and motif analysis was performed in this study. The results helped us to understand the functional homology among HD-Zip transcription factors, comparison of the tree with motifs that further depends upon the number of exons and introns. This classification and analysis further categorized the transcription factors of 4 HD-Zip subfamilies into different classes and revealed a deep evolutionary relationship among them and thus helped to further explore their functioning.

Keywords: HD-Zip family, *Arabidopsis thaliana*, Homeo box, phylogenetic analysis, motif

https://doi.org/10.36899/JAPS.2021.4.0312 *Published online December 18, 2020*

INTRODUCTION

Many biological processes such as development, growth, cell division and response against stress or environmental stimuli in organisms/cells are controlled by transcription factors (Shen *et al.,* 2019). Transcription factors (TFs) bind to the DNA and either up-regulate or down-regulate gene expression at the transcriptional level (Todaka *et al.,* 2015). There are more than 60 families of TFs in plants having different functions. The homeo-box (HB) family of TFs has a sub-family of homeodomain-leucine-zipper (HD-Zip) proteins that have been reported in *Arabidopsis thaliana* and many other plants. The HB is an important transcription factor family consisting of the homeodomain (HD) which binds to DNA at a specific site and a leucine zipper that acts as a dimerization motif (Wang and Zhuang, 2019). The HD-Zip proteins play fundamental roles in various plant developmental processes, from cell formation pattern to specifications of cell types (Shao *et al.,* 2018). The HD-containing proteins were grouped into six classes including KNOX (KNOTTED1-like homeobox), HD-Zip (homeodomain-leucine-zipper), Bell (bell domain), PHD-Finger (plant homeodomain-finger), ZF-HD (zinc finger-homeodomain) and WOX (Wuschel-related homeobox). Out of these classes, the HD-Zip proteins are abundant in plants and perform important functions in different processes of development and growth (Chen *et al.,* 2014).

In Arabidopsis, the HD-Zip family has 48 members that are further divided into 4 sub-families based on their biological roles and other conserved motifs (Gong *et al.,* 2019). HD-Zip I, II, III and IV proteins preferentially bind with the CAAT(A/T)ATTG, CAAT(C/G)ATTG, GTAAT(G/C)ATTAC and CATT(A/T)AATG motifs respectively (Sessa *et al.,* 2018). The HD-Zip family has a highly conserved HD-domain and a less conserved leucine zipper (LZ) motif and four sub-families are differentiated on the basis of downstream sequence of LZ domain (Ding *et al.,* 2017). LZ domain is a dimmer of Basic-region leucine zipper (bZip) class of TFs (Vinson *et al.,* 1989). The bZip domain is 60-80 amino acids in length with conserved DNA binding basic-region and a diversified LZ region (Zhang *et al.,* 2014). The localization of the leucines is very important for the DNA binding to the proteins. In Arabidopsis, the HD-Zip I subfamily is comprised of 17 candidate genes that do not show response to ABA signaling, sugar signaling and environmental stresses but are crucial for plant de-etiolation and embryogenesis (Henriksson *et al.,* 2005). The HD-Zip II subfamily of Arabidopsis has 9 members and all of them consist of a cellular redox status perceptive CPSCE motif located in the downstream region of LZ domain and most of these genes respond to shading, light and auxin as disclosed by biochemical and genetic analyses (Turchi *et al.,* 2013).

HD-Zip III has 5 TFs that are basically regulatory elements, involved in various developmental processes like apical embryo development, the formation of shoot meristematic tissues, determining the polarity of different organs, vascular diversity and transference of auxins

(Wenkel *et al.,* 2007). The fourth sub-family, HD-Zip IV, also known as HD-Gl2 is a larger one and contains 16 candidate genes. The genetic analysis illustrated that proteins belong to this subfamily play critical roles in the formation of the trichome, root propagation, accumulation of anthocyanin and differentiation of epidermal cells in Arabidopsis (Chen *et al.,* 2017). The HD-Zip family has been reported in many studies and considerable work has been done on its sub-families, but the focus was not driven on genome-wide analysis. The purpose of this research was to study 48 HD-Zip TFs, their phylogenetic analysis, conserved domain analysis, promoter analysis, gene structure analysis, analysis of *cis*-regulatory elements and chromosomal mapping by using various bioinformatics tools in Arabidopsis to understand their functional dynamics.

MATERIALS AND METHODS

Chromosome mapping: The physical location of TFs of the HD-Zip family was demonstrated in *A. thaliana*. The chromosomal mapping of all the 48 members of the HD-Zip family on the different chromosomes of Arabidopsis was performed by the online tool, TAIR (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp) by using standard procedures.

Identification of conserved domains: The core of the Homeo-domain was studied in all the 48 genes of the HD-Zip family of *A. thaliana* including all the subfamilies. For this purpose, multiple sequence alignment of all the members was performed by Unipro UGENE v. 34 (http://ugene.net/) (Rose *et al.,* 2018). Then protein sequences of all the candidate genes were compared for identification of the whole domain.

Phylogenetic analysis: The phylogenetic analysis and multiple sequence alignment (MSA) of protein sequences of HD-Zip genes were performed by using MEGA 6.0 software (https://www.megasoftware.net/). HD-Zip genes were grouped into different sub-families depending upon their biological functions and domain structure. The parameters used for alignment were as follow, penalty for gap opening: 10; penalty for gap extension: 0.2; weight matrix for protein: gonnet; penalties for specific residues: on; hydrophilic penalties: on; separation distance for gap: 4; end of gap separation: off; negative matrix usage: off; cut-off for delay divergent: 30%. An un-rooted phylogenetic tree was built by neighbor-joining (NJ) method by using multiple sequence alignment.

Synteny analysis: The protein sequences of all the HD-Zip genes were retrieved from an online database of plant transcriptional factors (http://planttfdb.cbi.pku.edu.cn/) for Synteny analysis. The sequences were saved in the FASTA format. Then, Synteny analysis was performed to identify the conservation of homologous genes by using an online tool by following default parameters (http://tools.bat.infspire.org/circoletto/).

Gene structure analysis: TAIR (https://www.arabidopsis.org/) and "Database for Transcription Factors of Arabidopsis" (DATF) (http://datf.cbi.pku.edu.cn/) were used to fetch the information and annotated the gene structure, e.g. number of exons and introns. The Microsoft PowerPoint and excel sheets were used to analyse the gene structure.

Analysis of conserved motifs: MEME 4.9.1 software (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) was used for the identification of conserved motifs in different protein domains. The complete dataset (protein sequences) of 48 TFs of HD-Zip was copied in corresponding box having features/properties to specify the results were; number of repetitions – any, number of repetitions, number of motifs - 15, minimum motif width -10, maximum motif width-56.

Promoter analysis and determination of *cis*-regulatory elements: Promoter analysis was done by retrieving 1000bp upstream region of the transcription initiation site of all 48 genes from Phytozome 9.1 database (http://www.phytozome.net/). A complete list of *cis*-acting elements was obtained by using PLACE, an online tool (Higo *et al.,* 1999) (http://www.dna.affrc.go.jp/PLACE/). We chose common *cis*-regulatory elements in promoter sequences of all HD-ZIP genes. The location of *cis*-regulatory elements was mapped by using Microsoft PowerPoint.

RESULTS AND DISCUSSION

Chromosomal mapping: The expression of HD-Zip gene family is highly regulated and shows response to environmental stimuli, organ-specific signals, to light, and to the stress conditions. All candidates of HD-Zip family were examined for the presence of LZ and HD domain. This family is further sub-divided into four sub-families HD-Zip I to IV. The first and forth sub-family have high number of genes while second family has a minimum number of genes and third family also have very few genes in Arabidopsis (Sessa *et al.,* 2018). We retrieved 48 candidate genes of HD-Zip TFs present in Arabidopsis. Different variants of these TFs were excluded (only those TFs were studied that have "1" value after decimal point e.g. AT3G01470.1). They were mapped on all the 5 chromosomes of *A. thaliana* (**Figure-1)** using an online chromosomal mapping software from TAIR (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp) database. First chromosome had 11 genes where 6 and 5 genes were present on short and long arms, respectively. The second chromosome contained 9 genes; 1 and 8 genes on short and long arms, respectively. The third chromosome had 6 genes; having equal numbers of genes

on both arms. The fourth and fifth chromosomes had 11 genes where forth chromosome have 2 genes on its short arm and fifth chromosome has 4 genes on its short arm while 9 and 7 genes are present on the long arm of fourth and fifth chromosome, respectively (**Figure-1**).

Depending upon the information retrieved through 'Phytozome', the chromosomal location of all the 48 genes belonging to HD-Zip family of TFs was determined. The genes of HD-Zip family of TFs were randomly dispersed on all of the 5 chromosomes of *A. thaliana*. It was found that 11 genes of HD-Zip family were located on chromosome number 1, 4 and 5 each while 9 genes were present on chromosome number 2. Moreover, 6 genes were found on chromosome number 3. The above data indicated the uneven distribution of HD-Zip genes on Arabidopsis chromosomes (**Figure-1**). Our results showed the same pattern as of the distribution of Trithorax homolog genes in chickpea (Qasim *et al.,* 2018), WRKY TFs in Arabidopsis (Sultan *et al.,* 2016), Potassium transporter genes in chickpea (Azeem *et al.,* 2018) and WRKY TFs in chickpea (Waqas *et al.,* 2019).

Figure-1: The mapping of HD-Zip transcription factors on *A. thaliana* chromosomes. The numbers of chromosomes are shown at the top of each chromosome and transcription factors are mapped according to their positions. The length of each chromosome is indicating its size.

Identification of conserved domain: HD-Zip family members of Arabidopsis were aligned by using Unipro UGENE software and analysed for the conserved domain. The signature box named "WFQNrr" was identified by UGENE in which "WFQN" residues were highly conserved while "rr" residues were semi-conserved. When protein sequences of all the 48 genes of this family were compared for the identification of whole domain, the total of 44 residues before the WFQNrr and 6 residues after the WFQNrr were identified as conserved with few alternative residues in some genes. So, the Homeo-domain of 56 residues was identified (**Figure-2**). The conserved amino acids were not present on the same position in all genes. They were present in different positions but found as a box which shows presence of amino acids as a conserved box in the domain in all the genes.

Moreover, 44 residues at upstream side and 6 residues at downstream side of this box were identified as conserved in all genes with few exceptions and thus it was investigated that the complete domain was present in distributed form and on different positions in every gene. So, to draw a consensus sequence of the domain of all the genes as well as to identify the complete domain as a whole, we deleted the amino acid residues one by one from both; the upstream and the downstream side of WFQNrr from each gene to bring this box on the same position in

every gene. By this, WFQNrr were placed on the same position in every gene when there were only 44 residues before this box and 6 residues after this box. Thus, Homeo-domain of 56 residues was identified in all genes having WFQNrr amino acids on position number 45, 46, 47, 48, 49 and 50 respectively (**Figure 2**).

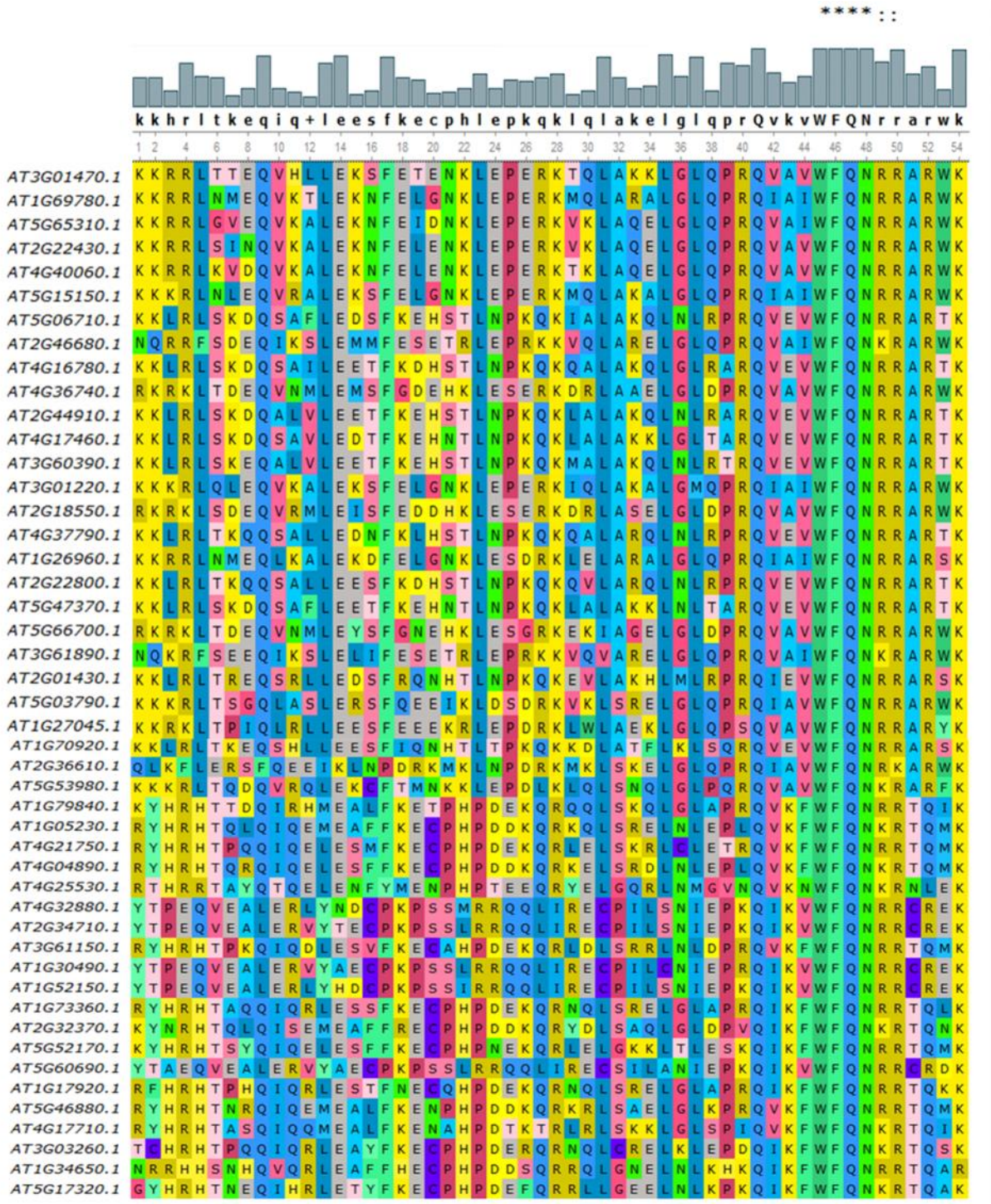

Figure-2: Multiple sequence alignment of HD-Zip transcription factors of *A. thaliana* by Unipro UGENE v. 34 software to identify the conserved domain. The signature sequence WFQNrr is identified and amino acids are shown in figure by " * " and " : " shape.

The proteins of HD-Zip I subfamily were identified via existence of homeo-domain which was closely related to leucine zipper motif. It was found in many studies that some proteins of the first sub-family

were intricated to the sucrose signalling and abscisic acid pathways. These proteins also played a critical role in the abiotic stress responses of the plants and cotyledon as well as involved in the leaf development and embryogenesis (Valdes *et al.,* 2012; Sessa *et al.,* 2018). Most of the genes of HD-Zip II sub-family were incriminated in phytochrome-mediated organ development e.g. leaf morphogenesis. Some genes were also responsive to the changes of light quality and auxin and shade avoidance as exhibit by biochemical and genetic analyses (Turchi *et al.,* 2013, 2015).

Phylogenetic analysis: A total of 4 sub families were observed in a phylogenetic tree. All the 4 sub families had conserved homeo-domain. HD-Zip subfamily I, II, III and IV have 17, 10, 5 and 16 genes respectively and each gene in each subfamily was categorized in different clades/classes according to their ancestor and evolutionary relationship as well as functions. It was concluded from the phylogenetic analysis that specific subfamilies of HD-Zip played specific role in specific type of responses in growth and development of *A. thaliana*. For phylogenetic analysis, an un-rooted tree was generated from all the 48 genes of HD-Zip members. First sub-family of HD-Zip TFs was divided into 7 clades that were named as clade alpha (α), beta one (β1), beta two (β2), phi (Φ), epsilon (ε), gamma (γ) and delta (δ). The HD-Zip II subfamily has 10 members that were divided into 5 clades and named as clade α, β, ε, γ and δ (**Figure-3**). Proteins of all the members of this family have a conserved set of cysteine residue on inner side and a leucine zipper motif at the outer side.

The HD-Zip III and IV subfamilies were characterized with the help of 2 more conserved domains; first domain was steroidogenic-acute-regulatory-protein related lipid-transfer (START) and second domain was START-adjacent-domain (SAD). Both subfamilies were differentiated by a fifth domain that was the C-terminal-MEKHLA-motif (Zhang *et al.,* 2014). This domain was present in third subfamily of HD-Zip proteins, but it was absent in forth subfamily. The third subfamily has 5 members that were divided into 3 clades and named as clade α, β and γ. The HD-Zip IV subfamily has 16 members that were divided into 6 clades and named as alpha (α), beta (β), epsilon (ε), gamma (γ), delta (δ) and zeta (ζ) (**Figure-3**).

Members of third sub-family were the key developmental regulators for the Arabidopsis shoot radial patterning and apical embryo. They were also involved in the development of lateral organ polarity, formation of shoot meristem, vascular differentiation and transportation of auxin (Qiu *et al.,* 2006). The proteins of the HD-Zip IV subfamily were involved to play role in epidermal processes, formation of trichome, accumulation of anthocyanin and development of root (Roodbarkelari and Groot, 2017). It helps to understand the evolutionary relationship among all the genes in more detail as each gene is further classified into different clades (Sessa *et al.,* 2018). This may help in further functional and comparative annotation of all the transcription factors of HD-Zip family.

Comparison of HD-Zip Genomic Sequences: Synteny relationship among all the 48 transcription factors of HD-Zip family of *A. thaliana* was observed in a circular figure as colour bars. Synteny analysis revealed tandem duplication and segmental duplication events in most of the transcription factors within the HD-Zip family. It was observed that transcription factors AT5G17320.1, AT5G52170.1, AT3G03260.1, AT2G32370.1, AT1G34650.1, AT1G52150.1, 0890 and 0333 have no synteny relationship at all. It was confirmed by synteny analysis of HD-Zip family where duplication events were shown by colour ribbons. Transcription factor "AT5G53980.1" that fall in Φ clade of HD-Zip I subfamily has been observed to play a prominent role in these duplication events. It was also observed that some transcription factors were not involved in segmental and tandem duplication events and thus these factors have their unique sequences and belonged to HD-Zip III and IV subfamilies (**Figure-4**).

Synteny analysis showed the tandem, whole genome and segmental duplication events that have significance in the evolution of many organisms. Genome duplication events played a significant role in the expansion of family (Hofberger *et al.,* 2015) and in Arabidopsis there are three duplication events. Based on the comprehensive analysis of chromosomal mapping, structure analysis of genes, phylogenetic analysis and analysis of motifs, we observed segmental and tandem duplication in most of the transcription factors of all the 4 subfamilies of HD-Zip transcription factors family.

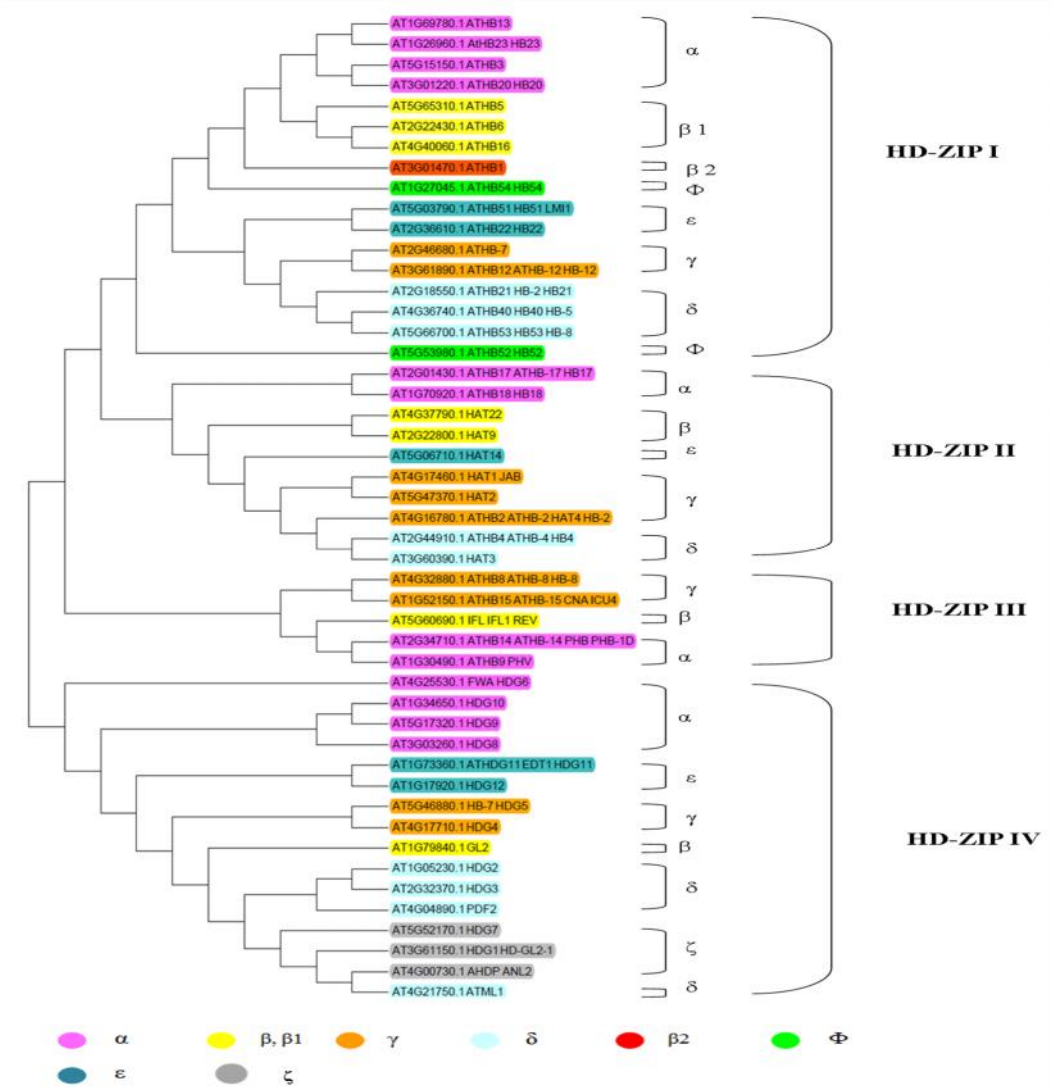

Figure-3: Rectangular phylogenetic tree of HD-Zip family of *A. thaliana* indicating the division of HD-Zip family into four sub families on the basis of evolution and functions.

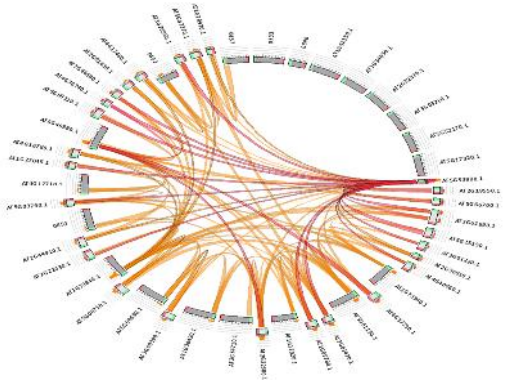

Figure-4: Comparison of gene sequences and identification of homologs of HD-Zip transcription factors by Synteny analysis in *A. thaliana*

Analysis of *cis*-regulatory elements in promoter region: Transcription factors regulate the expression of genes under biotic and abiotic stress conditions. The first step for promoter analysis was the retrieval of 1000 bp upstream to the start codon for each gene of HD-Zip family of *A. thaliana*. Then data were put into the PLACE database to find out the transcription factor binding sites (motifs) in the promoter region (**Supplementary Table-1**). We selected five most abundant *cis*-regulatory elements for each nucleotide sequence in all members of HD-Zip. Three out of five *cis*-regulatory elements named as CCAATBOX1, WORKY710S & DOFCOREZM were found in whole family, where they were present on positive strand of DNA. GATABOX was present almost 85% in promoter region of 48 genes while there was variability at 5th*cis*-regulatory element (**Figure-5)**, which was mostly present on negative strand of DNA.

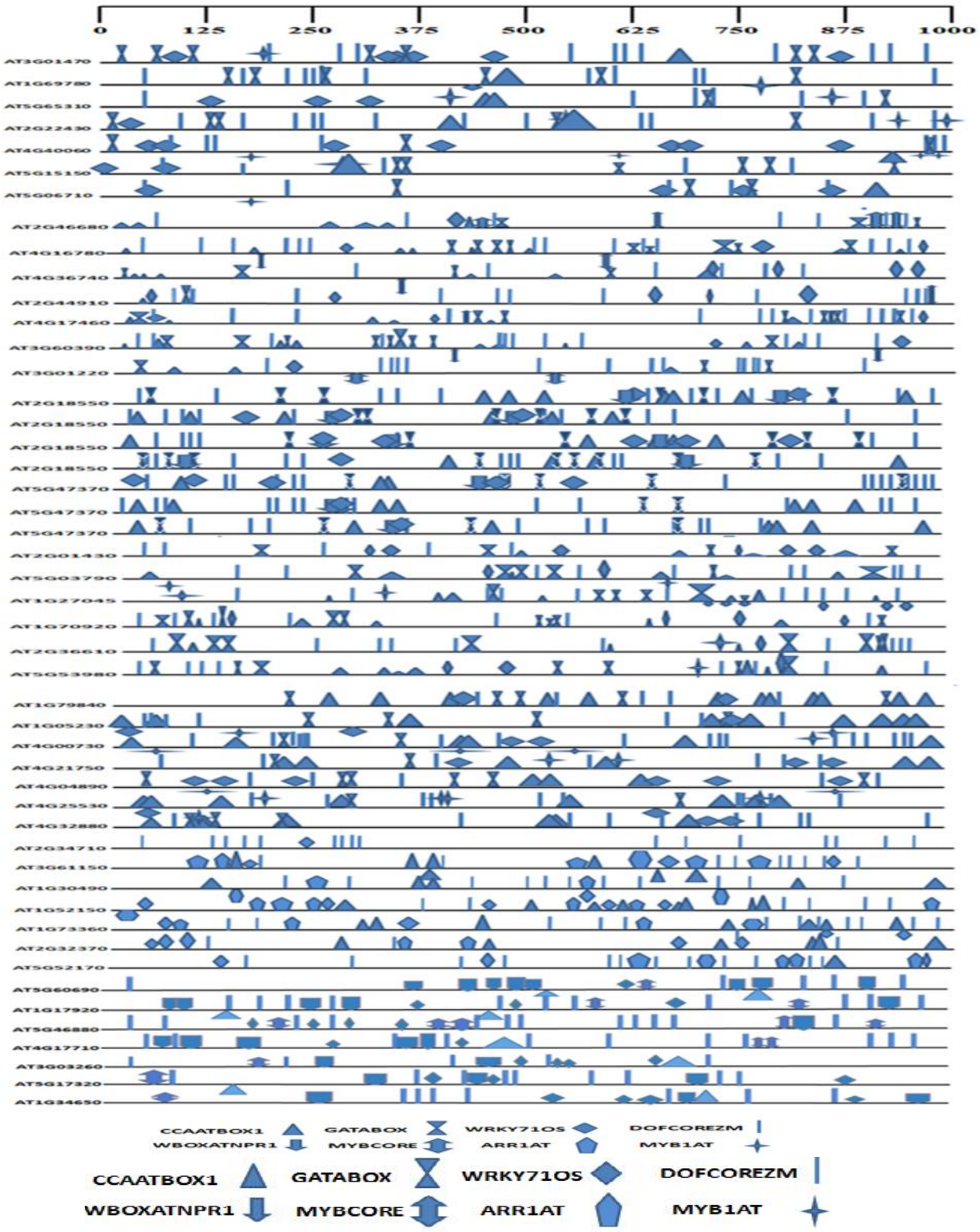

Figure-5: Promoter analysis; schematic representation of various PLACE-based motifs in 1kb promoter region of HD-Zip family. Most of motifs were present on + strand as shown in the figure. The motif codes and respective sequences are indicated in the figure.

CCAATBOX1, WORKY710S, GATABOX & DOFCOREZM were selected for analysis of *cis*-regulatory elements. CCAATBOX1 with site CCAAT was present in all members of HD-Zip family. The CCAAT Box regulates the flowering of A. *thaliana* and is mostly present in non-coding region of heat shock proteins of eukaryotes. DOFCOREZM with site AAAG involved in tissue specific expression. WORKY710S is a stress responsive *cis*-regulatory element and it was found in whole family of *Arabidopsis thaliana*. Promoter analysis of *cis*-regulatory element had revealed that almost every gene contains repetitive binding sites of transcriptional factors on the promoter region (**Figure-5**).

Identification of conserved motifs: Conserved motif analysis has revealed the homology between sub-families of HD-Zip TFs. As sub-families I and II have same number of motifs and exons while, sub-families III and IV have same number of motifs and exons. Thus, besides the phylogenetic analysis, motif analysis also showed the evolutionary relationship between the HD-Zip sub-families. Motif 1 links all the sub-families as it is conserved in all the transcription factors of all 4 sub-families (**Figure-6**). Based on presence of motifs, we concluded that genes belong to I and II family of HD-Zip TFs were closely related while members of subfamily III and IV also had homology. Motif 1 was conserved in 47 members. Majority of members of this family almost first 28 members were containing 3-4 exons and few of remaining members containing 7-10 exons while there were 2-3 exceptional genes containing 17-18 exons which is showing their evolutionary relatedness and differences (**Figure-6)**.

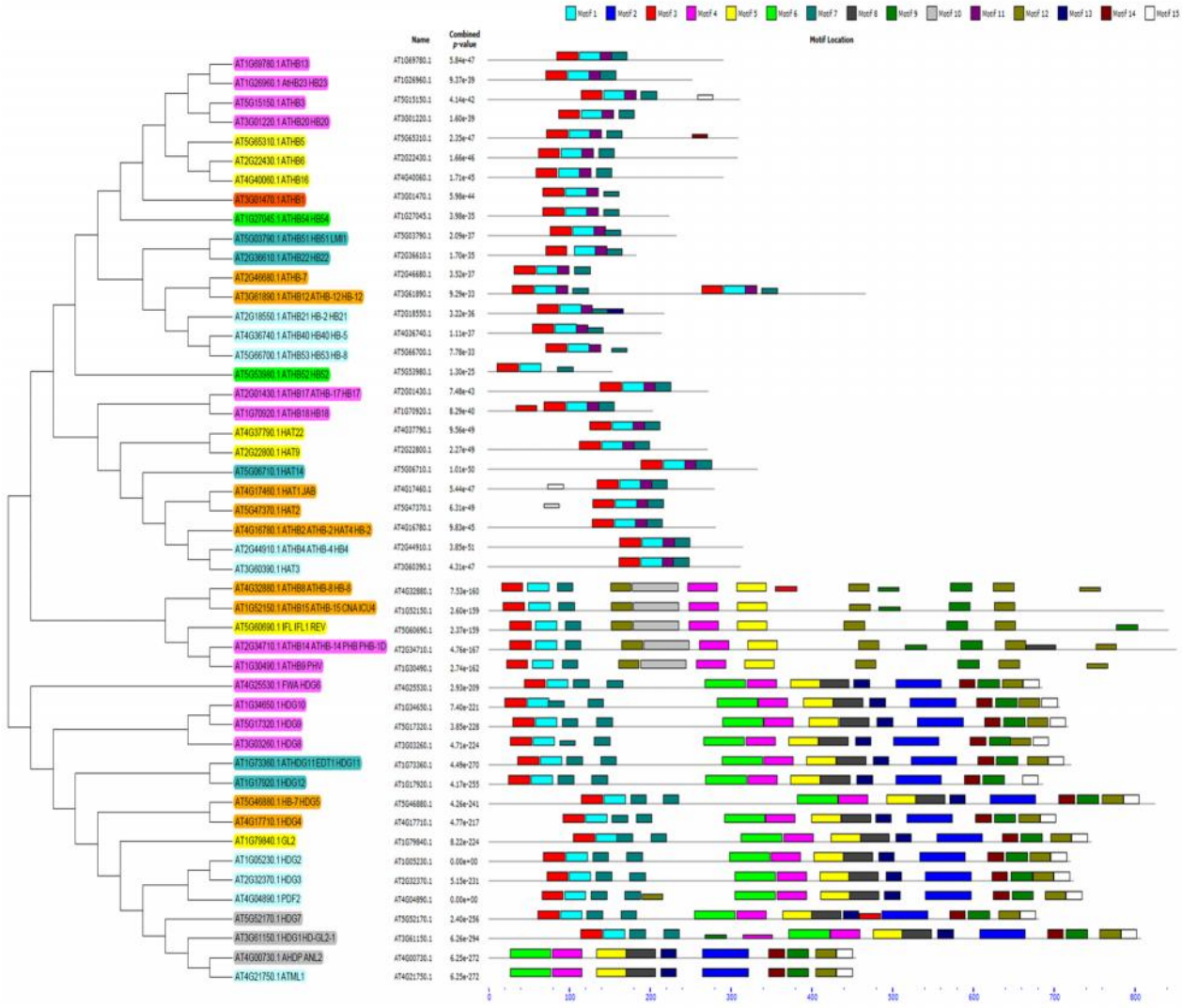

Figure-6: Phylogenetic relationship and motif composition; Conserved motif analysis done by using MEME online software and tree was constructed by using MEGA 6.0 software through multiple sequence alignment.

Analysis of gene Structure (Exons and Introns): Analysis of gene structure was used for mapping of exons, introns and UTR regions on the gene and thus to find out the location. These regions especially the coding regions that could be further used for functional annotation. The structural analysis of HD-Zip genes was done to find evidence of evolution from this family in *A. thaliana*. The observations revealed that number of exons was 1 to 18 in each gene. AT2G46680.1, AT3G61890.1, AT5G53980.1 had just one exon, while there were three genes that contain 18 exons including: AT1G30490.1, AT1G52150.1 & AT5G60690.1. The conserved number of introns and exons in genes was an indication for evolutionary relatedness (**Figure-7**).

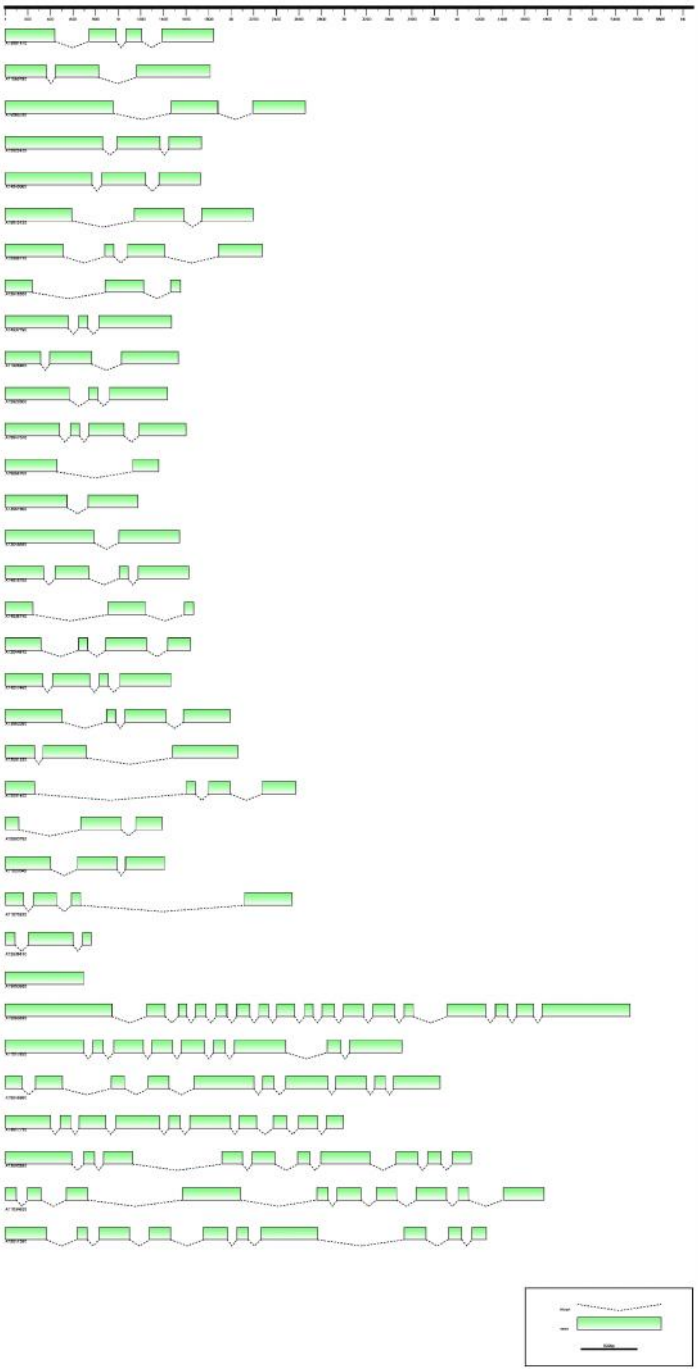

Figure-7: Gene structure analysis; Exon/intron structures of HD-Zip genes in *A. thaliana*. Colored boxes showing exons while line between boxes representing introns. Location of intron/exon could be estimated by using scale on the top.

Gene structure analysis revealed similarities and differences between genes of HD-Zip family to explain their evolutionary relationship. According to phylogenetic analysis this family was further divided in 4 sub-families depending upon the functional and structural similarities so to validate that fact we compared the results of structure analysis of genes, phylogenetic analysis and conserved motif analysis. In the result of comparison, majority of members of HD-Zip subfamily I and II were rich in few motifs like motif 1,3,11 and 15 while there were few exceptions that contain repeats of these motifs. Members of HD-Zip family III had 10 out of 15 motifs and members of HD-Zip family IV were composed of almost all 15 motifs which were selected to perform conserved analysis.

Conclusion: Here in the present study, we presented computational analysis of HD-Zip transcription factors family of *A. thaliana*, providing information about chromosomal mapping to locate the genes on chromosomes, motif analysis and analysis of conserved domains. The analysis of conserved domains represented that 44 residues were conserved before the signature box WFQNrr and 6 residues were conserved after the WFQNrr with few alternative residues in some genes, a total of 4 sub families were observed in the phylogenetic analysis. All the 4 sub families had conserved homeobox domain. This classification and analysis further categorized the transcription factors of 4 HD-Zip subfamilies into different clades and revealed a deeper evolutionary relationship among them. Thus, it may help to explore further functions of these transcription factors. The structural analysis of genes was performed for mapping location of exons and introns in gene. Promoter analysis of *cis*-regulatory element had revealed that almost every gene contains repetitive binding sites for transcriptional factors on the promoter region.

Acknowledgements: The authors gratefully acknowledge the financial support from Higher Education Commission of Pakistan through a research project grant.

REFERENCES

Azeem, F., B. Ahmad, R.M. Atif, M.A. Ali, H. Nadeem, S. Hussain, H. Manzoor, M. Azeem and M. Afzal (2018). Genome-Wide Analysis of Potassium Transport-Related Genes in Chickpea (Cicer arietinum L.) and Their Role in Abiotic Stress Responses. Plant Mol. Biol. Rep. 36:451-468.

Chen, E., X. Zhang, Z. Yang, X. Wang, Z. Yang, C. Zhang, Z. Wu, D. Kong, Z. Liu, G. Zhao, H. I. Butt, X. Zhang and F. Li (2017). Genome-wide analysis of the HD-ZIP IV transcription factor family in *Gossypium arboreum* and GaHDG11 involved in osmotic tolerance in transgenic Arabidopsis. Mol. Genet. Genome 292:593-609.

Chen, X., Z. Chen, H. Zhao, Y. Zhao, B. Cheng and Y. Xiang (2014). Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. PloS One 9:e87156.

Ding, Z., L. Fu, Y. Yan, W. Tie, Z. Xia, W. Wang, M. Peng, W. Hu and J. Zhang (2017). Genome-wide characterization and expression profiling of HD-Zip gene family related to abiotic stress in cassava. PloS One 12:e0173043.

Gong, S., Y. Ding, S. Hu, L. Ding, Z. Chen and C. Zhu (2019). The role of HD-Zip class I transcription factors in plant response to abiotic stresses. Physiol. Plant. 167:516-525.

Henriksson, E., A. S. B. Olsson, H. Johannesson, H. Johansson, J. Hanson, P. Engström and E. Söderman (2005). Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. Plant Physiol. 139:509-518.

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

Hofberger, J. A., D.L. Nsibo, F. Govers, K. Bouwmeester and M.E. Schranz (2015). A complex interplay of tandem-and whole-genome duplication drives expansion of the L-type lectin receptor kinase gene family in the brassicaceae. Genome Biol. Evol. 7:720-734.

Qasim, M., M. Azeem, M. Zubair, U. Qaiser, A.A. Shah, H. Nadeem, H. Manzoor, S. Hussain, B. Ahmad and F. Azeem (2018). Phylogenetic Analyses of the Trithorax Homologs and Related Proteins in *Cicer arietinum*. Int. J. Agric. Biol. 20:575-582.

Qiu, C. X., K.J. Zuo and K.X. Tang (2006). Class III Homeodomaiii-leuciiie Zipper Genes in Plant Development. Pakistan J. Biol. Sci. 9: 568-573.

Roodbarkelari, F. and E.P. Groot (2017). Regulatory function of homeodomain-leucine zipper (HD-ZIP) family proteins during embryogenesis. New Phytol. 213:95–104.

Rose, R., O. Golosova, D. Sukhomlinov, A. Tiunov and M. Prosperi (2018). Flexible design of multiple metagenomics classification pipelines with UGENE. Bioinformatics, bty901.

Sessa, G., M. Carabelli, M. Possenti, G. Morelli and I. Ruberti (2018). Multiple links between HD-Zip proteins and hormone networks. Int. J. Mol. Sci. 19:e4047.

Shao, J., I. Haider, L. Xiong, X. Zhu, R.M.F. Hussain, E. Övernäs, A.H. Meijer, G. Zhang, M. Wang, H.J. Boumeestr and P.B.F. Ouwerkerk (2018). Functional analysis of the HD-Zip transcription factor genes Oshox12 and Oshox14 in rice. PloS One 13:e0199248.

Shen, W., H. Li, R. Teng, Y. Wang, W. Wang and J.

Zhuang (2019). Genomic and transcriptomic analyses of HD-Zip family transcription factors and their responses to abiotic stress in tea plant (*Camellia sinensis*). Genomics 111:1142-1151.

Sultan, S., M.A. Ali, R.M. Atif, F. Azeem, H. Nadeem, M.H. Siddique, E. Filiz, K. Hussain and A. Abbas (2016). Genome wide analysis of stress responsive WRKY transcription factors in *Arabidopsis thaliana*. Turk. J. Agric-Food Sci. Technol. 4:279-290.

Todaka, D., K. Shinozaki and K. Yamaguchi-Shinozaki (2015). Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. Front. Plant Sci. 6:1–20.

Turchi, L., S. Baima, G. Morelli and I. Ruberti (2015). Interplay of HD-Zip II and III transcription factors in auxin-regulated plant development. J. Exp. Bot. 66:5043-5053.

Turchi, L., M. Carabelli,V. Ruzza,M. Possenti,M. Sassi,A. Peñalosa,G. Sessa,S. Salvi,V. Forte,G. Morelli and I. Ruberti (2013). Arabidopsis HD-Zip II transcription factors control apical embryo development and meristem function. Development 140:2118-2129.

Valdes, A.E., E. Övernäs, H. Johansson, A.R. Iglesias and P. Engström (2012). The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities. Plant Mol. Biol. 80:405-418.

Vinson, C.R., P.B. Sigler and S.L. McKnight (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Sci. 246:911-916.

Waqas, M., M.T. Azhar, I.A. Rana, F. Azeem, M.A. Ali, M.A. Nawaz, G. Chung and R.M. Atif (2019). Genome-wide identification and expression analyses of WRKY transcription factor family members from chickpea (*Cicer arietinum* L.) reveal their role in abiotic stress-responses. Genes & genomics, 41:467-481.

Wang, J. and L. Zhuang (2019). Identi fi cation and characterization of novel homeodomain leucine zipper (HD-Zip) transcription factors associated with heat tolerance in perennial ryegrass. Environ. Exp. Bot. 160:1–11.

Wenkel, S., J. Emery, B.H. Hou, M.M.S. Evans and M.K. Barton (2007) A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes. The Plant Cell 19:3379-3390.

Zhang, C.H., R.J. Ma, Z.J. Shen, X. Sun, N.K. Korir and M.L. Yu (2014). Genome-wide analysis of the homeodomain-leucine zipper (HD-ZIP) gene family in peach (*Prunus persica*). Genet. Mol. Res, 13:2654-2668.

Supplementary Table 1: Regulatory elements present in all the 48 HD-Zip TFs and their locations on promoter.

S. No.	Accession Number	Factor Name	Sequence of Factor	Strand	Repetition	Location on Promoter	
1	AT3G01470	CCAATBOX1	CCAAT	+	1	656	
		WRKY71OS	TGAC	+	6	103,320,352,368,447,861	
		DOFCOREZM	AAAG	+	12	179,209,270,309,549,601,605, 919,966	629,791,897,
		GATABOX	GATA	+	7	26,56,115,312,361,814,842	
		MYB1AT	WAACCA	+	1	187	
2	AT1G69780	CCAATBOX1	CCAAT	+	1	445	
		WRKY71OS	TGAC	-	1	497	
		DOFCOREZM	AAAG	+	12	50,156,219,236,242,262,321, 700,985	574,605,693,
		GATABOX	GATA	+	6	151,180,252,452,580,805	
		MYB1AT	WAACCA	-	1	762	
3	AT5G65310	CCAATBOX1	CCAAT	+	2	422,427	
		WRKY71OS	TGAC	+	3	125,260,326	
		DOFCOREZM	AAAG	+	6	46,623,692,716,828,891	
		GATABOX	GATA	+	2	713,931	
		MYB1AT	WAACCA	+	2	407,859	
4	AT2G22430	CCAATBOX1	CCAAT	+	2	397,542	
		WRKY71OS	TGAC	+	1	45	
		DOFCOREZM	AAAG	+	13	33,96,166,231,251,255,326,422,499,628,640,904, 982	
		GATABOX	GATA	+	5	6,129,136,527,818	
		MYB1AT	WAACCA	+	3	539,932,990	
5	AT4G40060	CCAATBOX1	CCAAT	-	1	914	

		WRKY71OS	TGAC	+	7	49,77,281,377,675,687,856
		DOFCOREZM	AAAG	+	7	83,127,140,262,972,976,997
		GATABOX	GATA	+	3	10,268,979
		MYB1AT	WAACCA	-	4	170,618,962,987
6	AT5G15150	CCAATBOX1	CCAAT	+	1	277
		WRKY71OS	TGAC	+	2	2,88
		DOFCOREZM	AAAG	+	5	65,158,339,681,813
		GATABOX	GATA	+	6	357,378,610,753,793,919
		MYB1AT	WAACCA	+	1	274
7	AT5G06710	CCAATBOX1	CCAAT	+	1	912
		WRKY71OS	TGAC	+	4	74,651,752,870
		DOFCOREZM	AAAG	+	5	53,217,667,736,861
		GATABOX	GATA	+	3	339,700,764
		MYB1AT	WAACCA	-	1	173
8	AT2G46680	CAATBOX1	CAAT	+	6	11,19,252,284,298,413
		WRKY71OS	TGAC	+	1	384
		DOFCOREZM	AAAG	+	7	83,319,452,797,808,972,983
		GATABOX	GATA	+	3	476,878,969
		MYBCORE	CNGTTR	+	1	667
9	AT4G16780	CAATBOX1	CAAT	+	10	39,173,320,326,371,440,123, 596,855,962
		WRKY71OS	TGAC	+	4	261,558,770,976
		DOFCOREZM	AAAG	+	14	77,126,158,228,234,244,510, 534,635,668,678,911, 919,954
		GATABOX	GATA	+	9	389,407,486,492,640,691,750, 758,870
		MYBCORE	CNGTTR	-	2	178,606
10	AT4G36740	CAATBOX1	CAAT	+	6	129,148,159,426,581,724
		WRKY71OS	TGAC	+	4	727,797,920,955
		DOFCOREZM	AAAG	+	6	280,464,697,737,789,801
		GATABOX	GATA	+	4	22,169,362,633
		MYBCORE	CNGTTR	-	1	342
11	AT2G44910	CAATBOX1	CAAT	+	1	64
		WRKY71OS	TGAC	+	5	97,253,664,738,818
		DOFCOREZM	AAAG	+	11	103,132,235,399,414,456,611, 762,954,974,979
		GATABOX	GATA	+	1	119
		MYBCORE	CNGTTR	+	1	989
12	AT4G17460	CAATBOX1	CAAT	+	6	14,118,273,327,462,817
		WRKY71OS	TGAC	+	3	66,352,958
		DOFCOREZM	AAAG	+	15	97,155,209,364,418,704,788, 798,835,850,869,875, 887,926
		GATABOX	GATA	+	7	38,447,480,791,801,853,906
		MYBCORE				
13	AT3G60390	CAATBOX1	CAAT	+	9	15,149,243,279,423,487,572, 756,825
		WRKY71OS	TGAC	+	2	746,968
		DOFCOREZM	AAAG	+	13	38,49,205,263,449,467,482,515,609,783,842,858, 905
		GATABOX	GATA	+	9	120,163,229,274,290,313,348, 353,776
		MYBCORE	CNGTTR	-	2	431,953
14	AT3G220	CAATBOX1	CAAT	+	3	113,160,673
		WRKY71OS	TGAC	+	1	222
		DOFCOREZM	AAAG	+	13	193,302,320,335,341,548,641, 685,736,744,757,763,892
		GATABOX	GATA	+	3	77,721,787
		MYBCORE	CNGTTR	-	2	580,264
15.	AT2G18550	CAATBOX1	CAAT	+	7	436,460,515,668,678,760,941
		WBOXATNPR1	TTGAC	+	2	625,848
		DOFCOREZM	AAAG	+	12	89,125,332,351,388,630,653, 704,775,802,918, 990
		GATABOX	GATA	+	6	59,206,250,639,694,874
		WRKY71OS	TGAC	+	2	626,849
16.	AT4G37790	CAATBOX1	CAAT	+	6	46,119,209,419,560,613
		WBOXATNPR1	TTGAC	+	2	260,498
		DOFCOREZM	AAAG	+	11	33,75,89,191,235,556,576,634,646,879,942

		GATABOX	GATA	+	6	275,289,459, 519,579,620
		WRKY71OS	TGAC	+	3	159,261 ,499
17.	AT1G26960	CAATBOX1	CAAT	+	5	23,573,641,657,712
		WBOXATNPR1	TTGAC	+	3	252,301,666
		DOFCOREZM	AAAG	+	8	59,85,104,115,358,455,928,950
		GATABOX	GATA	+	6	210,374,588, 791,844,888
		WRKY71OS	TGAC	+	5	253,302,626,667,784
18.	AT2G22800	CAATBOX1	CAAT	+	4	382,529,598,939
		WBOXATNPR1	TTGAC	-	1	767
		DOFCOREZM	AAAG	+	12	58,64,132,207,238,404,429,459,612,623,793,851
		GATABOX	GATA	+	9	71,76,118,396,547,562,602,657,662
		WRKY71OS	TGAC	+	1	261
19.	AT5G47370	CAATBOX1	CAAT	+	3	101 ,351,359
		WBOXATNPR1	TTGAC	+	2	96,449
		DOFCOREZM	AAAG	+	16	71,132,139,154,183,215,744, 888,894,916,925,929, 940, 947, 971, 993
		GATABOX	GATA	+	5	235,481,509,614,968
		WRKY71OS	TGAC	+	5	35 ,97,165,450,542
20.	AT5G66700	CAATBOX1	CAAT	+	7	22,79,324,467,836,863,882
		WBOXATNPR1	TTGAC	+	1	252
		DOFCOREZM	AAAG	+	15	7,93,97,164,175,197,230,284, 290, 509,565,820,926, 932,976
		GATABOX	GATA	+	2	631,650
		WRKY71OS	TGAC	+	1	253
21.	AT3G61890	CAATBOX1	CAAT	+	7	21,279,343,787,840 ,863,977
		WBOXATNPR1	TTGAC	+	1	353
		DOFCOREZM	AAAG	+	10	113,151,363,492,569,605,674, 697,706,773
		GATABOX	GATA	+	4	61,252,402,680
		WRKY71OS	TGAC	+	1	354
22	AT2G01430	CAATBOX1	CAAT	+	4	386,619,765,884
		WRKY71OS	TGAC	+	6	259,276,473,755,800,849
		DOFCOREZM	AAAG	+	4	39,47,209,295
		GATABOX	GATA	+	5	160,352,409,680
		MYB1AT	WAACCA	+	1	424
23	AT5G03790	CAATBOX1	CAAT	+	6	43,282,363,613,690,868
		WRKY71OS	TGAC	+	2	401,505
		DOFCOREZM	AAAG	+	11	68,100,186,383,461,634,804, 810,988,992,996
		GATABOX	GATA	+	5	248,393,429,735,969
		MYB1AT	WAACCA	-	2	55,166
24	AT1G27045	CAATBOX1	CAAT	+	7	234,314,362,475,787,830,971
		WRKY71OS	TGAC	-	6	708,736,785,889,913,942
		DOFCOREZM	AAAG	+	10	145,213,434,455,496,613,796, 852,880,940
		GATABOX	GATA	+	6	123,370,509,539,553,656
		MYB1AT	WAACCA	+	2	841,846
25	AT1G70920	CAATBOX1	CAAT	+	5	177,207,252,603,752
		WRKY71OS	TGAC	+	5	122,137,613,704,863
		DOFCOREZM	AAAG	+	6	1,47,90,381,494,874
		GATABOX	GATA	+	9	38,84,106,181,214,367,371,497,936
		MYB1AT	WAACCA	+	1	173
26	AT2G36610	CAATBOX1	CAAT	+	3	49,588,772
		WRKY71OS	TGAC	+	1	768
		DOFCOREZM	AAAG	+	8	219,255,344,876,901,943,972, 986
		GATABOX	GATA	+	8	23,101,109,349,763,873,879,964
		MYB1AT	WAACCA	+	2	729,3
27	AT5G53980	CAATBOX1	CAAT	+	7	227,257,262,279,771,791,942,
		WRKY71OS	TGAC	+	3	339,373,788
		DOFCOREZM	AAAG	+	9	5,41,72,84,291,545,704,724,895
		GATABOX	GATA	+	8	14,123,151,454,581,649,667,810
		MYB1AT	WAACCA	+	1	652
28	AT1G79840	CCAATBOX1	CAAT	+	12	268,293,400,516,542,726,787, 792,845,853,958,995
		WRKY71OS	TGAC	+	2	395,739

		DOFCOREZM	AAAG	+	5	432,526,633,667,797
		GATABOX	GATA	+	5	219,454,491,612,940
29	AT1G05230	CCAATBOX1	CAAT	+	9	9,47,368,711,754,876,938,963, 973
		WRKY71OS	TGAC	+	1	742
		DOFCOREZM	AAAG	+	7	51,58,65,101,666,691,805
		GATABOX	GATA	+	4	241,344,510,738
		MYB1AT	WAACCA	-	2	146,868
30	AT4G00730	CCAATBOX1	CAAT	+	6	27,152,420,424,679,983
		WRKY71OS	TGAC	+	2	486,510
		DOFCOREZM	AAAG	+	19	97,174,192,196,211,219,410, 465,614,703,721,725,867,884, 906, 943,959,964,974
		GATABOX	GATA	+	2	185,351
		MYB1AT	WAACCA	+	1	847
31	AT4G21750	CCAATBOX1	CAAT	+	6	210,435,477
		WRKY71OS	TGAC	+	3	424,824,863
		DOFCOREZM	AAAG	+	7	68,176,392,573,772,809,854
		GATABOX	GATA	+	2	196,368
		MYB1AT	WAACCA	+	2	527,592
32	AT4G04890	CCAATBOX1	CAAT	+	3	502,534,649
		WRKY71OS	TGAC	+	6	105,141,219,654,699,873
		DOFCOREZM	AAAG	+	5	176,247,349,849,924
		GATABOX	GATA	+	6	43,289,296,426,455,913
		MYB1AT	WAACCA	-	2	118,867
33	AT4G25530	CCAATBOX1	CAAT	+	8	50,55,137,287,559,730,773,807
		WRKY71OS	TGAC	-	2	48,654
		DOFCOREZM	AAAG	+	10	177,256,374,386,517,554,749, 754,800,876
		GATABOX	GATA	+	2	278,685
		MYB1AT	WAACCA	+	4	183,394,400,776
34		CCAATBOX1	CAAT	+	5	48,113,202,534,686,
	AT4G32880	WRKY71OS	TGAC	+	2	718,743
		DOFCOREZM	AAAG	+	10	77,424,548,600.664,751,778, 825,837,979
		GATABOX	GATA	+	3	106,125,196
		MYB1AT	WAACCA	+	1	109
35	AT2G34710	CCAATBOX1	CAAT	+	3	240,263,534
		WRKY71OS	TGAC	+	1	234
		DOFCOREZM	AAAG	+	17	41,119,127,133,167,272,277, 283,295,647,692,760,767,868, 873, 923,951
		ARR1AT	NGATT	+	3	297,710,934
36	AT3G61150	CCAATBOX1	CAAT	+	4	160,272,363,583
		WRKY71OS	TGAC	+	3	171,651,868
		DOFCOREZM	AAAG	+	8	188,395,718,757,761,805,824, 846
		ARR1AT	NGATT	+	5	102,130,565,704,772
		AMYBOX1	TAACARA	+	1	623
37	AT1G30490	CCAATBOX1	CAAT	+	5	127,375,379,823,990
		WRKY71OS	TGAC	-	2	142,587
		DOFCOREZM	AAAG	+	11	205,278,428,501,516,565,590, 629,717,752,880
		ARR1AT	NGATT	+	2	250,567
38	AT1G52150	CCAATBOX1	CAAT	+	7	289,502,577,603,671,683,782,
		WRKY71OS	TGAC	+	5	39,280,439,582,680
		DOFCOREZM	AAAG	+	6	371,426,800,818,901,939
		ARR1AT	NGATT	+	6	185,202,258,561,613,984
		AMYBOX1	TAACARA	-	1	17
39	AT1G73360	CCAATBOX1	CAAT	+	6	320,451,474,738,836,927
		WRKY71OS	TGAC	+	1	337
		DOFCOREZM	AAAG	+	14	85,143,169,518,594,612,732, 778,801,841,886,895,972, 992
		ARR1AT	NGATT	+	3	95,211,637
		AMYBOX1	TAACARA	+	1	749
40	AT2G32370	CCAATBOX1	CAAT	+	5	276,455,725,841,995
		WRKY71OS	TGAC	+	5	57,64,101,686,728
		DOFCOREZM	AAAG	+	3	113,347,866

		ARR1AT	NGATT	+	2	354,420
41	AT5G52170	CCAATBOX1	CAAT	+	1	875
		WRKY71OS	TGAC	+	4	141,479,872,970
		DOFCOREZM	AAAG	+	13	170,292,421,484,601,608,658, 692,718,723,773,815,865
		ARR1AT	NGATT	+	4	622,963,804,809
		ARR1AT AMYBOX1	TAACARA	+	1	714
42	AT5G60690	CCAATBOX1	CCAAT	-	2	535,786
		WRKY71OS	TGAC	+	1	606
		DOFCOREZM	AAAG	+	5	127,726,818,839,941
		GATABOX	GATA	+	8	376,430,458,488,505,744,794, 902
		MYBCORE	CNGTTR	+	1	448
43	AT1G17920	CCAATBOX1	CCAAT	-	2	172,445
		WRKY71OS	TGAC	+	3	372,491,674
		DOFCOREZM	AAAG	+	11	179.191,243,274,482,514.547, 735,889.898,950
		GATABOX	GATA	+	5	87.96.226,277,935
		MYBCORE	CNGTTR	+	2	570,814
44	AT5G46880	CCAATBOX1	CCAAT	--	--	--
		WRKY71OS	TGAC	+	3	182,247,271
		DOFCOREZM	AAAG	+	12	43,66,241,464,422,490,495,607,610,631,637,880
		GATABOX	GATA	+	1	822
		MYBCORE	CNGTTR	+	4	221,396,409.915
45	AT4G17710	CCAATBOX1	CCAAT	+	1	486
		WRKY71OS	TGAC	+	2	275,416
		DOFCOREZM	AAAG	+	11	56,87,352,426,499,593,710,750,928,935,951
		GATABOX	GATA	+	5	59,98,146,355,368
		MYBCORE	CNGTTR	+	3	782,790,803
46	AT3G03260	CCAATBOX1	CCAAT	+	1	692
		WRKY71OS	TGAC	+	4	496,536,543,657
		DOFCOREZM	AAAG	+	5	26,220,414,526,706
		GATABOX	GATA	+	2	265,457
		MYBCORE	CNGTTR	+	1	192
47	AT5G17320	CCAATBOX1	CCAAT	-	1	157
		WRKY71OS	TGAC	+	3	383,447,876
		DOFCOREZM	AAAG	+	9	98,362,405,470,491,572,629, 733,748
		GATABOX	GATA	+	3	311,440,718
		MYBCORE	CNGTTR	+	1	70
48	AT1G34650	CCAATBOX1	CCAAT	+	1	612
		WRKY71OS	TGAC	+	4	529,603,628,876
		DOFCOREZM	AAAG	+	3	372,376,397
		GATABOX	GATA	+	3	255,682,977
		MYBCORE	CNGTTR	+	1	78