Heliyon 10 (2024) e28045

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# In silico genome-wide analysis of homeodomain-leucine zipper transcription factors in *Cannabis sativa* L

Zhan-Ping Zhang [a,1], Zhen Wang [a,1], Jia-Xin Lu [a,1], Song Yan [a], Lian-Qing He [a], Pan-Pan Wang [a], Chen Qin [a], Wei-Chao Ren [a], Jiao Xu [b,d], Jian-Li Wu [e], Xiu-Bo Liu [b,**], Wei Ma [a,c,*]

[a] *Pharmacy of College, Heilongjiang University of Chinese Medicine, Harbin, China*
[b] *Department of Chinese Medicine, Jiamusi Campus, Heilongjiang University of Chinese Medicine, Jiamusi, China*
[c] *Key Laboratory of Basic and Application Research of Beiyao (Heilongjiang University of Chinese Medicine), Ministry of Education, Harbin, China*
[d] *Institute of Economic Crops, Heilongjiang Academy of Agricultural Sciences, Harbin, China*
[e] *Academy of traditional Chinese medicine, Heilongjiang University of Chinese Medicine, Harbin, China*

ARTICLE INFO

*Keywords:*
*Cannabis sativa* L.
Homeodomain-leucine zipper
Genome-wide analysis
bioinformatics

ABSTRACT

HD-Zip (Homeodomain-Leucine Zipper) is a family of transcription factors unique to higher plants and plays a vital role in plant growth and development. Increasing research results show that HD-Zip transcription factors are widely involved in many life processes in plants. However, the HD-Zip transcription factor for cannabis, a valuable crop, has not yet been identified. The sequence characteristics, chromosome localization, system evolution, conservative motif, gene structure, and gene expression of the HD-Zip transcription factor in the cannabis genome were systematically studied. Real-time quantitative polymerase chain reaction (qRT-PCR) was used to verify its function. The results showed that cannabis contained 33 *HD-Zip* gene members. The number of amino acids is 136–849aa, the isoelectric point is 4.54–9.04, and the molecular weight is 23264.32–93147.87Da. Many *cis*-acting elements are corresponding to hormone and abiotic stress in the HD-Zip family promoter area of cannabis. Sequencing of the transcriptome at 5 tissue sites of hemp, stems, leaves, bracts, and seeds showed similar levels of expression of 33 members of the *HD-Zip* gene family at 5 tissue sites. Bioinformatics results show that HD-Zip expression is tissue-specific and may be influenced by hormones and environmental factors. This lays a foundation for further research on the gene function of *HD-Zip*.

## 1. Introduction

HD-Zip (homeodomain-leucine zipper) proteins, homeobox family proteins, are transcription factors unique to higher plants that play vital roles in plant growth and development. They have a highly conserved (HD) domains and leucine zipper (LZ) domains that can form "helix-loop-helix-turned-helical" folding structures [1]. The LZ domain is a stable zipper-like hydrophobic domain formed by specific polypeptide chains, including multiple leucine residues, that interact with other proteins and perform various functions. Based

* Corresponding author. Pharmacy of College, Heilongjiang University of Chinese Medicine, Harbin, China.
** Corresponding author.
*E-mail addresses:* zyylxb@foxmail.com (X.-B. Liu), mawei@hljucm.net (W. Ma).
[1] The first three authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e28045
Received 9 August 2023; Received in revised form 4 March 2024; Accepted 11 March 2024
Available online 26 March 2024

on the amino acid structural domain characteristics and sequence similarity, the HD-Zip gene family can be divided into four subfamilies: HD-Zip I, II, III, and IV, all of which contain conserved HD and LZ structural domains [2].

*Cannabis sativa* L. is an annual herb belonging to the Cannabaceae and is one of the oldest domesticated plants worldwide [3]. Cannabis contains many active medicinal ingredients such as terpenes, flavonoids, phenolic terpenoids, and alkaloids [4]and was first used as a medicinal plant in the Middle East and Asia as early as the 6th century BC. Cannabis was introduced to Western medicine in the early 19th century [5]. To date, 565 compounds have been isolated from cannabis, which can be classified as cannabinoids or non-cannabinoids, 120 of which are phytocannabinoids. Cannabinoids are the only phenolic terpenoid secondary metabolites in cannabis whose structure is derived from geranyl pyrophosphate [6]. Tetrahydrocannabinol (THC) was the first cannabinoid to be discovered and studied and is known for its psychoactive properties [7]. Due to its hallucinogenic and addictive properties, it can only be used for medical research under state laws and regulations. Cannabidiol (CBD) is a high-level, fat-soluble, non-psychoactive cannabinoid and the second major active ingredient in cannabis. Compared to THC, CBD is not hallucinogenic, safer, reduces THC side effects, and does not affect movement or memory function.

HD-Zip transcription factors are extensively involved in plant growth, morphogenesis, and biological and abiotic stresses, among other life processes [8]. The function and number of members of the four HD-Zip subfamilies vary depending on their structures [9]. Among them, the HD-Zip I subfamily proteins are involved in the regulation of plant processes such as response to adversity in response to abiotic and biotic environmental factors, hormone response, response to light signals, and organ development. The HD-Zip II subfamily is involved in plant organ development and is related to light response and hormone response. The HD-Zip III subfamily is mainly involved in apical meristematic tissue differentiation, vascular tissue formation, apical meristem differentiation regulation, vascular tissue formation, polar transport of growth hormones, embryonic development, and lateral organ development in the proximal region [10]. Finally, the HD-Zip IV subfamily controls trichome production, anthocyanin accumulation, and epithelial cell differentiation. In the *Arabidopsis thaliana* model plant, the Arabidopsis subfamily I has 17 members, the Arabidopsis subfamily HD-Zip II has 10 members [11], and the HD-Zip III and HD-Zip IV subfamilies have five and 16 members [12], respectively. During disease resistance, plants exhibit various physiological and biochemical responses controlled by transcription factors [13]. The *HD-Zip* gene, which is associated with various plant growth processes and stress responses, is considered a key transcription factor for plant germplasm improvement. For example, *Oshox22* binds to CAAT (G/C)ATTG elements to enhance rice sensitivity to ABA signaling and regulate drought resistance and salt tolerance [14]. Furthermore, the cotton HD-ZIP Type I transcription factor (GhHB12) is specifically expressed in axillary buds and interacts with GhSPL10 and GhSPL13 to inhibit GhFT, GhFUL, and GhSOC1 expression and regulate cotton flowering time [15]. In *Medicago truncatula* Gaertn., the HD-Zip transcription factor HB1 is expressed in primary and lateral root meristem tissues and enhances tolerance to salt stress by regulating the contact area of the root system with saline soil [16]. However, studies on HD-Zip transcription factors in cannabis have not been reported; therefore, we aimed to use bioinformatics to predict the potential biological function of HD-Zip transcription factors and identify regulatory genes that influence the biosynthesis of cannabinoids, providing a rationale for selecting and cultivating high-quality cannabis varieties with high CBD and low THC content.

## 2. Materials and methods

### 2.1. Data sources and plant materials

The genomic data from the study was obtained from female strains of the cannabis CRBRx strain from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/)(GCA_900626175.1) [17,18]. Transcriptome data included RNA-seq data of flowers, bracts, leaves, stems, and roots of female DK varieties (PRJNA498707). The fragments per kilobase of transcript per million mapped reads (FPKM) value of *HD-Zip* was extracted from the RNA sequencing (RNA-seq) data of *Cannabis sativa* L. The *HD-Zip* gene expression was visually analyzed by TBtools software, and all FPKM values were processed by row scaling [19].

### 2.2. HD-Zip transcription factor identification and analysis in cannabis

The PlantTFDB website (http://planttfdb.gao-lab.org/) predicted the possible HD-Zip transcription factors, further screened the domains of the candidate sequences through the CD-Search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and Pfam database (PF02365), and removed the redundancy to obtain the *HD-Zip* gene family members. The ExPASy website (https://web.expasy.org/protparam/) was used to predict the size, relative molecular mass, theoretical isoelectric point, and average hydrophilic coefficient of the HD-Zip cannabis protein sequence. HD-Zip subcellular localization was observed using the WoLFPSORT website (http://www.genscript.com/wolf-psort.html), and the chromosomal distribution was visualized using TBtools software [20].

### 2.3. Multiple sequence comparison and phylogenetic analysis of cannabis HD-Zip transcription factors

Comparative analysis of the conserved structural domains of the cannabis HD-Zip proteins was performed using the DNAMAN 9.0 software. Obtain the HD-Zip protein data of *Arabidopsis thaliana* in the PlantTFDB database, and merge it with the marijuana HD-ZIP protein file. The MUSCLE program in MEGA7.0 software (http://www.megasoftware.net) was used to perform multiple sequence alignment and use MEGA-X software to use the Neighbor-joining (NJ), the Bootstrap value Set to 1000 times to build a system evolution.

### 2.4. Structural and motif analysis of the cannabis HD-Zip transcription factor

Gene structure was analyzed using TBtools software, and conserved motifs were predicted using the online website MEME (http://meme-suite.org/tools/meme) [19], the maximum number of motifs were set to 18 and other settings of default parameters, and the evolutionary tree of gene structure and motifs combined with the cannabis HD-Zip transcription factor was visualized using TBtools for visualization. Conserved domains in CsHD-Zip proteins were forecasted using the NCBI website's CD-Search tool, displayed alongside the *CsHD-Zip* phylogenetic tree.

### 2.5. Prediction of cis-acting elements of the cannabis HD-Zip transcription factor

The 2000 bp sequence upstream of the cannabis *HD-Zip* coding genes was extracted using TBtools software as a *cis*-element sequence in the promoter region, and the sequence was submitted to PLANTCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for *cis*-acting element prediction [19]. The resulting files were filtered and categorized for visualization.

### 2.6. Expression analysis of cannabis HD-Zip transcription factors

Using the transcriptome data of flowers, bracts, leaves, stems, roots, and seeds [21], the Dinafem (named Dinamed Kush) variety female plant was obtained by the group using a cross between Purple Kush and Dinamed Autoflowering CBD. The heat map was plotted using TBtools software for clustering and differential expression analysis [19].

### 2.7. Real-time fluorescence quantitative PCR validation of cannabis HD-Zip transcription factors

Fragments per kilobase of transcript per million mapped (FPKM) reads values of HD-Zip transcription factors were extracted from HD-Zip transcription factors expression heat maps established using TBtools software. The FPKM values were treated using row-scale transformation. Transcriptome data were validated by qRT-PCR. Total RNA was first extracted using the RNA Easy Fast Plant Tissue Kit (Tiangen Biotech Co., Ltd., Beijing, China), followed by reverse transcription of the RNA to cDNA using the Fastking RT Kit (With DNase) (Tiangen Biotech Co., Ltd., Beijing, China). Finally, StarLighter HP SYBR Green qPCR Mix (Universal) (Beijing Foreverstar Biotech Co., Ltd., Beijing, China) was used for qRT-PCR assays using cDNA as a template. The *EF1α* gene was used as the reference gene. The qRT-PCR primers used in this study are listed in Table S1. The PCR amplification conditions used for all reactions were as follows: 3 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 20 s at 60 °C and 15 s at 72 °C. The relative expression of each gene was calculated using the $2^{-\triangle\triangle Ct}$ algorithm, and three technical replicates were set up for each experiment. These experimental steps were performed according to the manufacturer's instructions. The qPCR primers were designed using the qPCR primers Primer Premier (version 5.0) software and synthesized for this study at Rebbosinko Biotech Ltd. In Beijing.

## 3. Results

### 3.1. Cannabis HD-Zip and its physicochemical properties

By identifying the structural domains, 33 members of the cannabis *HD-Zip* family were identified. Based on the chromosomal localization of *HD-Zip*, the 33 members were named *CsHD1-CsHD33* (Fig. 1). In the cannabis *HD-ZIP* family, each *HD-ZIP* gene can be found to find matched chromosomal information. The distribution is evenly distributed on 10 chromosomes. Chromosomes 2 and 10

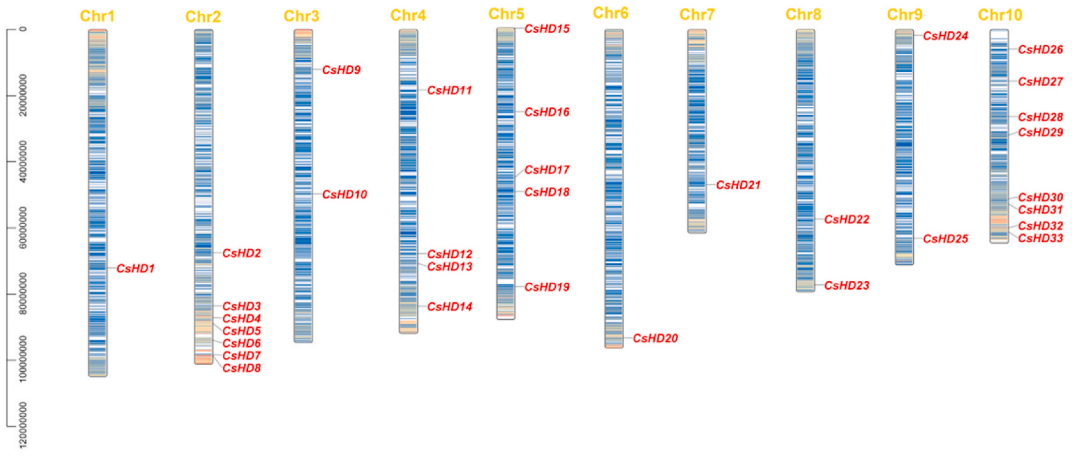


**Fig. 1.** Chromosomal location of the CsHD gene. Chromosome length (bp) is denoted on the y-axis. The concentration of genes is depicted through colored streaks on every chromosome.

**Table 1**
Information and characteristics of *CsHD* gene family.

| Gene name | Locus name | Accession NO. | Location | aa | Subcellular location | pI | M.W (Da) | GRAVY |
|---|---|---|---|---|---|---|---|---|
| *CsHD1* | LOC115714257 | rna-XM_030642892.1 | 72110347-72111990 | 327 | nuclear | 6.07 | 38 005.78 | −1.128 |
| *CsHD2* | LOC115706074 | rna-XM_030633590.1 | 67468999-67470620 | 320 | nuclear | 5.42 | 36 805.48 | −1.090 |
| *CsHD3* | LOC115706377 | rna-XM_030634009.1 | 83514970-83521795 | 333 | nuclear | 7.30 | 36 865.70 | −0.826 |
| *CsHD4* | LOC115706571 | rna-XM_030634260.1 | 87067285-87073409 | 840 | nuclear | 5.69 | 91 329.15 | −0.348 |
| *CsHD5* | LOC115704522 | rna-XM_030631725.1 | 88864933-88865886 | 247 | nuclear | 9.04 | 28 809.07 | −1.041 |
| *CsHD6* | LOC115707515 | rna-XM_030635513.1 | 93979858-93986884 | 837 | nuclear | 5.94 | 92 068.43 | −0.134 |
| *CsHD7* | LOC115708255 | rna-XM_030636476.1 | 98418784-98423894 | 772 | nuclear | 5.81 | 84 561.23 | −0.364 |
| *CsHD8* | LOC115707829 | rna-XM_030635910.1 | 98581016-98585765 | 755 | nuclear | 5.64 | 83 575.46 | −0.457 |
| *CsHD9* | LOC115709290 | rna-XM_030637356.1 | 12044206-12046199 | 347 | nuclear | 4.54 | 39 126.85 | −0.755 |
| *CsHD10* | LOC115708712 | rna-XM_030636707.1 | 49693567-49696348 | 249 | nuclear | 7.71 | 27 860.30 | −0.847 |
| *CsHD11* | LOC115714770 | rna-XM_030643524.1 | 18199201-18204059 | 734 | nuclear | 6.33 | 79 965.59 | −0.328 |
| *CsHD12* | LOC115715244 | rna-XM_030644084.1 | 67793604-67796239 | 348 | nuclear | 5.27 | 39 339.10 | −0.859 |
| *CsHD13* | LOC115712233 | rna-XM_030640495.1 | 70722710-70724404 | 301 | nuclear | 8.63 | 32 984.09 | −0.654 |
| *CsHD14* | LOC115714457 | rna-XM_030643176.1 | 83654078-83655827 | 214 | nuclear | 6.16 | 25 043.36 | −0.907 |
| *CsHD15* | LOC115715621 | rna-XM_030644277.1 | 60988-62153 | 204 | nuclear | 8.87 | 24 629.62 | −1.264 |
| *CsHD16* | LOC115716687 | rna-XM_030645541.1 | 25253647-25260730 | 761 | nuclear | 5.62 | 83 208.70 | −0.281 |
| *CsHD17* | LOC115716879 | rna-XM_030645791.1 | 45211201-45218371 | 814 | nuclear | 5.66 | 90 670.40 | −0.458 |
| *CsHD18* | LOC115716550 | rna-XM_030645362.1 | 49440555-49447624 | 814 | nuclear | 5.62 | 90 647.37 | −0.458 |
| *CsHD19* | LOC115716393 | rna-XM_030645179.1 | 78150464-78152717 | 330 | nuclear | 8.21 | 36 771.94 | −0.817 |
| *CsHD20* | LOC115719260 | rna-XM_030648226.1 | 93220975-93226670 | 845 | cytoplasmic | 5.80 | 92 266.92 | −0.107 |
| *CsHD21* | LOC115723870 | rna-XM_030653335.1 | 46829690-46836379 | 737 | nuclear | 6.41 | 81 729.85 | −0.329 |
| *CsHD22* | LOC115725296 | rna-XM_030654760.1 | 57241171-57243123 | 399 | nuclear | 7.17 | 44 591.80 | −1.114 |
| *CsHD23* | LOC115695550 | rna-XM_030622607.1 | 77222014-77223744 | 139 | nuclear | 6.14 | 16 154.02 | −1.069 |
| *CsHD24* | LOC115697662 | rna-XM_030624750.1 | 1743268-1743785 | 136 | nuclear | 8.46 | 15 942.65 | −1.204 |
| *CsHD25* | LOC115697202 | rna-XM_030624117.1 | 63113580-63120239 | 849 | nuclear | 5.76 | 93 147.87 | −0.187 |
| *CsHD26* | LOC115701102 | rna-XM_030628797.1 | 5892080-5893934 | 321 | nuclear | 4.88 | 36 235.93 | −0.815 |
| *CsHD27* | LOC115699352 | rna-XM_030626707.1 | 15574919-15577141 | 309 | nuclear | 6.42 | 35 242.97 | −0.990 |
| *CsHD28* | LOC115701160 | rna-XM_030628868.1 | 26309613-26311521 | 357 | nuclear | 8.57 | 39 434.83 | −0.795 |
| *CsHD29* | LOC115700801 | rna-XM_030628454.1 | 31944245-31947663 | 231 | nuclear | 8.37 | 26 638.14 | −0.823 |
| *CsHD30* | LOC115698490 | rna-XM_030625561.1 | 51193527-51194520 | 205 | nuclear | 8.77 | 23 264.32 | −0.660 |
| *CsHD31* | LOC115699242 | rna-XM_030626545.1 | 52728392-52735653 | 740 | nuclear | 5.63 | 80 726.23 | −0.322 |
| *CsHD32* | LOC115700934 | rna-XM_030628613.1 | 60038579-60042676 | 328 | nuclear | 4.88 | 36 590.95 | −0.933 |
| *CsHD33* | LOC115700384 | rna-XM_030627943.1 | 61131724-61133917 | 380 | nuclear | 6.51 | 43 560.28 | −1.244 |

have many genes, containing seven and eight *CSHD* genes, Chromosome 1 contained one *CsHD* gene, chromosome 3 contained two *CsHD* genes, chromosome 4 contained four *CsHD* genes, chromosome 5 contained five *CsHD* genes, chromosome 6 contained one *CsHD* gene, chromosome 7 contained one *CsHD* gene, chromosome 8 contained two *CsHD* genes, and chromosome 9 contained two *CsHD* genes.

Analysis of members of the cannabis *HD-Zip* gene family using correlation analysis sites (Table 1) revealed amino acid counts ranging from a low of 136 aa to a high of 849 aa. The isoelectric points from 4.54 to 9.04, the relative molecular mass ranged from 23264.32 Da to 93147.87 Da, and the GRAVY values were negative (from −0.107 to −1.264), indicating that all hemp HD-Zip proteins are hydrophilic proteins but have different degrees of hydrophilicity. The prediction of subcellular localization showed that all the proteins except CsHD20 were located in the cytoplasm. These results may provide a theoretical basis for the study of the family functions of the *HD-Zip* gene in cannabis.

### 3.2. Evolutionary relationships of cannabis HD-Zip genes and analysis of covariance

To understand the evolutionary relationship between the *HD-Zip* gene family and cannabis, a phylogenetic tree was constructed based on 81 proteins from *Cannabis sativa* and *Arabidopsis thaliana* (Fig. 2). The cannabis *HD-Zip* family comprises four subfamilies. Of these, 14 were *HD-ZipI*, seven were *HD-ZipII*, three were *HD-ZipIII* and nine were *HD-ZipIV*.

Collinearity analysis of *Cannabis sativa* with *Oryza sativa* (rice) and *Arabidopsis thaliana* revealed that there were 20 homologous gene pairs, with 12 pairs of homologous genes common to these three species (Fig. 3). Between *C. sativa* and rice, four pairs of similar genes were discovered, and an additional 16 pairs of similar genes were found between *C. sativa* and *Arabidopsis*.

### 3.3. Conserved motifs and gene structure of cannabis HD-Zip

The MEME website (https://meme-suite.org/meme/tools/meme) was used to analyze the conserved motifs of the cannabis HD-Zip protein online, and the upper limit of the number of conserved motifs identified was 18, named motif 1–18. Utilizing TBtools software for visualization, the findings revealed (Fig. 4A&4B) that *CsHD-Zip* conserved motif count varied between 3 and 15, among which, the members of *HD-Zip* subfamily I and subfamily II both contained only The three members of *HD-Zip* subfamily III all contain motif 1, motif 2, motif 6, motif 14, motif 9, motif 4, motif 7, motif 12, motif 18, and *CsHD-Zip20* also contains motif 13. The members of the *HD-Zip* IV subfamily include motif 2, motif 1, motif 16, motif 5, motif 14, motif 9, motif 4, motif 15, motif 7, motif 10, motif 17, motif 3, motif 11, motif 8, and motif 13.

Intron and exon analysis of *CsHD-Zip* gene family members revealed that the number of introns in members of the same subfamily

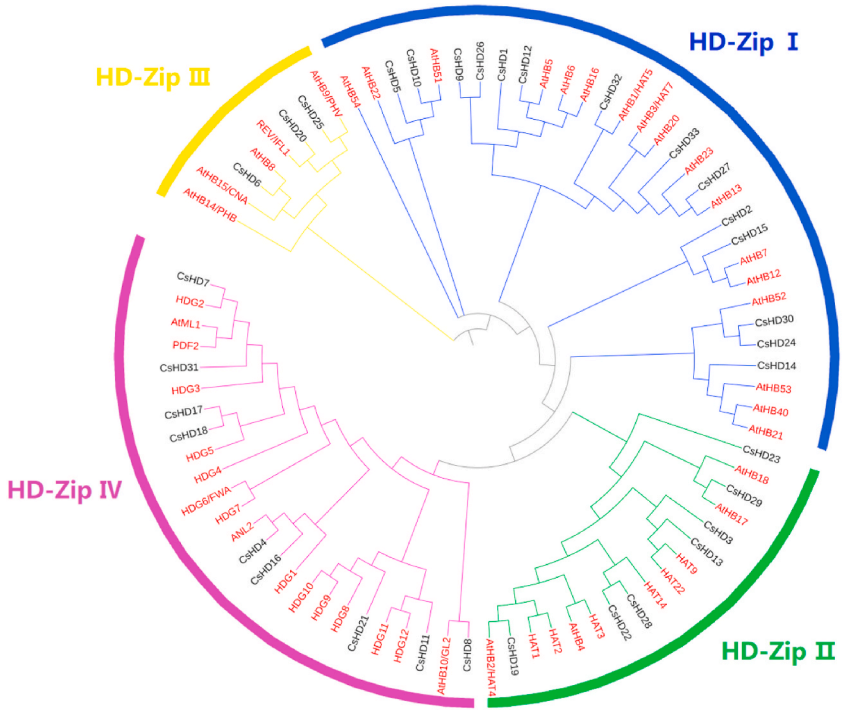


**Fig. 2.** Phylogenetic analysis of homeodomain-leucine zipper (HD-Zip) family in *Cannabis sativa* and *Arabidopsis thaliana.* The neighbor-joining phylogenetic tree was constructed using thirty-three *HD-Zips* from *C. sativa* and 48 *HD-Zips* from *Arabidopsis*. Four distinct subfamilies were identified for the HD-Zip proteins.

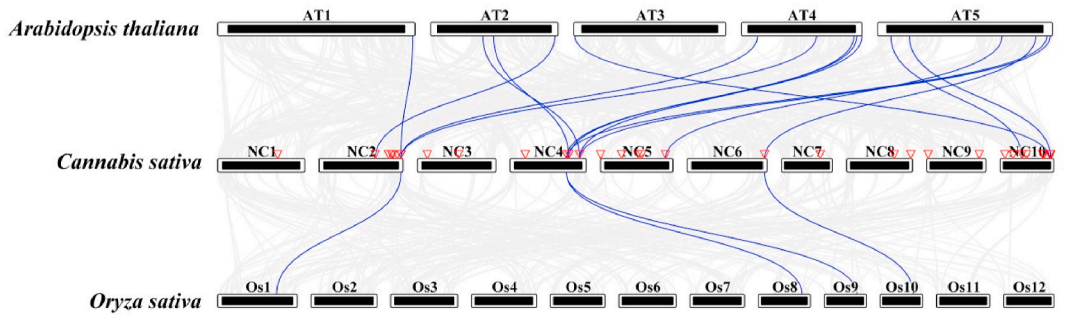


**Fig. 3.** Collinearity analysis of *homeodomain-leucine zipper* (*HD-Zip*) gene between *Cannabis sativa*, *Oryza sativa,* and *Arabidopsis thaliana*. Colored lines denote syntenic regions between cannabis chromosomes and others.

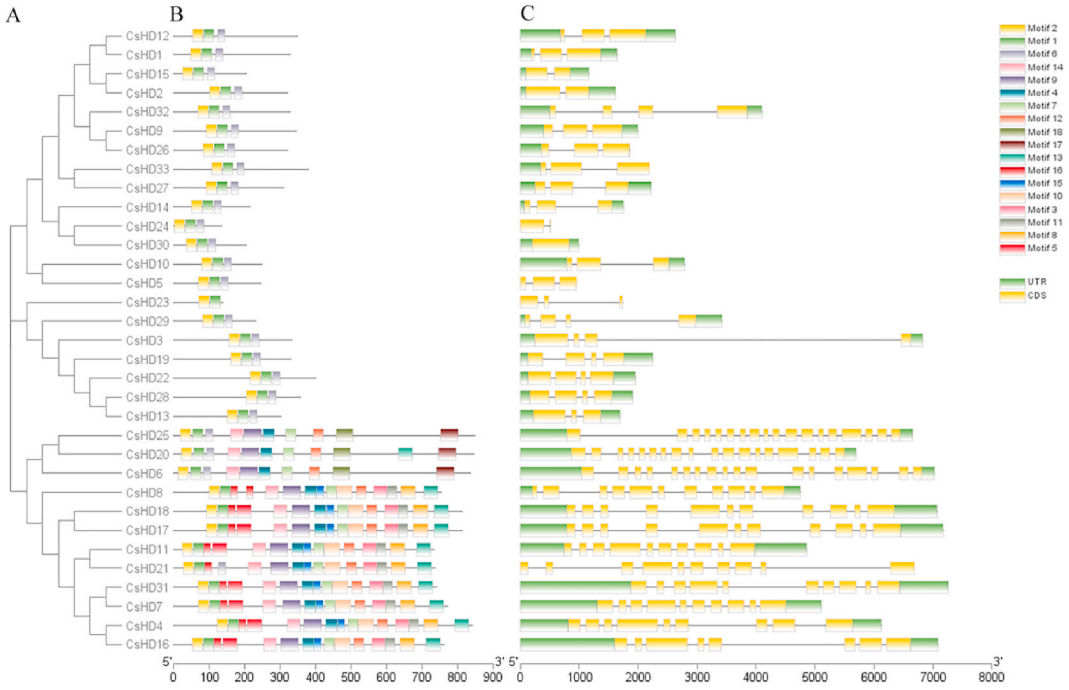


**Fig. 4.** The structure of the *HD-Zip* genes. A. Phylogenetic tree of CsHD proteins. B. Conserved domains of CsHD proteins. C. The gene structures and conserved domains of CsHD. Introns are shown as black lines.

was relatively consistent, however, the number of introns and exons in members of different subfamilies differed significantly (Fig. 4C). Differences in intron and exon numbers in the cannabis *HD-Zip* gene reflect structural diversity among members of the cannabis *HD-Zip* family, whereas exon numbers are more uniform among members of the same branch of the evolutionary tree.

### *3.4. Cannabis HD-Zip gene cis-elements prediction*

*Cis*-acting elements and *trans*-interactions can regulate gene expression, and analysis of their upstream *cis*-acting elements can predict the potential function of the corresponding genes. In the cannabis *HD-Zip* gene family, *cis*-elements are present in the promoter regions of each gene. Statistical analysis of the *cis*-elements revealed that photoreactive elements were the most abundant (Fig. 5), followed by hormone-like *cis*-elements (gibberellin, baicalin, salicylic acid, growth hormone, and methyl jasmine). Three genes, *CsHD-Zip16*, *CsHD-Zip17*, and *CsHD-Zip18*, had the highest number of MeJA-responsive elements at positions 3, 5, and 5, respectively.

### *3.5. Heat map of cannabis HD-Zip expression pattern with qRT-PCR validation*

Analysis of the expression patterns of cannabis *HD-Zip* gene family members in different tissue sites revealed that in flowers *CsHD-1*, *CsHD-2*, *CsHD-3*, *CsHD-4*, *CsHD-8*, *CsHD-9*, *CsHD-10*, *CsHD-12*, *CsHD-13*, *CsHD-14*, *CsHD-15*, *CsHD 17*, *CsHD-18*, *CsHD-19*, *CsHD-20*,

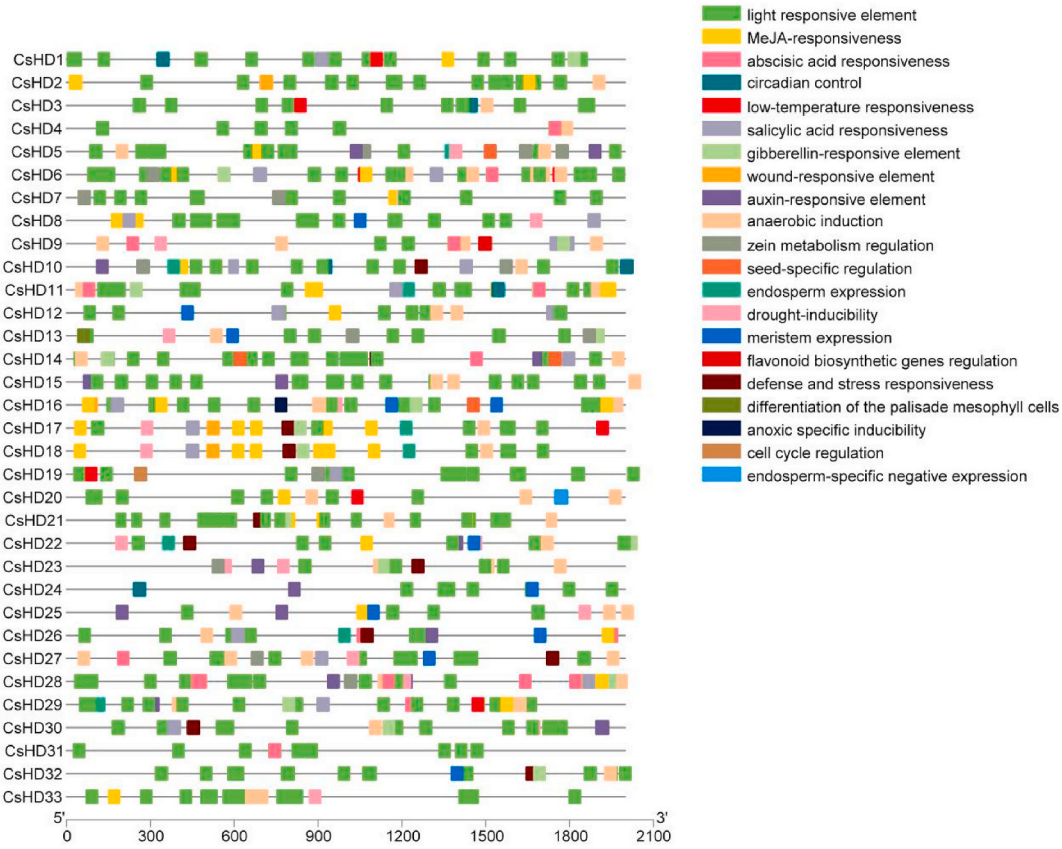


**Fig. 5.** Distribution of *cis*-acting elements in the *CsHD-Zip* genes. Distribution of the 21 *cis*-acting elements in the 33 CsHD-Zip promoter regions. The elements are represented by different symbols.

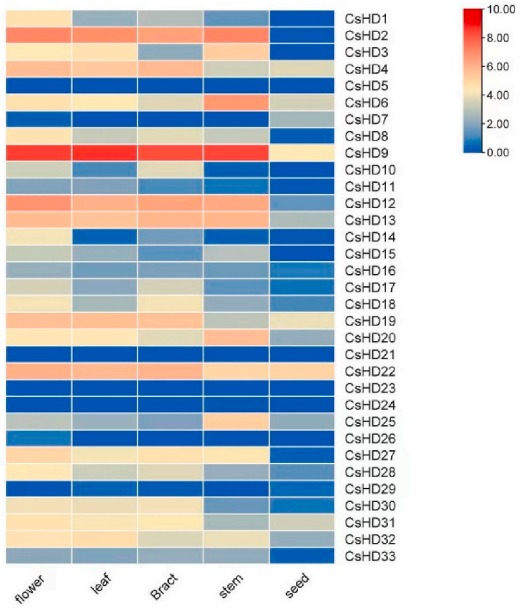


**Fig. 6.** The expression analysis of the *CsHD* gene. *CsHD* gene expression patterns in flower, leaf, bract, stem and seed.

*CsHD-22*, *CsHD-25*, *CsHD-27*, *CsHD-28*, *CsHD-30*, *CsHD-31*, *CsHD-32*, *CsHD-33,* 23 genes were highly expressed (Fig. 6). *CsHD-5*, *CsHD-1*, *CsHD-11*, *CsHD -16*, *CsHD-21*, *CsHD-23*, CsHD-24, CsHD-26, and CsHD-29, a total of nine genes, exhibited low expression in all five tissues. The highest expression of *CsHD-9* was observed in the stems, leaves, flowers, and bracts.

In this study, five *CsHD* genes were randomly selected for qRT-PCR validation (Fig. 7), and their expression of 5 genes at different tissue sites was consistent with the transcriptome data.

## 4. Discussion

*Cannabis sativa* L. has attracted the interest of scientists in recent years because of its abundance of phenolic terpenoids, flavonoids, terpenoids, and other important pharmacologically active components [22]. HD-Zip, a family of transcription factors unique to higher plants, plays a significant role in plant growth and development and is extensively involved in plant growth and development [23], abiotic stress response [24], and other life processes, including inflorescence development [25], flowering time regulation, lateral root development regulation, and hormone and stress resistance regulation [26–31]. Therefore, the present study analyzed cannabis HD-Zip transcription factors genome-wide using bioinformatics to predict the potential biological functions of HD-Zip transcription factors.

The correlation analysis website research of the cannabis *HD-Zip* gene family members indicated that the relative molecular weights ranged from 23264.32Da to 93147.87Da, the related amino acid counts ranged from 136 to 849aa, and their isoelectric points ranged from 4.54 to 9.04. The results indicated that all cannabis HD-Zip proteins were hydrophilic, but with different degrees of hydrophilicity. The expected subcellular localization of the hashish *HD-Zip* gene household participants confirmed that all proteins were positioned in the nucleus, except for the CsHD20 protein, which was positioned in the cytoplasm.

In this study, 33 members of the cannabis *HD-Zip* gene family were identified using bioinformatical analysis. According to the phylogenetic tree, the cannabis *HD-Zip* family was divided into four subfamilies: 14 in the *HD-Zip I* family, seven in the *HD-Zip II*, three in the *HD-Zip III* family, and nine in *HD-Zip IV*. Covariance analysis of cannabis with rice and Arabidopsis revealed 20 homologous gene pairs between the genes of the *HD-ZIP* family of cannabis and those of Arabidopsis and rice, of which 12 pairs were common to all three species. Homologous genes in different species maintain the same or similar functions during evolution, and the tissue expression pattern of genes is generally related to gene function [32]. Four homologous gene pairs were identified between cannabis and rice, while there were 16 additional homologous gene pairs of the HD-ZIP transcription factor family between cannabis and Arabidopsis, suggesting that homologous genes might have similar functions. To better verify transcriptome-related information, five randomly selected *CsHD* genes were identified by qRT-PCR. The results showed that the expression levels of the five genes in various tissues were consistent with the transcriptome-related data. These genes were similar to the Arabidopsis subfamily and could respond to light signal transduction and play a crucial role in the development of cannabis.

Cannabis *HD-Zip* gene family proteins contain multiple motifs; motifs 1 and 2 constitute the homeodomain and LZ structural domains, respectively, suggesting that motifs 1 and 2 are characteristic of hemp *HD-Zip* gene family proteins. Members of the same subfamily had similar motif types and numbers.

Medicinal cannabis contains a variety of natural plant products, and cannabinoids are the most important class of known compounds [33]. The process of cannabinoid metabolism and synthesis has been clarified, but the molecular regulation mechanism is unknown. Currently, the most studied cannabinoids are THC and CBD, both terpenoids and isomers of each other [34], and both active substances in cannabis with high levels and medicinal properties. In plant tissues, cannabinoids are synthesized in the carboxylated form. Cannabigerolic acid (CBGA), a common precursor of tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA), has been studied extensively. These are synthesized under the catalysis of THCA synthase and CBDA synthase, respectively, and are gradually decarboxylated into THC and CBD upon drying [35]. Hormones are important endogenous growth regulators in plants and are trace signaling molecules that have a non-negligible effect on plant growth, development, and the associated gene expression [36–38]. Different concentrations of exogenous hormones affect various plant tissues and regulate the endogenous hormone levels, which in turn regulate plant growth and development [39].

The study still has some limitations, such as the small scope involved, and further careful verification and research are needed to determine its biological significance. In the subsequent experimental study, interaction analysis of hemp HD-Zip transcription factors could be carried out, and recombinant plasmids containing target genes and promoters were introduced into Agrobacterium and injected into tobacco. After culture, firefly fluorescence and sea kidney fluorescence luminescence values were measured on the injected leaves. The transient expression technique is one of the methods to study plant gene function, which has become more and more mature over the years. The target gene was constructed into the transient expression vector through the enzyme digestion link and then transferred into the plant. The target gene enters the plant and will be expressed in a very short time, and the gene product will be detected, which makes this technology have the characteristics of short test cycle, good repeatability, and high conversion efficiency. Now this instantaneous expression system has been established in many plants. When the reporter vector and the expression vector are transferred into the plant host cell by agrobacterium transformation, if the target gene in the expression vector is specifically bound to the promoter sequence in the reporter vector, the expression of the tagged gene in the reporter vector will be enhanced, and the fluorescence signal will be amplified. The interaction between the target gene and the target promoter can be proved by detecting the fluorescence value of the enzyme label. Agrobacterium-mediated transient expression system can make gene expression efficient, fast and simple. In the present study, this system has been established in a variety of plants, including model plants Arabidopsis and tomato, and other plants such as grapes and potatoes [40]. The technical principle of the Agrobacterium-mediated transient expression system is to construct foreign genes into cointegrated vectors or bivectors, and transform them into Agrobacterium tumefaciens by means of infection, vacuuming, osmosis and injection, so that the host system can come into contact with Agrobacterium, thereby

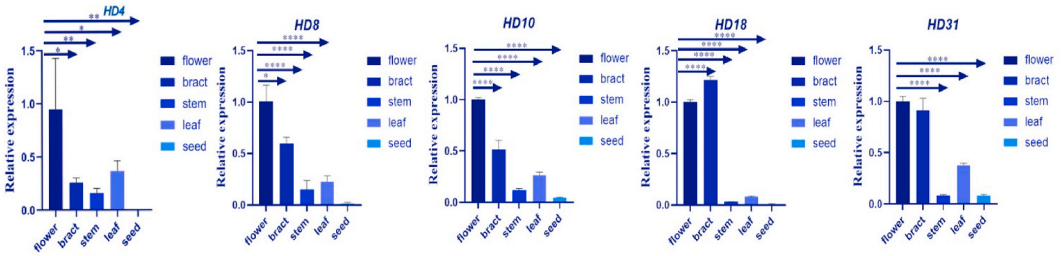


**Fig. 7.** Differential expression analysis of *CsHD* family genes in different tissue parts of *C.sativa*. RNA-seq data screening was used to identify the five genes: *CsHD4*, *CsHD8*, *CsHD10*, *CsHD18*, and *CsHD31*. *EF1a* served as the internal reference gene. On the x-axis, each organ and tissue is depicted, while the y-axis illustrates the comparative expression levels of genes. The asterisks (*) indicate significant differences ($p < 0.05$), the double asterisks (**) indicate significant differences ($p < 0.01$), the three asterisks (***) indicate a significant difference ($p < 0.001$), and the four asterisks (***) indicate a significant difference ($p < 0.0001$).

allowing T-DNA to enter the plant somatic cell nucleus and achieve short-term expression [41].

Different *cis*-acting elements in gene promoter regions usually imply that the gene can respond to the corresponding hormone or stress induction [42]. In the cannabis *HD-Zip* gene family, each gene promoter region contains *cis*-elements, with the largest number of light-responsive elements, followed by hormone-like *cis*-elements. Analysis of the expression patterns of the members of the cannabis *HD-Zip* gene family in different tissue sites revealed that 23 genes were highly expressed in flowers. We sequenced the transcriptome of five tissue sites of cannabis, including flowers, stems, leaves, bracts, and seeds, and found that the expression of 33 cannabis *HD-Zip* gene family members converged among the five tissue sites. Bioinformatics suggested that the expression of this gene family may be affected by hormonal and external environmental factors, that cannabis *HD-Zip* genes have tissue expression specificity, and that genes specifically highly expressed in flowers may have important regulatory roles in the development of female cannabis flowers. This provides the theoretical basis for the subsequent functional study of *HD-Zip* genes.

## 5. Conclusions

We systematically investigated the sequence characteristics, chromosomal localization, gene structure, conserved sequences, phylogeny, and differential gene expression of HD-Zip transcription factors in the cannabis genome. These results indicated that the *HD-Zip* gene in hemp has tissue-specific expression and may be influenced by hormones and external environmental factors. This study contributes to a more comprehensive understanding of the biological function of the cannabis gene family and lays a foundation for further study of *HD-Zip* gene function and the improvement of cannabis varieties.

## Funding statement

Talent training project supported by General project of Heilongjiang Postdoctoral Fund (LBH-Z21028), the Central Government for the Reform and Development of Local Colleges and Universities (ZYRCB2021008) and Heilongjiang Touyan Innovation Team Program (HLJTYTP2019001). Heilongjiang Provincial Natural Science Foundation of China (YQ2020H029)

**Supplementary Table S1**
The primers used for qRT-PCR in this study

| Gene name | Forward primer sequence (5′–3′) | Revers primer sequence (5′–3′) |
|---|---|---|
| *CsHD4* | CATTAATGGAATCGAGCCGT | AGAGAGGACTTGTAGTTCCG |
| *CsHD8* | CGAGAAGTTCGTTGGAGTTT | ACCAGTTTCCACACTCCTAA |
| *CsHD10* | GAACGCTTGTATGATGCTCT | TCTCTTAGCATCCCCTTCAG |
| *CsHD18* | AATGTTGCCTCATCAGGAAC | TGCATGCTCTTCAACGTTAA |
| *CsHD31* | AATGAAGACCCAGCATGAAC | GCATTTTCGAGCCTCAAATG |
| *EF1α* | ACCAAGATTGACAGGCGTTC | CCTTCTTCTCCACAGCCTTG |

## CRediT authorship contribution statement

**Zhan-Ping Zhang:** Formal analysis, Data curation, Conceptualization. **Zhen Wang:** Formal analysis, Data curation, Conceptualization. **Jia-Xin Lu:** Data curation, Conceptualization. **Song Yan:** Investigation. **Lian-Qing He:** Project administration. **Pan-Pan Wang:** Funding acquisition. **Chen Qin:** Project administration. **Wei-Chao Ren:** Resources. **Jiao Xu:** Validation. **Jian-Li Wu:** Writing – original draft, Visualization. **Xiu-Bo Liu:** Project administration, Methodology. **Wei Ma:** Writing – review & editing, Writing – original draft, Visualization, Software.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] C.R. L, G.G. M, P.C. M, et al., Homeoboxes in plant development, Biochim. Biophys. Acta 1442 (1) (1998) 1–19, https://doi.org/10.1016/s0167-4781(98)00119-5.
[2] F.D. Ariel, P.A. Manavella, C.A. Dezar, et al., The true story of the HD-Zip family, Trends Plant Sci. 12 (9) (2007) 419–426, https://doi.org/10.1016/j.tplants.2007.08.003.
[3] T. Long, M. Wagner, D. Demske, et al., Cannabis in Eurasia: origin of human use and Bronze Age trans-continental connections, Veg. Hist. Archaeobotany 26 (2) (2017) 245–258, https://doi.org/10.1007/s00334-016-0579-6.
[4] I.J. Flores-Sanchez, Y.H. Choi, R. Verpoorte, Metabolite analysis of Cannabis sativa L. by NMR spectroscopy, Methods Mol. Biol. 815 (2012) 363–375, https://doi.org/10.1007/978-1-61779-424-7_27.
[5] A.W. Zuardi, History of cannabis as a medicine: a review, Br. J. Psychiatry 28 (2) (2006) 153–157, https://doi.org/10.1590/s1516-44462006000200015.
[6] S. Chandra, H. Lata, M.A. ElSohly, et al., Cannabis cultivation: Methodological issues for obtaining medical-grade product, Epilepsy Behav. 70 (Pt B) (2017) 302–312, https://doi.org/10.1016/j.yebeh.2016.11.029.
[7] C. Muller, P. Morales, P.H. Reggio, Cannabinoid Ligands targeting TRP Channels, Front. Mol. Neurosci. 11 (2018) 487, https://doi.org/10.3389/fnmol.2018.00487.
[8] R. Sharif, A. Raza, P. Chen, et al., HD-ZIP gene family: potential roles in improving plant growth and regulating stress-responsive mechanisms in plants, Genes 12 (8) (2021) 1256, https://doi.org/10.3390/genes12081256.
[9] G. He, P. Liu, H. Zhao, et al., The HD-ZIP II transcription factors regulate plant architecture through the auxin pathway, Int. J. Mol. Sci. 21 (9) (2020) 3250, https://doi.org/10.3390/ijms21093250.
[10] K. Nakashima, Y. Ito, K. Yamaguchi-Shinozaki, Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses, Plant Physiol. 149 (1) (2009) 88–95, https://doi.org/10.1104/pp.108.129791.
[11] J.C. Harris, M. Hrmova, S. Lopato, et al., Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli, New Phytol. 190 (4) (2011) 823–837, https://doi.org/10.1111/j.1469-8137.2011.03733.x.
[12] M.J. Prigge, D. Otsuga, J.M. Alonso, et al., Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development, Plant Cell 17 (1) (2005) 61–76, https://doi.org/10.1105/tpc.104.026161.
[13] F.D. Ariel, P.A. Manavella, C.A. Dezar, et al., The true story of the HD-Zip family, Trends Plant Sci. 12 (9) (2007) 419–426, https://doi.org/10.1016/j.tplants.2007.08.003.
[14] S. Zhang, I. Haider, W. Kohlen, et al., Function of the HD-Zip I gene Oshox22 in ABA-mediated drought and salt tolerances in rice, Plant Mol. Biol. 80 (6) (2012) 571–585, https://doi.org/10.1007/s11103-012-9967-1.
[15] X. He, T. Wang, Z. Xu, et al., The cotton HD-Zip transcription factor GhHB12 regulates flowering time and plant architecture via the GhmiR157-GhSPL pathway, Commun. Biol. 1 (2018) 229, https://doi.org/10.1038/s42003-018-0234-0.
[16] K. Liu, X. Han, Z. Liang, et al., Genome-wide identification, Classification, and expression analysis of the HD-Zip transcription factor family in Apple (Malus domestica Borkh.), Int. J. Mol. Sci. 23 (5) (2022), https://doi.org/10.3390/ijms23052632.
[17] D.V. Romanov, G.I. Karlov, M.G. Divashuk, Developing Oligo Probes for chromosomes identification in hemp (cannabis sativa L.), Plants 11 (15) (2022), https://doi.org/10.3390/plants11151900.
[18] J.J. Zager, I. Lange, N. Srividya, et al., Gene networks Underlying cannabinoid and terpenoid accumulation in cannabis, Plant Physiol. 180 (4) (2019) 1877–1897, https://doi.org/10.1104/pp.18.01506.
[19] Z. Wang, Z. Zhang, P. Wang, C. Qin, L. He, L. Kong, W. Ren, X. Liu, W. Ma, Genome-wide identification of the NAC transcription factors family and regulation of metabolites under salt stress in Isatis indigotica, Int. J. Biol. Macromol. 240 (2023 Jun 15) 124436, https://doi.org/10.1016/j.ijbiomac.2023.124436.
[20] C. Chen, H. Chen, Y. Zhang, et al., TBtools: an Integrative Toolkit developed for interactive Analyses of Big biological data, Mol. Plant 13 (8) (2020) 1194–1202, https://doi.org/10.1016/j.molp.2020.06.009.
[21] M.M. Radwan, S. Chandra, S. Gul, et al., Cannabinoids, phenolics, terpenes and alkaloids of cannabis, Molecules 26 (9) (2021), https://doi.org/10.3390/molecules26092774.
[22] R.E. M, P.L. A, Constituents of cannabis sativa, Adv. Exp. Med. Biol. (2021) 1264, https://doi.org/10.1007/978-3-030-57369-0_1.
[23] J.V. Cabello, A.L. Arce, R.L. Chan, The homologous HD-Zip I transcription factors HaHB1 and AtHB13 confer cold tolerance via the induction of pathogenesis-related and glucanase proteins, Plant J. 69 (1) (2012) 141–153, https://doi.org/10.1111/j.1365-313X.2011.04778.x.
[24] Y. Zhao, Q. Ma, X. Jin, et al., A novel maize homeodomain-leucine zipper (HD-Zip) I gene, Zmhdz10, positively regulates drought and salt tolerance in both rice and Arabidopsis, Plant Cell Physiol. 55 (6) (2014) 1142–1156, https://doi.org/10.1093/pcp/pcu054.
[25] O. Son, Y.S. Hur, Y.K. Kim, et al., ATHB12, an ABA-inducible homeodomain-leucine zipper (HD-Zip) protein of Arabidopsis, negatively regulates the growth of the inflorescence stem by decreasing the expression of a gibberellin 20-oxidase gene, Plant Cell Physiol. 51 (9) (2010) 1537–1547, https://doi.org/10.1093/pcp/pcq108.
[26] D.A. Ré, C.A. Dezar, R.L. Chan, et al., Nicotiana attenuata NaHD20 plays a role in leaf ABA accumulation during water stress, benzylacetone emission from flowers, and the timing of bolting and flower transitions, J. Exp. Bot. 62 (1) (2011) 155–166, https://doi.org/10.1093/jxb/erq252.
[27] F. Ariel, A. Diet, M. Verdenaud, et al., Environmental regulation of lateral root emergence in Medicago truncatula requires the HD-Zip I transcription factor HB1, Plant Cell 22 (7) (2010) 2171–2183, https://doi.org/10.1105/tpc.110.074823.
[28] Y. Li, S. Zhang, R. Dong, et al., The grapevine homeobox gene VvHB58 influences seed and fruit development through multiple hormonal signaling pathways, BMC Plant Biol. 19 (1) (2019) 523, https://doi.org/10.1186/s12870-019-2144-9.
[29] S. Sawa, M. Ohgishi, H. Goda, et al., The HAT2 gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in Arabidopsis, Plant J. 32 (6) (2002) 1011–1022, https://doi.org/10.1046/j.1365-313x.2002.01488.x.
[30] C. Gu, Z.H. Guo, H.Y. Cheng, et al., A HD-ZIP II HOMEBOX transcription factor, PpHB.G7, mediates ethylene biosynthesis during fruit ripening in peach, Plant Sci. 278 (2019) 12–19, https://doi.org/10.1016/j.plantsci.2018.10.008.
[31] J. Wei, H. Choi, P. Jin, et al., GL2-type homeobox gene Roc4 in rice promotes flowering time preferentially under long days by repressing Ghd7, Plant Sci. 252 (2016) 133–143, https://doi.org/10.1016/j.plantsci.2016.07.012.
[32] J.X. Lu, J.Y. Sun, Z. Wang, W.C. Ren, N.N. Xing, M.Q. Liu, Z.P. Zhang, L.Y. Kong, X.Y. Su, X.B. Liu, W. Ma, In Silico genome-wide analysis of B3 transcription factors in cannabis sativa L, Cannabis Cannabinoid Res (2022 Dec 13), https://doi.org/10.1089/can.2022.0168.
[33] Y. Yang, R. Vyawahare, M. Lewis-Bakker, et al., Bioactive Chemical Composition of cannabis Extracts and cannabinoid Receptors, Molecules 25 (15) (2020), https://doi.org/10.3390/molecules25153466.
[34] N. Happyana, S. Agnolet, R. Muntendam, et al., Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR, Phytochemistry 87 (2013) 51–59, https://doi.org/10.1016/j.phytochem.2012.11.001.
[35] F.E. Dussy, C. Hamberg, M. Luginbühl, et al., Isolation of Delta9-THCA-A from hemp and analytical aspects concerning the determination of Delta9-THC in cannabis products, Forensic Sci. Int. 149 (1) (2005) 3–10, https://doi.org/10.1016/j.forsciint.2004.05.015.

[36] E. Shani, O. Yanai, N. Ori, The role of hormones in shoot apical meristem function, Curr. Opin. Plant Biol. 9 (5) (2006) 484–489, https://doi.org/10.1016/j.pbi.2006.07.008.
[37] G. Ma, A.K. Zelman, P.V. Apicella, et al., Genome-wide identification and expression analysis of homeodomain leucine zipper subfamily IV (HD-ZIP IV) gene family in cannabis sativa L, Plants 11 (10) (2022), https://doi.org/10.3390/plants11101307.
[38] S. Proietti, V. Scariot, S. De Pascale, et al., Flowering mechanisms and environmental stimuli for flower Transition: Bases for production Scheduling in Greenhouse Floriculture, Plants 11 (3) (2022), https://doi.org/10.3390/plants11030432.
[39] P. Grappin, D. Bouinot, B. Sotta, et al., Control of seed dormancy in Nicotiana plumbaginifolia: post-imbibition abscisic acid synthesis imposes dormancy maintenance, Planta 210 (2) (2000) 279–285, https://doi.org/10.1007/pl00008135.
[40] J.J. Choi, S.J. Klosterman, L.A. Hadwiger, A promoter from Pea gene DRR206 is Suitable to regulate an Elicitor-coding gene and Develop disease resistance, Phytopathology 94 (6) (2004 Jun) 651–660, https://doi.org/10.1094/PHYTO.2004.94.6.651.
[41] C.W. Simmons, J.S. VanderGheynst, S.K. Upadhyaya, A model of Agrobacterium tumefaciens vacuum infiltration into harvested leaf tissue and subsequent in planta transgene transient expression, Biotechnol. Bioeng. 102 (3) (2009 Feb 15) 965–970, https://doi.org/10.1002/bit.22118.
[42] P. Bhalothia, C. Sangwan, A. Alok, et al., PP2C-like promoter and its Deletion Variants are induced by ABA but not by MeJA and SA in Arabidopsis thaliana, Front. Plant Sci. 7 (2016) 547, https://doi.org/10.3389/fpls.2016.00547.