

Article

Genome-Wide Identification and Expression Analysis of HD-ZIP I Gene Subfamily in *Nicotiana tabacum*

Yueyue Li ¹, Bingchuan Bai ¹, Feng Wen ¹, Min Zhao ², Qingyou Xia ^{3,4,5} , Da-Hai Yang ^{6,*}  and Genhong Wang ^{1,4,5,*}

¹ State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing 400716, China

² Chongqing Institute of Tobacco Science, Chongqing 400716, China

³ Biological Science Research Center, Southwest University, Chongqing 400716, China

⁴ Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing 400716, China

⁵ Chongqing Key Laboratory of Sericulture, Southwest University, Chongqing 400716, China

⁶ Tobacco Breeding and Biotechnology Research Center, Yunnan Academy of Tobacco Agricultural Sciences, Key Laboratory of Tobacco Biotechnological Breeding, National Tobacco Genetic Engineering Research Center, Kunming 650021, China

* Correspondence: bioresearch2013@126.com (D.-H.Y.); wanggh168@swu.edu.cn (G.W.); Tel.: +86-23-6825-1573 (G.W.)

Received: 18 June 2019; Accepted: 28 July 2019; Published: 30 July 2019



Abstract: The homeodomain-leucine zipper (HD-Zip) gene family, whose members play vital roles in plant growth and development, and participate in responding to various stresses, is an important class of transcription factors currently only found in plants. Although the HD-Zip gene family, especially the HD-Zip I subfamily, has been extensively studied in many plant species, the systematic report on HD-Zip I subfamily in cultivated tobacco (*Nicotiana tabacum*) is lacking. In this study, 39 HD-Zip I genes were systematically identified in *N. tabacum* (Nt). Interestingly, that 64.5% of the 31 genes with definite chromosome location information were found to originate from *N. tomentosiformis*, one of the two ancestral species of allotetraploid *N. tabacum*. Phylogenetic analysis divided the *NtHD-Zip* I subfamily into eight clades. Analysis of gene structures showed that *NtHD-Zip* I proteins contained conserved homeodomain and leucine-zipper domains. Three-dimensional structure analysis revealed that most *NtHD-Zip* I proteins in each clade, except for those in clade η , share a similar structure to their counterparts in *Arabidopsis*. Prediction of cis-regulatory elements showed that a number of elements responding to abscisic acid and different abiotic stresses, including low temperature, drought, and salinity, existed in the promoter region of *NtHD-Zip* I genes. The prediction of *Arabidopsis* ortholog-based protein–protein interaction network implied that *NtHD-Zip* I proteins have complex connections. The expression profile of these genes showed that different *NtHD-Zip* I genes were highly expressed in different tissues and could respond to abscisic acid and low-temperature treatments. Our study provides insights into the evolution and expression patterns of *NtHD-Zip* I genes in *N. tabacum* and will be useful for further functional characterization of *NtHD-Zip* I genes in the future.

Keywords: HD-Zip I; gene subfamily; *Nicotiana tabacum*; abiotic stress; expression analysis

1. Introduction

As regulatory proteins for the transcriptional activation or repression of target genes, transcription factors (TFs) are the main players in development and differentiation in eukaryotic organisms [1]. The TFs regulate gene expression by specifically interacting with cis-regulatory regions [2]. The homeodomain-leucine zipper (HD-Zip) gene family is a relatively pivotal class of TFs, present

only in the plant kingdom [3,4]. This gene family has been identified in multiple plant species, such as soybean (*Glycine max*), tomato (*Solanum lycopersicum*), grape (*Vitis vinifera*), rice (*Oryza sativa*), maize (*Zea mays*), and wheat (*Triticum aestivum*) [5–10]. Based on their corresponding gene structure, including some specific cis-elements, conserved sequences, and biological function, they can be divided into four subfamilies: HD-Zip I to HD-Zip IV [4]. The number of members in each HD-Zip subfamily varies. All HD-Zip proteins from HD-Zip I, II, III, and IV subfamilies contain HD and leucine zipper (LZ) motifs [11]. The HD can specifically bind to DNA, and the LZ can act as a dimerization motif. Members of HD-Zip I contain these two basic motifs. Besides the HD and LZ in the HD-Zip family members, other motifs have been described and functionally characterized in HD-Zip II, III, and IV. For example, some extra motifs called CPSCE (consisting of five conserved amino acid sequences Cys, Pro, Ser, Cys, and Glu) and N-terminal conserved sequence have been shown in the HD-Zip II subfamily [12,13]. The START (star-related lipid transfer) motif, a steroidogenic acute regulatory protein related to lipid transfer, was found in HD-Zip III and IV subfamilies [14,15]. It was reported that HD-Zip I and II proteins can recognize a similar pseudopalindromic sequence CAAT(C/G)ATTG [4,12,13], whereas HD-Zip III and IV proteins recognize sequences GTAAT(G/C)ATTAC and TAAATG(C/T)A, respectively [11,14,15].

The HD-Zip I subfamily has been shown to play important roles in plant growth, development, de-etiolation, and response to abiotic and biotic stresses [16–19]. There are 17, 13, 14, and 22 HD-Zip I members in *Arabidopsis thaliana*, *Cucumis sativus*, *O. sativa*, and *S. lycopersicum*, respectively [3,6,8,20]. In *Arabidopsis*, the ectopic expression of one HD-Zip I member *ATHB12* resulted in larger leaves with enlarged cells, strongly suggesting its positive effect on leaf growth [21]; another HD-Zip I member *ATHB1* was involved in cell wall composition and elongation [22]. In *C. sativus*, HD-Zip I member *CsGL1* was identified to function in trichome formation [23]. Ariel et al. demonstrated that *Medicago truncatula* HD-Zip I TF *HB1* was required for the adaptive developmental response of lateral roots [24]. Overexpression of rice HD-Zip I genes, *Oshox12*, and *Oshox14*, led to reduced panicle length and a dwarf phenotype [25]. In addition to function in plant growth and development, HD-Zip I proteins were shown to play important roles in abiotic stress responses. For instance, *ATHB7* and *ATHB12*, which were strongly up-regulated by water-deficit and abscisic acid, function as positive regulators of PP2C in *Arabidopsis* [26]. *HaHB4* was involved in the regulation of tolerance to water deficit through ethylene-mediated senescence in sunflower [27]. Both rice HD-Zip I members *Oshox22* and *OsSLI1* function in abscisic acid (ABA)-mediated drought and salt tolerance [28,29]. Overexpression of *Zmhdz10* or *TaHDZipI-5* could increase plant tolerance to low temperature, drought, or salt stress [30,31]. Soybean HD-Zip I gene *Gshdz4* was significantly up-regulated by alkali stress, and further results showed that it enhanced bicarbonate tolerance and response to osmotic pressure [32]. It was also reported that over-expressing one chickpea HD-Zip I gene, *CaHDZ12*, resulted in improved tolerance to abiotic stress, and *CaHDZ12* expression was regulated by *CaWRKY70* [33]. The expression of *Nicotiana attenuata* HD-Zip I member *NaHD20* was also shown to be positively correlated with ABA accumulation in leaves during water deficit and the expression of some dehydration-responsive genes [34]. In addition, HD-ZIP I members have also been shown to play a defensive role against some biotic stresses. An *Arabidopsis* mutant with constitutively overexpressed *ATHB13* exhibited resistance to downy mildew (*Hyaloperonospora arabidopsidis*) and green peach aphid (*Myzus persicae*) [35]. A HD-ZIP I protein in pepper was characterized as having a positive role in regulating resistance to *Ralstonia solanacearum* infection [36].

Although HD-ZIP I genes have been widely studied in other plant species, there are no reports concerning HD-ZIP I in allotetraploid *N. tabacum*. In this study, a genome-wide identification of HD-ZIP I genes was performed in *N. tabacum* (Nt). The phylogenetic relationship, conserved domains, genome organization, and gene structure of NtHD-ZIP I members were investigated. Also, possible cis-acting elements in the promoter sequences of NtHD-ZIP I genes and three-dimensional modeling of NtHD-ZIP I proteins were analyzed. Furthermore, expression profiles of NtHD-ZIP I genes were investigated in different tissues and under low temperature or ABA treatment using quantitative real-time PCR

(qRT-PCR). This study provided essential information concerning the *NtHD-Zip I* subfamily genes in *N. tabacum* and enhanced our understanding of *HD-Zip I* genes in plants.

2. Materials and Methods

2.1. Sequence Search and Identification of *HD-Zip I* Subfamily

Predicted tobacco coding sequences and protein sequences were downloaded from the Sol Genomics Network (<http://solgenomics.net/organism/>). Sequences of *Arabidopsis*, *S. lycopersicum*, *Camellia sinensis*, *O. sativa*, *Z. mays*, and *Manihot esculenta* were downloaded from PlantGDB (<http://www.plantgdb.org/>) and were named according to previous reports (Table S1). Prediction of subcellular localization was based on WoLF (<https://www.genscript.com/wolf-psort.html>). Each *NtHD-ZIP I* gene was mapped to its corresponding chromosome and ancestral species based on the reported tobacco genome [37–39].

2.2. Phylogenetic, Gene Structure, and Conserved Domain Analyses of *HD-Zip Class I* Subfamily of Tobacco

MEGA 7.0 was used to construct the phylogenetic tree of *HD-Zip I* family proteins in *S. lycopersicum*, *C. sinensis*, *O. sativa*, *Z. mays*, *M. esculenta*, and *Arabidopsis* using the neighbor-joining algorithm with 1000 bootstrap replications. Gene structure was obtained by online program Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn>) using *NtHD-Zip I* subfamily cDNA and genomic sequences. The online site Multiple Em for Motif Elicitation (MEME, <http://meme.nbcr.net/meme/cgibin/meme.cgi>) was used to identify conserved domains of *NtHD-Zip I* proteins, with a total of 20 motifs investigated, then processed through an online website (<http://www.omicshare.com/>). Multiple sequence alignments were done together by ClustalW and Genedoc (<https://www.psc.edu/biomed/genedoc/>) [40].

2.3. Analysis of Promoter, Prediction of Three-Dimensional Modeling, and Interacting Networks of *HD-Zip Class I* Proteins between *N. tabacum* and *Arabidopsis*

About 2000 bp of genomic DNA sequence at upstream of the starting codon of *NtHD-Zip I* genes were selected for analysis of *cis*-elements in promoter through the online website Genomatix (<http://www.genomatix.de/solutions/genomatix-software-suite.html>). In this study, ABA-, low temperature-, drought stress-, and salinity stress-regulated elements were selected to analyze promoter *cis*-elements. Three-dimensional modeling of *HD-Zip I* proteins was performed by Phyre2 server (Protein Homology/analogy Recognition Engine, <http://www.sbg.bio.ic.ac.uk/phyre2>) [41]. The prediction of the interacting networks of proteins was constructed in STRING (https://string-db.org/?tdsourcetag=s_pctim_aiom).

2.4. Plant Material and Stress Treatment

The cultivated *N. tabacum* variety Honghuadajingyuan (HHDJY) was used to detect the expression of *NtHD-Zip I* subfamily genes in various tissues. Seeds were sterilized in 75% ethanol for 30 s, washed 1–2 times with sterile water, then sterilized in 10% sodium hypochlorite for 8–10 min, washed 4–5 times with sterile water and thereafter seeded on MS solid medium, and cultured at 25 °C. The seedlings were cultured in 16/8 h of light/dark photoperiod at 25 °C. Roots, stems, and leaves were collected from tobacco seedlings about 2 months after germination. Furthermore, flowers and seeds were collected from mature tobacco in soil. Roots, stems, leaves, flowers, and seeds of tobacco were temporarily stored in liquid nitrogen and then stored in a –80 °C refrigerator for the detection of tissue expression profiles.

For low-temperature stress, tobacco seedlings with four leaves (approximately 2 months after germination) grew under 16/8 h of light/dark photoperiod at 25 °C in soil, were placed in a 4 °C incubator for low-temperature treatment. Tobacco leaves at time points of 0, 1, 6, 12, and 24 h were sampled for RNA extraction and quantitative analysis. For ABA treatment, seedlings with four leaves (approximately 2 months after germination) grew in the same conditions as those of low-temperature stress, and the leaves of plants were sprayed with 200 µM ABA; leaves were collected at 0, 1, 6, 12,

and 24 h after the ABA treatment. Three biological replicates were taken at each time point under 4 °C and ABA treatments. All 30 treated plants were used for sample harvesting and the following qPCR analysis.

2.5. RNA Extraction and Quantitative RT-PCR

Total RNA in tobacco was extracted through Trizol reagent (Invitrogen). The cDNA was synthesized by extracting RNA using GoScript™ Reverse Transcription System (Promega). Gene-specific primers were designed in online tools (PrimerQuest Tool, <https://sg.idtdna.com/Primerquest/Home/Index>). Primers of target and reference genes (*NtEF1α*) were used for quantitative RT-PCR (Table S2) using Takara's SYBR Premix Ex Taq™II.

The real-time PCR analyses were made by using a qTOWER2.2 real-time PCR system (Analytik Jena AG, Jena, Germany), and the procedure was as follows: denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 10 s and annealing/extension at 60 °C for 1 min [42]. Each sample was run through three technical repeats, and quantitative results were analyzed using the $2^{-\Delta\Delta C_t}$ algorithm [43]. Pictures were drawn using GraphPad Prism 6. The heat map used for tissue expression profiling was made by an online website (<http://www.omicshare.com/>) based on row scale-transformed expression values.

2.6. Statistical Analysis

GraphPad Prism 6 was utilized for all statistical analyses. ****, ***, **, and * indicate significant differences compared to the control (0 h) at $p < 0.0001$, $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

3. Results

3.1. Identification of HD-Zip Class I Members in *N. tabacum*

To identify the *NtHD-Zip* I subfamily gene in *N. tabacum*, a BLASTP search was implemented against the tobacco genome database using HD-Zip I protein sequences from multiple plants, and conserved HD and LZ domains were further analyzed following BLASTP search. A total of 39 *NtHD-Zip* I candidate genes were identified in tobacco and named *NtHDZI1*–*NtHDZI39* (Table 1). The length of the deduced protein sequences of the *NtHD-Zip* I members ranged from 172 (*NtHDZI38* and *NtHDZI39*) to 361 (*NtHDZI3*) amino acids. Results of subcellular localization prediction indicated that *NtHD-Zip* I members were all located in the nucleus (Table 1).

Table 1. Basic characteristics of tobacco homeodomain-leucine zipper (HD-ZIP) I genes.

Gene Name	Length of Deduced Protein (aa)	Number of Exons	Subcellular Location	Chromosome	Origin
<i>NtHDZI1</i>	342	3	nucl: 13.5, cyto_nucl: 7.5	Chr13	T
<i>NtHDZI2</i>	342	3	nucl: 13.5, cyto_nucl: 7.5	Chr13	T
<i>NtHDZI3</i>	361	3	nucl: 14	Chr13	T
<i>NtHDZI4</i>	301	2	nucl: 14	Chr13	T
<i>NtHDZI5</i>	289	3	nucl: 14	Chr12	T
<i>NtHDZI6</i>	290	3	nucl: 14	Chr12	T
<i>NtHDZI7</i>	270	3	nucl: 14	unknown	unknown
<i>NtHDZI8</i>	302	3	nucl: 13.5, cyto_nucl: 7.5	Chr24	T
<i>NtHDZI9</i>	270	3	nucl: 14	unknown	unknown
<i>NtHDZI10</i>	260	2	nucl: 14	Chr12	T
<i>NtHDZI11</i>	308	3	nucl: 13.5, cyto_nucl: 7.5	Chr17	unknown
<i>NtHDZI12</i>	308	3	nucl: 13.5, cyto_nucl: 7.5	Chr3	S
<i>NtHDZI13</i>	328	3	nucl: 14	Chr4	T

Table 1. Cont.

Gene Name	Length of Deduced Protein (aa)	Number of Exons	Subcellular Location	Chromosome	Origin
<i>NtHDZI14</i>	300	3	nucl: 14	Chr20	S
<i>NtHDZI15</i>	298	3	nucl: 14	Chr24	T
<i>NtHDZI16</i>	289	3	nucl: 14	unknown	unknown
<i>NtHDZI17</i>	332	3	nucl: 14	unknown	unknown
<i>NtHDZI18</i>	312	3	nucl: 14	Chr12	T
<i>NtHDZI19</i>	289	3	nucl: 14	unknown	unknown
<i>NtHDZI20</i>	245	2	nucl: 14	Chr9	T
<i>NtHDZI21</i>	288	3	nucl: 14	Chr20	S
<i>NtHDZI22</i>	282	3	nucl: 13.5, cyto_nucl: 7.5	Chr17	unknown
<i>NtHDZI23</i>	237	2	nucl: 14	Chr17	unknown
<i>NtHDZI24</i>	226	3	nucl: 14	Chr2	T
<i>NtHDZI25</i>	226	3	nucl: 14	Chr2	T
<i>NtHDZI26</i>	210	3	nucl: 14	Chr17	unknown
<i>NtHDZI27</i>	218	3	nucl: 14	Chr17	unknown
<i>NtHDZI28</i>	210	3	nucl: 14	Chr17	unknown
<i>NtHDZI29</i>	212	3	nucl: 14	Chr17	unknown
<i>NtHDZI30</i>	213	3	nucl: 14	unknown	unknown
<i>NtHDZI31</i>	212	3	nucl: 14	Chr14	T
<i>NtHDZI32</i>	250	2	nucl: 14	Chr7	S
<i>NtHDZI33</i>	226	2	nucl: 14	unknown	unknown
<i>NtHDZI34</i>	221	2	nucl: 14	unknown	unknown
<i>NtHDZI35</i>	225	2	nucl: 14	Chr4	T
<i>NtHDZI36</i>	225	2	nucl: 14	Chr23	T
<i>NtHDZI37</i>	225	2	nucl: 14	Chr23	T
<i>NtHDZI38</i>	172	1	nucl: 14	Chr2	T
<i>NtHDZI39</i>	172	1	nucl: 14	Chr2	T

“T” and “S” represent the two origins of tobacco, *N. tomentosiformis* and *N. sylvestris*, respectively. “Unknown” represented no judgment could be made on the origins of *NtHDZI* genes based on the available genome sequences.

There were 37 *NsylHD-Zip I* genes found in ancestral species *N. sylvestris*, and 39 *NtomHD-Zip I* genes in *N. tomentosiformis* according to two reported ancestral genomic databases (*N. sylvestris* and *N. tomentosiformis*) (Table S3) [39]. However, the complete genomic sequences *NtHDZI1–6, 8, 10, 13, 15, 18, 20, 24, 25, 31*, and *35–39* only appeared in chromosomes of the T-genome (*tomentosiformis*) of *N. tabacum*, and *NtHDZI12, 14, 21*, and *32* were exclusively in chromosomes of the S-genome (*sylvestris*) of *N. tabacum*. The genomic origins of *NtHDZI7, 9, 11, 16, 17, 19, 22, 23, 26–30, 33*, and *34* could not be determined using the current genomic maps [37,38]. Approximately 80% of *NtHD-Zip I* genes in tobacco could be located on the corresponding chromosome. In detail, there were four *NtHD-Zip I* genes on each chromosome 2, 12, and 13. Chromosome 17 contained the maximal number of *NtHD-Zip I* genes, seven in total. Each of chromosomes 4, 20, 23, and 24 contained two *NtHD-Zip I* genes respectively. Each of chromosomes 3, 7, 9, and 14 contained only one *NtHDZ I* gene (Table 1).

3.2. Phylogenetic and Evolutionary Analysis of HD-Zip Class I Proteins in *N. tabacum*

A phylogenetic tree was constructed to assess the evolutionary relationships of *NtHD-Zip I* proteins among publicly available HD-Zip I proteins from other plants, including *A. thaliana*, *S. lycopersicum*, *O. sativa*, *Cassava*, *Maize*, and *C. sinensis* (Figure 1). Similar to previous results [3], there were 11 clades identified in the phylogenetic tree: α , $\beta 1$, $\beta 2$, γ , δ , $\epsilon 1$, $\epsilon 2$, η , $\varphi 1$, $\varphi 2$, and ζ . *NtHD-Zip I* members appeared in eight clades: α , $\beta 1$, $\beta 2$, γ , δ , $\epsilon 1$, η , and $\varphi 1$. No *NtHD-Zip Class I* members were in clades $\epsilon 2$, $\varphi 2$, and ζ . Additionally, *NtHD-Zip Class I* proteins were used alone to construct a phylogenetic tree, the classification results of which were consistent with that of the phylogenetic tree including HD-Zip Class I members from other plants (Figure 2A). Clade α contained the maximal number (eight) of *NtHD-Zip Class I* members: *NtHDZI11, 12, 14, 15, 19, 20, 22*, and *23*. Each Clade $\beta 1$, $\epsilon 1$, and $\varphi 1$ had the fewest number (two) of *NtHD-Zip I* members.

In addition, similar to a phylogenetic tree for HD-Zip Class I members [20], clade β was divided into $\beta 1$ and $\beta 2$ sub-clades in the current analysis. Clade ϵ was also divided into sub-clades $\epsilon 1$ and $\epsilon 2$, and clade φ was divided into $\varphi 1$ and $\varphi 2$ (Figure 1). As found in the previous report, members of clade ζ belonged to monocots (rice and maize); members of clades $\beta 2$, γ , $\epsilon 2$, η , $\varphi 1$, and $\varphi 2$ were only from dicots (*Arabidopsis*, tomato, *C. sinensis*, and *M. esculenta*) [20].

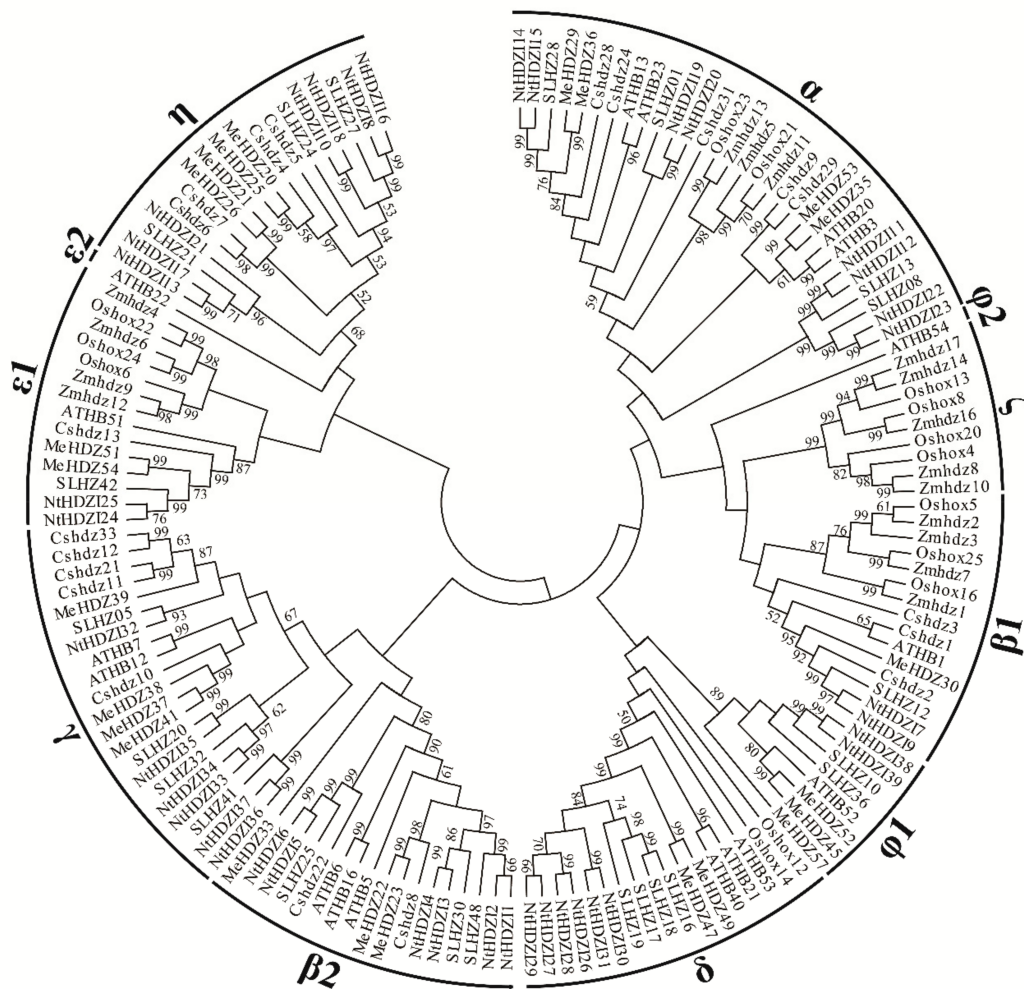


Figure 1. Phylogenetic analysis of HD-Zip I proteins from tobacco and other plants. The phylogenetic tree was constructed using 39, 17, 22, 14, 23, 16, and 20 HD-Zip I protein sequences from *N. tabacum*, *A. thaliana*, *S. lycopersicum*, *O. sativa*, *Cassava*, *Maize*, and *C. sinensis*, respectively. Plant HD-Zip I members were divided into eleven clades.

3.3. Genomic Structure, Conserved Domain, and Motif Analysis of HD-Zip Class I Proteins in *N. tabacum*

The exon–intron structure analysis showed that the *NtHD-Zip* Class I genes had either two or three exons, except for *NtHDZI38* and *NtHDZI39*, which had no intron (Figure 2B). Results of multiple sequence alignments showed that all *NtHD-Zip* I proteins contained conserved HD and LZ domains (Figure S1). Predicted motif (Figure 2C; Figure S2; Table S4) analysis revealed that motif 1 and motif 3 encoded the HD domain, and motif 2 encoded the LZ domain. Distribution features of the 20 predicted motifs in deduced proteins were consistent with the phylogenetic analysis (Figure 2).

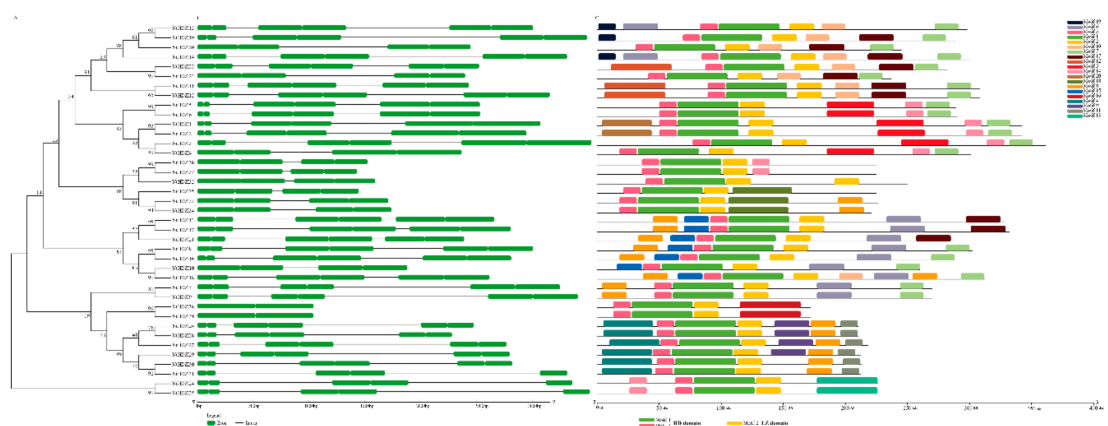


Figure 2. Gene structure, motif prediction of NtHD-Zip I proteins in *N. tabacum*. (A) The phylogenetic tree is constructed using the full-length amino acid sequences from *N. tabacum* alone by the NJ method. (B) Intron/exon structure of *NtHDZI* genes. Introns and exons are represented by black lines and green boxes, respectively. (C) Conserved motifs distribution of NtHDZI proteins. All motif predictions are obtained through the MEME online website (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>). Different color boxes at the top right-hand corner represent different motifs detected. The characteristic motifs, motif 1 and motif 3, corresponding to HD domain and motif 2 corresponding to LZ domain, are highlighted at the bottom.

3.4. Promoter Analysis of the HD-Zip Class I Genes in *N. tabacum*

To assess the possible response patterns of the *NtHD-Zip* I subfamily gene to different stress treatments, cis-acting elements including components related to stress treatment, such as the ABA-, low temperature-, drought-, and salinity-regulated elements were predicted in the promoter region of these genes (Table 2; Figure S4). The type and number of cis-components, which should respond to salinity, drought, and ABA in the promoter region of each gene, are listed in Table 2. The maximum number of such elements was found in *NtHDZI28* (107). Furthermore, the number of predicted cis-acting elements varied from 168 (*NtHDZI34*) to 237 (*NtHDZI28*) (Figure S3; Table S5).

The promoter sequences of both *NtHDZI16* and *NtHDZI30* contained the fewest number of low temperature-regulated elements, with only 43 (Table S5). However, *NtHDZI22* contained the maximal number of low-temperature-regulated components with 69, followed by *NtHDZI21* with 67. There were more elements responding to ABA compared with those to low temperature. *NtHDZI28*, 13, 38, and 11 ranked in the top four for ABA-responding-elements, with 154, 155, 156, and 161, respectively. *NtHDZI23* contains the fewest number (107) of ABA-related components.

Arabidopsis homeobox protein (AHBP), a classical cis-acting element, can respond to salinity, drought, and ABA [19,26,44–46], and was abundantly distributed in each of the NtHD-Zip Class I members. For example, there were 59 AHBP elements in *NtHDZI11*, which was the most compared to other *NtHD-Zip* I genes (Table S5). DNA binding with one finger (DOFF), responding to low-temperature stress [47], was another highly distributed element in NtHD-Zip Class I members—the promoter region of *NtHDZI3* contained the most number (24) of DOFF elements (Table S5). GT-box element (GTBX) is abundantly distributed among all the tested drought stress elements [48] and the promoter region of *NtHDZI12* has the maximum number of GTBX element. The numbers of cis-elements of *NtHDZI7* and *NtHDZI9* belonging to clade β 1 were extremely similar (Figure 1), with 178 and 172, respectively. Likewise, the number of cis-elements in *NtHDZI38* and *NtHDZI39* of clade ϕ 1 were 211 and 203, respectively (Figure 1; Table S5).

Table 2. The number of stress-responsive cis-acting elements in the promoter region of each *NtHDZI* gene.

Gene Name	Respond to ABA (AREF, GLKF, LEGB, TCPF, SWNS)	Respond to Low Temperature (CGCG, DOFF, MADS, OCSE)	Respond to Salinity (ASRC, NIGS, NTMF)	Respond to Drought (CAAT, GARP, GTBX, PNRE)	Respond to Salinity, Drought, and ABA (ABRE, AHBP, MIIG, MYBL, MYBS, NACF)	Respond to Low Temperature, Salinity, and Drought (CCAF, MYCL)	Respond to Low Temperature, Salinity, and Drought (CE1F)	Respond to Low Temperature, Salinity, Drought, and ABA (DREB, EPFF, GCCF, WBXF)	Respond to Salinity and ABA (GBOX, JARE, TELO)
<i>NtHDZI1</i>	12	34	14	29	95	18	1	7	3
<i>NtHDZI2</i>	12	35	11	30	96	13	1	9	1
<i>NtHDZI3</i>	9	34	9	31	65	18	1	13	5
<i>NtHDZI4</i>	19	26	16	33	78	20	0	12	5
<i>NtHDZI5</i>	16	34	8	27	73	21	1	6	6
<i>NtHDZI6</i>	8	31	13	31	88	5	0	7	4
<i>NtHDZI7</i>	14	19	13	22	80	9	3	14	4
<i>NtHDZI8</i>	21	20	16	31	96	15	0	15	6
<i>NtHDZI9</i>	12	25	11	23	66	14	2	12	7
<i>NtHDZI10</i>	12	29	15	22	82	15	0	14	6
<i>NtHDZI11</i>	11	20	14	27	106	21	0	16	7
<i>NtHDZI12</i>	7	26	10	38	95	19	0	9	3
<i>NtHDZI13</i>	19	24	11	19	86	22	0	13	10
<i>NtHDZI14</i>	17	33	15	19	75	24	0	7	8
<i>NtHDZI15</i>	20	29	11	19	73	21	0	6	12
<i>NtHDZI16</i>	17	24	13	28	95	17	0	6	1
<i>NtHDZI17</i>	16	20	11	28	72	29	1	8	5
<i>NtHDZI18</i>	9	24	16	24	77	23	0	17	17
<i>NtHDZI19</i>	15	25	11	18	74	17	0	4	8
<i>NtHDZI20</i>	11	31	16	18	92	21	3	9	8
<i>NtHDZI21</i>	18	37	10	29	75	21	0	9	7
<i>NtHDZI22</i>	17	37	17	29	78	23	1	9	4
<i>NtHDZI23</i>	9	29	16	32	64	19	1	11	4
<i>NtHDZI24</i>	11	25	17	34	98	13	0	13	4
<i>NtHDZI25</i>	13	32	20	27	82	11	1	9	6
<i>NtHDZI26</i>	12	27	17	31	79	7	1	9	7
<i>NtHDZI27</i>	19	23	14	35	103	13	0	10	6
<i>NtHDZI28</i>	10	30	14	39	107	21	0	11	5
<i>NtHDZI29</i>	17	29	10	24	86	10	0	11	8
<i>NtHDZI30</i>	13	19	11	23	81	15	0	9	5
<i>NtHDZI31</i>	6	20	16	28	93	14	0	14	9
<i>NtHDZI32</i>	18	29	13	25	92	11	0	8	7
<i>NtHDZI33</i>	12	12	7	20	77	23	0	13	9
<i>NtHDZI34</i>	6	24	13	20	69	18	1	13	4
<i>NtHDZI35</i>	15	25	17	25	89	20	2	6	10
<i>NtHDZI36</i>	16	32	10	21	82	20	0	11	7
<i>NtHDZI37</i>	28	29	8	27	83	18	0	7	14
<i>NtHDZI38</i>	23	26	5	24	96	15	0	12	10
<i>NtHDZI39</i>	18	29	11	29	86	14	0	7	9

The numbers in the table represent the number of cis-acting elements. ABRE: ABA response elements; AHBP: Arabidopsis homeobox protein; AREF: Auxin response element; CCAF: Circadian control factors; DREB: Dehydration responsive element binding factors; EPFF: EPF-type zinc finger factors; GBOX: Plant G-box/C-box bZIP proteins; GCCF:GCC box family; GLKF: Golden2-like factors; JARE: Jasmonate response element; LEGB: Legumin Box family; MIIG: MYB IIG-type binding sites; MYBL:MYB-like proteins; MYCL: Myc-like basic helix-loop-helix binding factors; NACF: Plant specific NAC; SWNS: Secondary wall NACs; TCPF: DNA-binding proteins with the plant specific TCP-domain; TELO: Telo box (plant interstitial telomere motifs); WBXF: W Box family; CGCG:Calmodulin binding/CGCG box binding proteins; DOFF: DNA binding with one finger (DOF); CE1F:Coupling element 1 binding factors; MADS: MADS box proteins; OCSE: Enhancer element first identified in the promoter of the octopine synthase gene (OCS) of the *Agrobacterium tumefaciens* T-DNA; ASRC: AS1/AS2 repressor complex; MYBS: MYB proteins with single DNA binding repeat; NIGS: NACL-inducible genes; NTMF: NAC factors with transmembrane motif; CAAT: CAAT binding factors; GARP: MYB-related DNA binding proteins(Golden2, ARR, Psr); GTBX: GT-box elements.

3.5. Prediction of Three-Dimensional Modeling and Interaction Network of HD-Zip Class I Proteins in *N. tabacum* and *Arabidopsis*

To better understand the structural characteristics and interaction network of the HD-Zip I subfamily, three-dimensional models of all the NtHD-Zip I proteins were constructed using the Phyre2 server (Figure 3), and the protein–protein interactions of the HD-Zip I subfamily between *N. tabacum* and *Arabidopsis* were predicted using STRING (Figure 4).

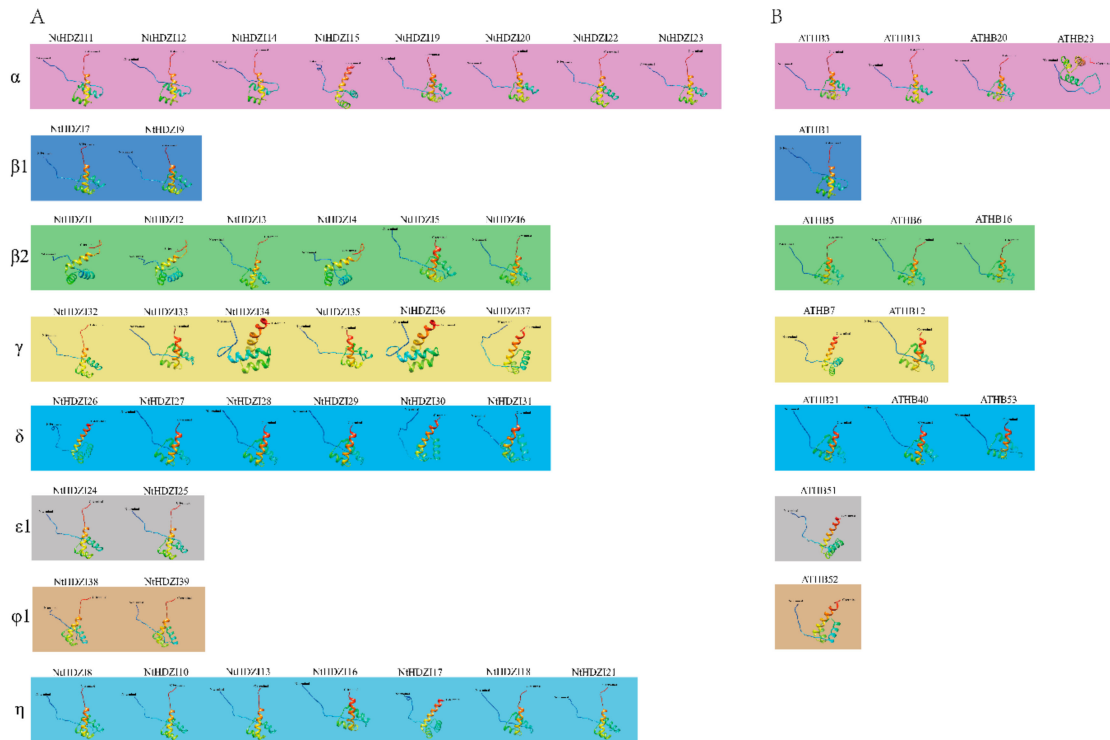


Figure 3. Three-dimensional modeling of HD-Zip I proteins in *N. tabacum* and *Arabidopsis*. The structure of NtHDZI proteins (A) and ATHBI proteins (B), with a confidence level > 90%, is shown and the activated sites are highlighted in light blue, yellow, and light green.

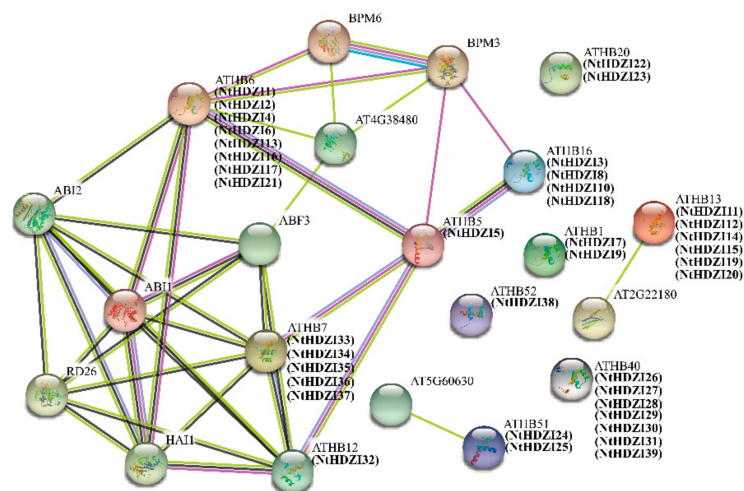


Figure 4. The prediction of the interaction network of NtHD-Zip I proteins based on the interactions of their orthologs in *Arabidopsis*.

All NtHD-Zip I proteins contained the α -helix and coil structure [49]. Most NtHD-Zip I members shared a similar three-dimensional structure, especially for those clustered in the same clade.

The predicted structures of NtHD-Zip proteins in each sub-clade were similar to their counterpart(s) in *Arabidopsis*, except for clade η members.

The protein–protein interactions of HD-Zip I subfamily between *N. tabacum* and *Arabidopsis* were determined using STRING (<http://string-db.org/>). All NtHD-Zip Class I proteins appeared in the known interaction network of *Arabidopsis* HD-Zip Class I proteins according to the prediction, implying that they had complex connections. The results showed that the protein structure and sequence of ATHB7 were similar to five NtHD-Zip I proteins (NtHDZI33–37), and ATHB12 was similar to NtHDZI32, suggesting that these NtHD-Zip I members may be involved in both regulating growth or development, and responding to water deficit [44]. In addition, like their counterpart ATHB20, NtHDZI22 and 23 might play a role in regulating ABA sensitivity and seed dormancy [50]. There were six *NtHD-Zip* I genes (*NtHDZI11*, 12, 14, 15, 19, and 20) similar to *ATHB13*, which may represent a component of the sucrose-signaling pathway [18]. *ATHB16* was shown to be a regulator of flowering time in response to photoperiod [17]. According to the predicted interaction network, NtHDZI3, 8, 10, and 18 might play a similar role in *N. tabacum*. Similarly, NtHDZI24 and 25 might function as a LATE MERISTEM IDENTITY1 (LMI1) homeodomain protein to regulate stipular ratio similar to the role of *ATHB51* [51], whereas NtHDZI1, 2, 4, 6, 13, 16, 17, and 21 might be involved in regulating the response to water-deficit in an ABA-dependent manner similar to the function of *ATHB6* [45].

3.6. Expression Profiles of Tobacco HD-Zip I Genes in Various Tissues

To gain insight into potential functions, quantitative RT-PCR was employed to determine the expression patterns of *NtHD-Zip* I genes in five tissues: roots, stems, leaves, flowers, and seeds (Figure 5). All 39 *NtHD-Zip* I genes were expressed in one or more tested tissues with different expression patterns.

NtHDZI1, 2, 14–16, 26, and 33–35 had the highest expression in flowers compared with other tissues. Fifteen NtHD-Zip I genes: *NtHDZI6*, 10, 11, 20, 21, 23, 27–31, and 36–39 showed higher expression levels in seeds and some were even expressed in a seed-specific manner. Interestingly, *NtHDZI24* and *NtHDZI25* were exclusively expressed in leaves, and *NtHDZI22* was predominantly expressed in roots. *NtHDZI7* and *NtHDZI12* had higher expression in roots, stems, and flowers, and *NtHDZI19* and *NtHDZI32* were mainly expressed in leaves, flowers, and seeds. In addition, *NtHDZI15* had higher expression levels in roots, stems, and seeds, and *NtHDZI8* was preferentially expressed in roots, leaves, and flowers. *NtHDZI9* and *NtHDZI17* showed higher expression in roots and seeds, and *NtHDZI13* and *NtHDZI18* were highly expressed in roots and leaves. Compared to other genes, *NtHDZI4* showed relatively wider expression patterns in roots and flowers. Furthermore, *NtHDZI3* showed a relatively stable expression pattern among all tissues tested. Interestingly, members within each clade did not always share similar expression patterns.

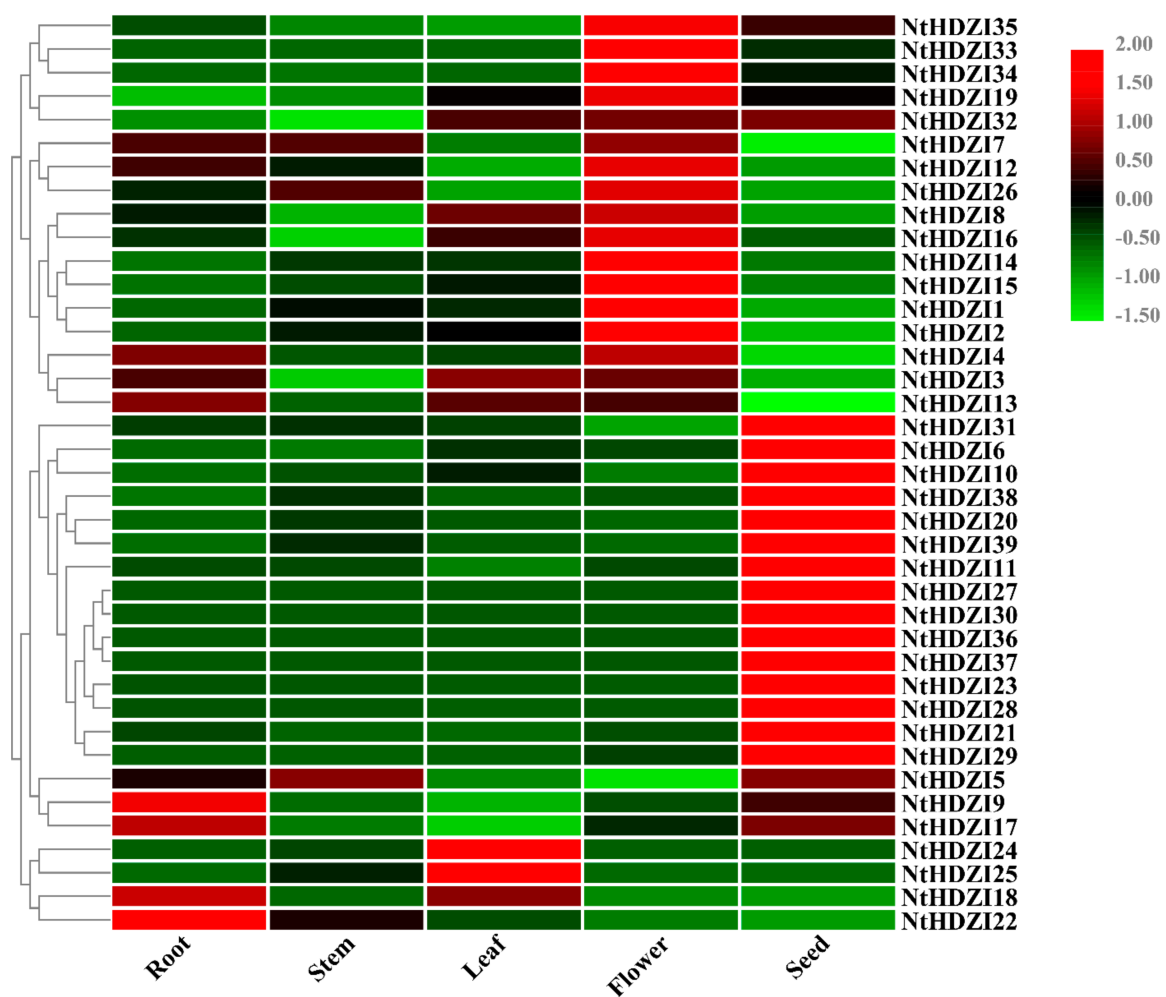


Figure 5. Relative expression levels of *NtHDZI* genes in various tissues. The roots, stems, leaves, flowers, and seeds were collected from the *N. tabacum* plants. All of the expression levels of the *NtHDZI* genes were normalized by the expression levels of *NtEF1 α* . The $2^{-\Delta\Delta CT}$ method was used to be an evaluation of the relative expression. The heat map was drawn in row scale-transformed expression values. Red or green colors represent the difference in expression levels in each sample, respectively.

3.7. Expression Profiling of Tobacco HD-Zip I Genes under Low Temperature and ABA Treatments

The HD-Zip I genes have been reported to play important roles in response to abiotic stress. Transcript abundances of *NtHD-Zip* I genes were investigated in leaves under low temperature (Figure 6) and ABA treatments (Figure 7). The results suggested that most *NtHD-Zip* I genes were responsive to low-temperature stress and ABA treatment, with the exception of *NtHDZI21* and *NtHDZI33*. Expression of *NtHDZI22*, 24, and 25 exhibited rapid down-regulation at 1 h after exposure to low temperature, but *NtHDZI2–7*, 9, 13, 15, 17, 20, 31, 38, and 39 were significantly down-regulated at 6 h after exposure to low temperature (Figure 6). Expression of some genes, for example, *NtHDZI27*, 29, 36, and 37, declined at a late stage after plant exposure to low temperature. With low-temperature treatment, *NtHDZI1*, 8, 10–12, 14, 16, 19, 23, 28, 30, and 34 were up-regulated at 1 h, whereas expression of *NtHDZI18*, 26, 32, and 35 increased greatly and significantly at 12 or 24 h. Interestingly, *NtHDZI21* and *NtHDZI33* expression did not significantly change at all time points tested after low-temperature treatment.

The expression patterns of *NtHD-Zip* I genes were roughly divided into three types under ABA treatment (Figure 7). The first type was characterized by an expression pattern with significant up-regulation at 1 h; these genes were *NtHDZI1–4*, 7–9, 11–13, 17, 21, 29, and 31–37. The second type of expression pattern was continuous down-regulation under ABA treatment, including *NtHDZI5*,

6, 14, 15, 18, 19, 22, 24, and 25. The third type of expression patterns was a transient decline at 1 h, recovered or even enhanced at 6 or 12 h, and then declined again at 24 h. Genes with the third type of expression pattern were *NtHDZ110*, 16, 20, 23, 26–28, 30, 38, and 39.

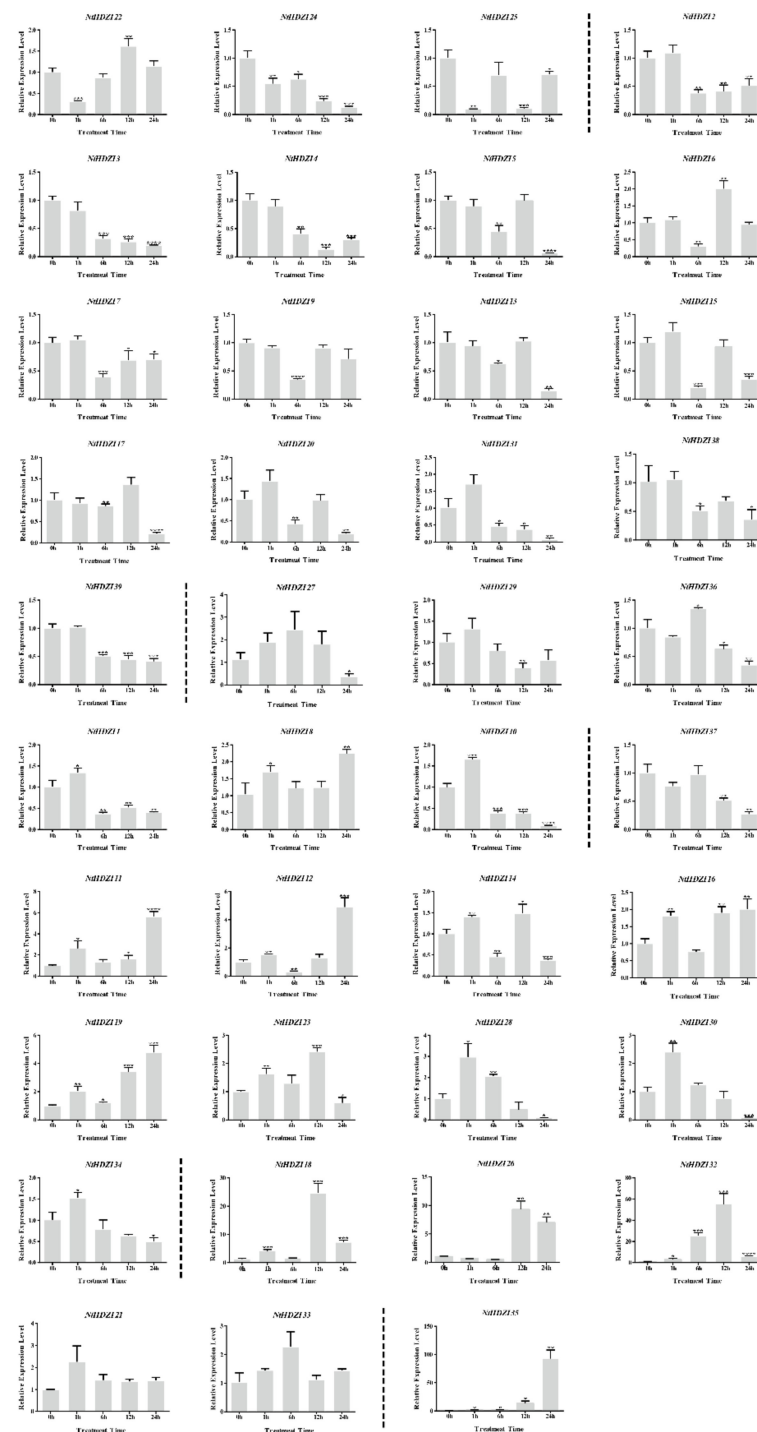


Figure 6. The expression levels of 39 *NtHD-Zip I* genes in *N. tabacum* under low-temperature (4 °C) treatment. Bars represent the mean values of three replicates \pm standard deviation (SD). All of the expression levels of the *NtHDZ1* genes were normalized by the expression levels of *NtEF1a*. Untreated leaves (0 h) were normalized as “1” at each graph. ****, ***, **, and * indicate significant difference compared to the control (0 h) at $p < 0.0001$, $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

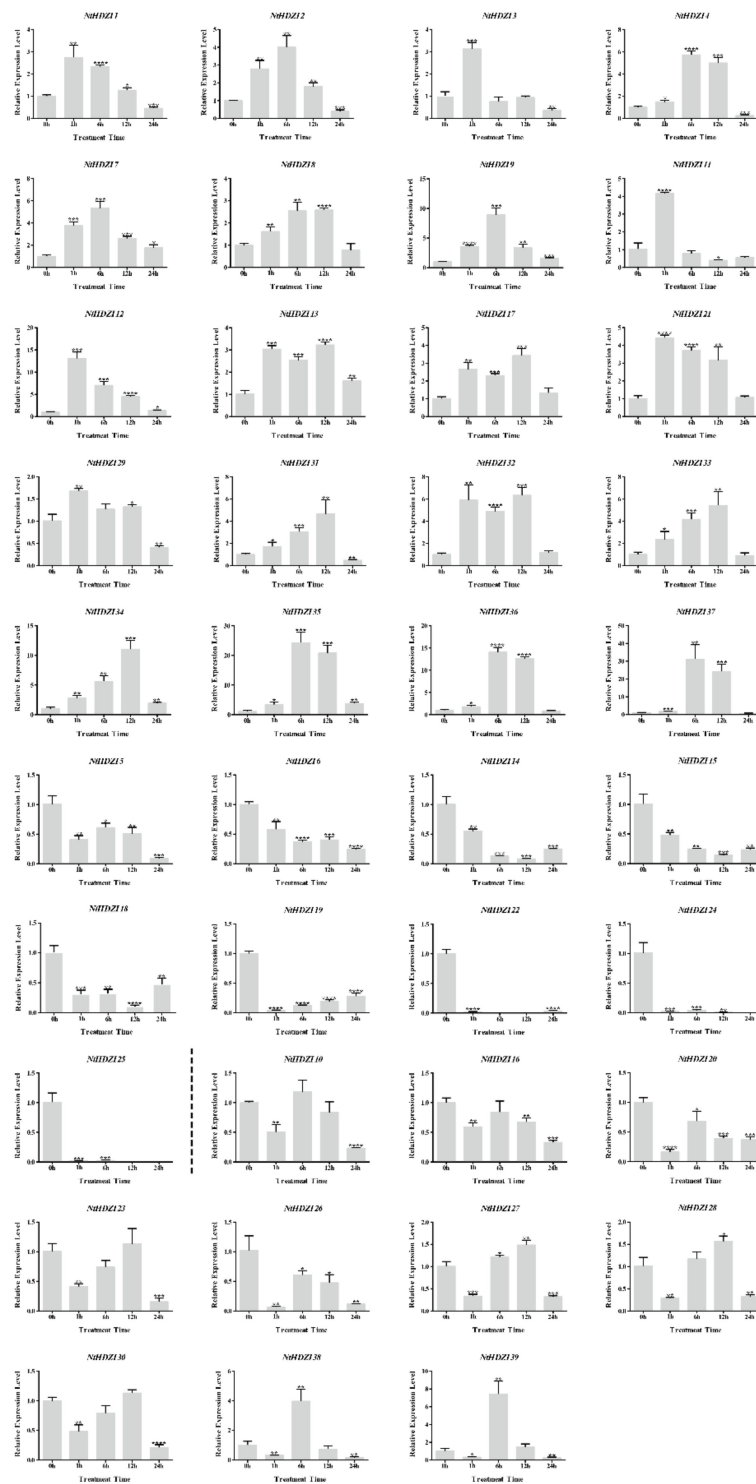


Figure 7. The expression levels of 39 *NtHD-Zip I* genes in *N. tabacum* under ABA (200 μ M) treatment. Bars represent the mean values of three replicates \pm standard deviation (SD). All of the expression levels of the *NtHDZi* genes were normalized by the expression levels of *NtEF1 α* . Untreated leaves (0 h) were normalized as “1” at each graph. ****, ***, **, and * indicate significant difference compared to the control (0 h) at $p < 0.0001$, $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

4. Discussion

Plants are confronted with major challenges involving biotic and abiotic stresses. The TFs, as regulatory proteins regulating gene expression by binding to cis-acting elements on the promoter [2],

are responsive to various stresses. The HD-Zip proteins are a type of TF found only in plants, and play important roles in plant growth and stress response [3,4]. The HD-Zip family consists of four subfamilies based on the reported identification in other species: HD-Zip I–IV [11]. The HD-Zip I subfamily is mainly involved in stress response, such as to low temperature, drought, salt, and ABA [5,6,19]. To date, although HD-Zip I members have been extensively studied in other species [6,20,52–54], no detailed identification has been performed in the model plant tobacco. In this study, 39 *NtHD-Zip* I genes were identified in *N. tabacum* by comprehensive analysis. Tobacco has more HD-Zip I genes (39) compared with other plants—17, 22, 20, 14, and 16 in *Arabidopsis*, tomato, tea plant, rice, and maize, respectively [3,6,8,9,55].

As an allotetraploid plant, cultivated tobacco was evolved through interspecific hybridization of *N. sylvestris* and *N. tomentosiformis*, and the origin of these HD-Zip I genes were surveyed for chromosome localization [37,38]. The complete genomic sequences of *NtHDZI1–6*, 8, 10, 13, 15, 18–20, 24, 25, 31, and 35–39 were only found in the T-genome, while those of *NtHDZI12*, 14, 21, and 32 only appeared in the S-genome of *N. tabacum* (Table 1). The uneven distribution of these *NtHD-Zip* I genes implied that complex gene rearrangement and deletion occurred following interspecific hybridization.

Importantly, similarly to HD-Zip I in other plant species [6–8,20], all identified *NtHD-Zip* I proteins contained conserved HD and LZ domains (Figure 2 and Figure S1). There were 2–3 exons for most *NtHD-Zip* I genes, except for *NtHDZI38* and *NtHDZI39*, which had just one exon. Phylogenetic analysis showed that *NtHD-Zip* I genes were distributed in 8 out of 11 plant HD-Zip I subclasses, indicating that the basic features of the plant HD-Zip I gene family were formed early, which was also independently confirmed by the intron–exon structure and motif organization patterns (Figure 1, Figure 2, and Figure S2; Table S4).

There were 14 genes in pairs in the phylogenetic tree: *NtHDZI1* and *NtHDZI2*, *NtHDZI8* and *NtHDZI16*, *NtHDZI13* and *NtHDZI17*, *NtHDZI24* and *NtHDZI25*, *NtHDZI33* and *NtHDZI34*, *NtHDZI36* and *NtHDZI37*, and *NtHDZI38* and *NtHDZI39* (Figure 2). In addition to high sequence similarities, similar genomic structure, motif composition, and three-dimensional structure (Figure 2; Figure S2; Table S4), most pairwise genes: *NtHDZI1* and *NtHDZI2*, *NtHDZI13* and *NtHDZI17*, *NtHDZI24* and *NtHDZI25*, *NtHDZI33* and *NtHDZI34*, *NtHDZI36* and *NtHDZI37*, and *NtHDZI38* and *NtHDZI39*, tended to have a similar expression pattern in response to low temperature stress or ABA treatment (Figures 6 and 7). For instance, both *NtHDZI36* and *NtHDZI37* were induced to the highest expression level at 6 h post-ABA stimuli, whereas *NtHDZI38* and *NtHDZI39* showed a dramatic reduction after 1 h for under low-temperature treatment and were significantly induced at 6 h by ABA treatment (Figure 7). In addition, *NtHDZI1* and *NtHDZI2* both showed a sharp reduction at 6 h under low-temperature stress (Figure 6). Interestingly, most of these paired HD-Zip I genes originated from *N. tomentosiformis* and derived from the same chromosome, for example, *NtHDZI1* and *NtHDZI2* in chromosome 13, *NtHDZI36* and *NtHDZI37* in chromosome 23, and *NtHDZI38* and *NtHDZI39* in chromosome 2 (Table 1). This implied that gene rearrangement or duplication occurred after interspecific hybridization. One exception was *NtHDZI13* and *NtHDZI17*, which exhibited similar expression patterns under low-temperature stress or ABA treatment. However, there were also paired genes which showed different expression patterns in response to low-temperature stress or ABA treatment, such as *NtHDZI8* and *NtHDZI16–NtHDZI8* were up-regulated, but *NtHDZI16* was down-regulated at 1 h under ABA treatment (Figure 7). Tomato *SLHZ08*, *SLHZ13*, and *Arabidopsis ATHB13* all belong to clade α , and expression of *SLHZ08*, *SLHZ13*, and *ATHB13* are up-regulated by low-temperature stress [6,56]. Expression of *SLHZ08* and *SLHZ13* can be up-regulated 3 h after the cold treatment [6]. Moreover, the expression level of *ATHB13* could also be induced, and overexpression of *ATHB13* conferred cold tolerance in *Arabidopsis* [56]. *NtHDZI11*, 12, 14, 19, 22 and 23 were *NtHD-Zip* I family members in clade α (Figure 1), and expression of these genes were all up-regulated by low-temperature stress. Furthermore, most genes in this clade had a similar motif composition (Figure 2). Thus, these *NtHD-Zip* I genes might be involved in low-temperature tolerance.

Analysis of three-dimensional structure and protein–protein interactions are helpful to gain insights into the function of NtHD-Zip I members. Results of three-dimensional modeling indicated that most genes in the same clades tended to have similar protein structures (Figure 3). For instance, NtHDZI20 and NtHDZI22 of clade α showed similar protein models; and NtHDZI28 and NtHDZI29 of clade δ also had similar protein structures. Furthermore, a predicted interaction network of HD-Zip I proteins between tobacco and *Arabidopsis* was constructed (Figure 4). Previous reports showed that AtHB7 and AtHB12 functioned as positive transcriptional regulators of PP2C genes that participate in negative regulation of ABA signals [26]. Therefore, the tobacco counterparts *NtHDZI32–37* might share similar functions to those of *AtHB7* and *AtHB12*, which was further supported by the evolutionary analysis. Phylogenetic analysis revealed that *NtHDZI32–37* belongs to the same clades γ with *AtHB7* and *AtHB12* (Figure 1). Furthermore, expression profiling indicated the significant induction of these six genes (*NtHDZI32–37*) after ABA treatment, which suggests that *NtHDZI32–37* might participate in ABA-mediated signaling pathways. In addition, *NtHDZI11, 12, 14, 15, 19, and 20* may play an active role in the sucrose-signaling pathway as *ATHB13* does [18], whereas *NtHDZI3, 8, 10, and 18* may participate in regulating the photoperiod process as does *ATHB16* does [17].

The expression profile analysis showed that *NtHD-Zip I* genes had large expression variations among different tissues, which implied that they might have diverse functions. *NtHDZI1, 2, 14–16, 26, and 33–35* had higher expression levels in flowers (Figure 5). It was previously reported that *GhHB12*, specifically expressed in axillary buds, could affect cotton morphological construction and flower development delay when overexpressed [57]. *NtHDZI6, 10, 11, 20, 21, 23, 27–31, and 36–39* showed a seed-specific expression pattern (Figure 5), implying a role in seed formation and seed dormancy similar to *ATHB20*, which was also abundantly expressed in seed germination of *Arabidopsis* [50]. The ectopic expression of *HaHB4*, an HD-Zip I TF from *Helianthus annuus*, could significantly change the morphology of the veins [58]. Given that NtHDZI24 and NtHDZI25 showed a relatively high expression in leaves, *NtHDZI24* and *NtHDZI25* might associate with tobacco leaf development.

Many studies have reported that the HD-Zip I subfamily participates in abiotic stress response [6,19,26–33,45,50,56,59–61]. Cabello et al. found that *HaHB1* or *AtHB13* can be up-regulated by drought and salt and can significantly improve salt tolerance and drought resistance by their ectopic expression to maintain cell membrane integrity [59]. Furthermore, silencing of *SIHB2* can significantly increase the water content of plants and malondialdehyde contents to enhance the tolerance of plants to salt and drought in tomato [60]. Overexpressing an HD-Zip I gene from wheat endosperm, *TaHDZip1-2*, improves the freezing tolerance of plants [61]. In this study, we investigated the potentially abiotic stress-responding elements, including ABA, low temperature, drought stress, and salinity stress-regulated elements, within 2000-bp upstream of the start codon ATG (Table 2; Figure S4). Results showed that all the *NtHD-Zip I* genes contained a number of these stress-responding elements. The ABA-responding component accounted for the greatest number in the promoter region of *NtHD-Zip I* genes, elements related to drought and salinity stresses ranked in second place, and were followed by low-temperature stress components. A quantitative RT-PCR based expression profile revealed that expression levels of all *NtHD-Zip I* genes, except for *NtHDZI21* and *NtHDZI33*, showed significant changes, either up-regulated or down-regulated under low temperature or ABA treatment (Figures 6 and 7), which is consistent with the results from the analysis of stress-regulated elements within the promoter sequence of *NtHD-Zip I* genes.

Although the potential functions were suggested based on the comparing of *NtHDZI* genes and deduced NtHDZI proteins with the members in other plant HD-Zip I subfamily (particularly with those in *Arabidopsis*) for exon–intron structure, predicted cis-acting elements associated with abiotic stress factors in the promoter regions, expression profiles, and homology, three-dimensional structures of proteins, putative protein–protein networks, respectively, the further functional characterizations of *NtHDZI* genes are required for understanding the biological significance of HD-Zip I subfamily in *N. tabacum*.

5. Conclusions

In total, 39 *NtHD-Zip* I genes were identified in this study. The *NtHD-ZIP* I gene subfamily were grouped into eight clades based on the evolutionary analysis. Gene structure, conserved motifs, and three-dimensional modeling were further analyzed. Genes grouped into the same clade tended to have similar exon-intron organization, motifs, and protein structure. Moreover, the upstream sequences of *NtHD-Zip* I genes contained abundant cis-acting elements associated with ABA response and abiotic stresses, such as low temperature, drought, or salinity response elements. The protein–protein interaction network was shown for tobacco and *Arabidopsis* HD-Zip I proteins. There were different expression patterns among different tissues, under ABA treatment, and low-temperature stress. This study will provide useful information for the functional characterizing *HD-Zip* I genes in *N. tabacum*.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/10/8/575/s1>, Figure S1: Multiple sequence alignments in NtHD-Zip I proteins, Figure S2: Sequence logos of NtHDZI motifs in *N. tabacum*. The overall height of the stack represents the level of sequence conservation. The height of the residue in the stack represents the relative frequency of each residue at that location, Figure S3: The distribution of different stress-related cis-elements predicted in the promoter region of NtHD-Zip I genes, Figure S4: Promoter Analysis of HD-ZIP I genes in tobacco. A sequence of 2000 bp upstream of the start codon ATG of HD-Zip I genes in *N. tabacum* were investigated to analyze cis-elements by using Genomatix. Putative cis-elements were represented by different colors and distributed on the black lines, Table S1: Protein sequences of HD-Zip I in *Nicotiana tabacum*, *Camellia sinensis*, *Solanum lycopersicum*, *Oryza sativa*, *Manihot esculenta*, *Zea mays*, and *Arabidopsis thaliana*, Table S2: Primer sequences used for the qPCR analysis of HD-Zip I gene expression in tobacco, Table S3: Distribution of NtHD-ZIP I gene in ancestral species *N. sylvestris* and *N. tomentosiformis*, Table S4: Predicted motif sequence in NtHD-Zip I proteins, Table S5: Classification and statistics of different kinds of cis-acting elements in the promoter regions of NtHDZI genes.

Author Contributions: Conceptualization, Q.X., D.Y., and G.W.; methodology, Q.X., D.Y., and G.W.; formal analysis, D.Y., G.W., and Y.L.; investigation, Y.L., B.B., F.W., M.Z., and Q.X.; supervision, D.Y. and G.W.; writing—original draft preparation, Y.L. and G.W.; writing—review and editing, G.W., Q.X., D.Y., Y.L., B.B., F.W., and M.Z.

Funding: This work was supported by grants from Fundamental Research Funds for the Central Universities (No. XDJK2018C007), 110201801025(JY-02), NY20180401070006; the National Natural Science Foundation of China (No. 31760072, 31860413); and the Yunnan Academy of Tobacco Agricultural Sciences (No. 2016YN22, 2018530000241002, 2019530000241003).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Franco-Zorrilla, J.M.; Solano, R. Identification of plant transcription factor target sequences. *Biochim. Biophys. Acta* **2017**, *1860*, 21–30. [[CrossRef](#)] [[PubMed](#)]
2. Todeschini, A.L.; Georges, A.; Veitia, R.A. Transcription factors: Specific DNA binding and specific gene regulation. *Trends Genet.* **2014**, *30*, 211–219. [[CrossRef](#)] [[PubMed](#)]
3. Henriksson, E.; Olsson, A.S.; Johannesson, H.; Johansson, H.; Hanson, J.; Engström, P.; Söderman, E. Homeodomain Leucine Zipper Class I Genes in Arabidopsis. Expression Patterns and Phylogenetic Relationships. *Plant Physiol.* **2005**, *139*, 509–518. [[CrossRef](#)] [[PubMed](#)]
4. Ariel, F.D.; Manavella, P.A.; Dezar, C.A.; Chan, R.L. The true story of the HD-Zip family. *Trends Plant Sci.* **2007**, *12*, 419–426. [[CrossRef](#)] [[PubMed](#)]
5. Chen, X.; Chen, Z.; Zhao, H.; Zhao, Y.; Cheng, B.; Xiang, Y. Genome-Wide Analysis of Soybean HD-Zip Gene Family and Expression Profiling under Salinity and Drought Treatments. *PLoS ONE* **2014**, *9*, e87156. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, Z.; Chen, X.; Guan, X.; Liu, Y.; Chen, H.; Wang, T.; Wang, J.L.; Wang, J. A genome-wide survey of homeodomain-leucine zipper genes and analysis of cold-responsive HD-Zip I members' expression in tomato. *Biosci. Biotechnol. Biochem.* **2014**, *78*, 1337–1349. [[CrossRef](#)]
7. Li, Z.; Zhang, C.; Guo, Y.; Niu, W.; Wang, Y.; Xu, Y. Evolution and expression analysis reveal the potential role of the HD-Zip gene family in regulation of embryo abortion in grapes (*Vitis vinifera* L.). *BMC Genom.* **2017**, *18*, 744. [[CrossRef](#)] [[PubMed](#)]

8. Agalou, A.; Purwantomo, S.; Övernäs, E.; Johannesson, H.; Zhu, X.; Estiati, A.; de Kam, R.J.; Engström, P.; Slamet-Loedin, I.H.; Zhu, Z.; et al. A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. *Plant Mol. Biol.* **2008**, *66*, 87–103. [[CrossRef](#)]
9. Mao, H.; Yu, L.; Li, Z.; Liu, H.; Han, R. Molecular evolution and gene expression differences within the HD-Zip transcription factor family of *Zea mays* L. *Genetica* **2016**, *144*, 243. [[CrossRef](#)]
10. Yue, H.; Shu, D.; Wang, M.; Xing, G.; Zhan, H.; Du, X.; Song, W.; Nie, X. Genome-Wide Identification and Expression Analysis of the HD-Zip Gene Family in Wheat (*Triticum aestivum* L.). *Genes* **2018**, *9*, 70. [[CrossRef](#)]
11. Elhiti, M.; Stasolla, C. Structure and function of homodomain-leucine zipper (HD-Zip) proteins. *Plant Signal. Behav.* **2009**, *4*, 86–88. [[CrossRef](#)] [[PubMed](#)]
12. Harris, J.C.; Hrmova, M.; Lopato, S.; Langridge, P. Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytol.* **2011**, *190*, 823–837. [[CrossRef](#)] [[PubMed](#)]
13. Meijer, A.H.; de Kam, R.J.; D’Erfurth, I.; Shen, W.; Hoge, J.H. HD-Zip proteins of families I and II from rice: Interactions and functional properties. *Mol. Genet. Evol.* **2000**, *263*, 12–21. [[CrossRef](#)]
14. Côté, C.L.; Boileau, F.; Roy, V.; Ouellet, M.; Lévasseur, C.; Morency, M.J.; Cooke, J.E.; Séguin, A.; Mackay, J.J. Gene family structure, expression and functional analysis of HD-Zip III genes in angiosperm and gymnosperm forest trees. *BMC Plant Biol.* **2010**, *10*, 273. [[CrossRef](#)]
15. Pandey, A.; Misra, P.; Alok, A.; Kaur, N.; Sharma, S.; Lakhwani, D.; Asif, M.H.; Tiwari, S.; Trivedi, P.K. Genome-Wide Identification and Expression Analysis of Homeodomain Leucine Zipper Subfamily IV (HDZ IV) Gene Family from *Musa accuminata*. *Front. Plant Sci.* **2016**, *7*, 20. [[CrossRef](#)] [[PubMed](#)]
16. Aoyama, T.; Dong, C.H.; Wu, Y.; Carabelli, M.; Sessa, G.; Ruberti, I.; Morelli, G.; Chua, N.H. Ectopic Expression of the Arabidopsis Transcriptional Activator *AtHB-1* Alters Leaf Cell Fate in tobacco. *Plant Cell* **1995**, *7*, 1773. [[PubMed](#)]
17. Yan, W.; Eva, H.; Eva, S.D.; Kerstin Nordin, H.; Eva, S.; Peter, E.M. The Arabidopsis homeobox gene, *ATHB16*, regulates leaf development and the sensitivity to photoperiod in Arabidopsis. *Dev. Biol.* **2004**, *264*, 228–239.
18. Hanson, J.; Johannesson, H.; Engström, P. Sugar-dependent alterations in cotyledon and leaf development in transgenic plants expressing the HDZhdip gene *ATHB13*. *Plant Mol. Biol.* **2001**, *45*, 247–262. [[CrossRef](#)]
19. Johannesson, H.; Wang, Y.; Hanson, J.; Engström, P. The Arabidopsis thaliana homeobox gene *ATHB5* is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol. Biol.* **2003**, *51*, 719–729. [[CrossRef](#)]
20. Liu, W.; Fu, R.; Li, Q.; Li, J.; Wang, L.; Ren, Z. Genome-wide identification and expression profile of homeodomain-leucine zipper Class I gene family in *Cucumis sativus*. *Gene* **2013**, *531*, 279–287. [[CrossRef](#)]
21. Hur, Y.S.; Um, J.H.; Kim, S.; Kim, K.; Park, H.J.; Lim, J.S.; Kim, W.Y.; Jun, S.E.; Yoon, E.K.; Lim, J. Arabidopsis thaliana homeobox 12 (*ATHB12*), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytol.* **2015**, *205*, 316–328. [[CrossRef](#)] [[PubMed](#)]
22. Capella, M.; Ribone, P.A.; Arce, A.L.; Chan, R.L. Arabidopsis thaliana HomeoBox 1 (*AtHB1*), a Homeodomain-Leucine Zipper I (HD-Zip I) transcription factor, is regulated by PHYTOCHROME-INTERACTING FACTOR 1 to promote hypocotyl elongation. *New Phytol.* **2015**, *207*, 669–682. [[CrossRef](#)] [[PubMed](#)]
23. Li, Q.; Cao, C.; Zhang, C.; Zheng, S.; Wang, Z.; Wang, L.; Ren, Z. The identification of *Cucumis sativus* *Glabrous 1* (*CsGL1*) required for the formation of trichomes uncovers a novel function for the homeodomain-leucine zipper I gene. *J. Exp. Bot.* **2015**, *66*, 2515. [[CrossRef](#)] [[PubMed](#)]
24. Ariel, F.; Diet, A.; Verdenaud, M.; Gruber, V.; Frugier, F.; Chan, R.; Crespi, M. Environmental Regulation of Lateral Root Emergence in *Medicago truncatula* Requires the HD-Zip I Transcription Factor *HB1*. *Plant Cell* **2010**, *22*, 2171–2183. [[CrossRef](#)]
25. Shao, J.; Haider, I.; Xiong, L.; Zhu, X.; Rmf, H.; E, Ö.; Meijer, A.H.; Zhang, G.; Wang, M.; Bouwmeester, H.J. Functional analysis of the HD-Zip transcription factor genes *Oshox12* and *Oshox14* in rice. *PLoS ONE* **2018**, *13*, e0199248. [[CrossRef](#)] [[PubMed](#)]
26. Valdés, A.E.; Övernäs, E.; Johansson, H.; Radaiglesias, A.; Engström, P. 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.* **2012**, *80*, 405–418. [[CrossRef](#)]

27. Manavella, P.A.; Arce, A.L.; Dezar, C.A.; Bitton, F.; Renou, J.P.; Crespi, M.; Chan, R.L. Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower *Hahb-4* transcription factor. *Plant J.* **2010**, *48*, 125–137. [[CrossRef](#)]
28. Zhang, S.; Haider, I.; Kohlen, W.; Jiang, L.; Bouwmeester, H.; Meijer, A.H.; Schluempmann, H.; Liu, C.M.; Ouwerkerk, P.B.F. Function of the HD-Zip I gene *Oshox22* in ABA-mediated drought and salt tolerances in rice. *Plant Mol. Biol.* **2012**, *80*, 571–585. [[CrossRef](#)]
29. Xi, H.; Min, D.; Liao, J.; Xi, Y.; Hui, C.; Feng, J.; Ji, H.; Zhang, H.S. OsSLI1, a Homeodomain Containing Transcription Activator, Involves Abscisic Acid Related Stress Response in Rice (*Oryza sativa* L.). *Sci. World J.* **2014**, *2014*, 809353.
30. Zhao, Y.; Ma, Q.; Jin, X.; Peng, X.; Liu, J.; Deng, L.; Yan, H.; Sheng, L.; Jiang, H.; Cheng, B. 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.* **2014**, *55*, 1142–1156. [[CrossRef](#)]
31. Yang, Y.F.; Luang, S.; Harris, J.; Riboni, M.; Li, Y.; Bazanova, N.; Hrmova, M.; Haefele, S.; Kovalchuk, N.; Lopato, S. Overexpression of the class I homeodomain transcription factor *TaHDZip1-5* increases drought and frost tolerance in transgenic wheat. *Plant Biotechnol. J.* **2017**, *16*, 1227–1240. [[CrossRef](#)] [[PubMed](#)]
32. Cao, L.; Yu, Y.; Duanmu, H.; Chen, C.; Duan, X.; Zhu, P.; Chen, R.; Li, Q.; Zhu, Y.; Ding, X. A novel *Glycine soja* homeodomain-leucine zipper (HD-Zip) I gene, *Gshdz4*, positively regulates bicarbonate tolerance and responds to osmotic stress in *Arabidopsis*. *BMC Plant Biol.* **2016**, *16*, 184. [[CrossRef](#)] [[PubMed](#)]
33. Sen, S.; Chakraborty, J.; Ghosh, P.; Basu, D.; Das, S. Chickpea *WRKY70* Regulates the Expression of a Homeodomain-Leucine Zipper (HD-Zip) I Transcription Factor *CaHDZ12*, which Confers Abiotic Stress Tolerance in Transgenic Tobacco and Chickpea. *Plant Cell Physiol.* **2017**, *58*, 1934–1952. [[CrossRef](#)] [[PubMed](#)]
34. Ré, D.A.; Dezar, C.A.; Chan, R.L.; Baldwin, I.T.; Bonaventure, G. *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.* **2011**, *62*, 155–166. [[CrossRef](#)] [[PubMed](#)]
35. Gao, D.; Appiano, M.; Huibers, R.P.; Chen, X.; Loonen, A.E.; Visser, R.G.; Wolters, A.M.; Bai, Y. Activation tagging of *ATHB13* in *Arabidopsis thaliana* confers broad-spectrum disease resistance. *Plant Mol. Biol.* **2014**, *86*, 641–653. [[CrossRef](#)] [[PubMed](#)]
36. Mou, S.; Liu, Z.; Gao, F.; Yang, S.; Su, M.; Shen, L.; Wu, Y.; He, S. *CaHDZ27*, a Homeodomain-Leucine Zipper I (HD-Zip I) Protein, Positively Regulates the Resistance to *Ralstonia solanacearum* Infection in Pepper. *Mol. Plant-Microbe Interact.* **2017**, *30*, 960. [[CrossRef](#)] [[PubMed](#)]
37. Edwards, K.D.; Fernandez-Pozo, N.; Drake-Stowe, K.; Humphry, M.; Evans, A.D.; Bombarely, A.; Allen, F.; Hurst, R.; White, B.; Kernodle, S.P.; et al. A reference genome for *Nicotiana tabacum* enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genom.* **2017**, *18*, 448. [[CrossRef](#)] [[PubMed](#)]
38. Sierro, N.; Battey, J.N.D.; Ouadi, S.; Bakaher, N.; Bovet, L.; Willig, A.; Goepfert, S.; Peitsch, M.C.; Ivanov, N.V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* **2014**, *5*, 3833. [[CrossRef](#)]
39. Sierro, N.; Battey, J.N.; Ouadi, S.; Bovet, L.; Goepfert, S.; Bakaher, N.; Peitsch, M.C.; Ivanov, N.V. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* **2013**, *14*, R60. [[CrossRef](#)]
40. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *25*, 4876–4882. [[CrossRef](#)]
41. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845. [[CrossRef](#)] [[PubMed](#)]
42. Zhang, X.; Cheng, T.; Wang, G.; Yan, Y. Cloning and evolutionary analysis of tobacco *MAPK* gene family. *Mol. Biol. Rep.* **2013**, *40*, 1407–1415. [[CrossRef](#)] [[PubMed](#)]
43. Livak, K.J.; Schmittgen, T.D. Analysis of Relative Gene Expression Data using Real-Time Quantitative PCR. *Method* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
44. Ré, D.A.; Capella, M.; Bonaventure, G.; Chan, R.L. Arabidopsis *AtHB7* and *AtHB12* evolved divergently to fine tune processes associated with growth and responses to water stress. *BMC Plant Biol.* **2014**, *14*, 1–14. [[CrossRef](#)] [[PubMed](#)]

45. Söderman, E.; Hjellström, M.; Fahleson, J.; Engström, P. The HD-Zip gene *ATHB6* in Arabidopsis is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions. *Plant Mol. Biol.* **1999**, *40*, 1073–1083. [[CrossRef](#)]
46. Gonzálezgrandío, E. Abscisic acid signaling is controlled by a *BRANCHED1/HD-ZIP I* cascade in Arabidopsis axillary buds. *Proc. Natl. Acad. Sci. USA* **2016**, *114*, E245. [[CrossRef](#)]
47. Fornara, F.; de Montaigu, A.; Sanchez-Villarreal, A.; Takahashi, Y.; Ver Loren van Themaat, E.; Huettel, B.; Davis, S.J.; Coupland, G. The GI-CDF module of Arabidopsis affects freezing tolerance and growth as well as flowering. *Plant J.* **2015**, *81*, 695–706. [[CrossRef](#)]
48. Shen, Q.; Ho, T.H. Functional Dissection of an Abscisic Acid (ABA)-Inducible Gene Reveals Two Independent ABA-Responsive Complexes Each Containing a G-Box and a Novel *cis*-Acting Element. *Plant Cell* **1995**, *7*, 295–307.
49. Chai, W.; Si, W.; Ji, W.; Qin, Q.; Zhao, M.; Jiang, H. Genome-Wide Investigation and Expression Profiling of HD-Zip Transcription Factors in Foxtail Millet (*Setaria italica* L.). *Biomed. Res. Int.* **2018**, 8457614. [[CrossRef](#)]
50. Barrero, J.M.; Millar, A.A.; Griffiths, J.; Czechowski, T.; Scheible, W.R.; Udvardi, M.; Reid, J.B.; Ross, J.J.; Jacobsen, J.V.; Gubler, F. Gene expression profiling identifies two regulatory genes controlling dormancy and ABA sensitivity in Arabidopsis seeds. *Plant J.* **2010**, *61*, 611–622. [[CrossRef](#)]
51. Vuolo, F.; Kierzkowski, D.; Runions, A.; Hajheidari, M.; Mentink, R.A.; Gupta, M.D.; Zhang, Z.; Vlad, D.; Wang, Y.; Pecinka, A.; et al. LMI1 homeodomain protein regulates organ proportions by spatial modulation of endoreduplication. *Genes Dev.* **2018**, *32*, 1361–1366. [[CrossRef](#)] [[PubMed](#)]
52. Li, Z.; Jiang, H.; Zhou, L.; Lin, D.; Lin, Y.; Peng, X.; Yan, H.; Cheng, B. Molecular evolution of the HD-ZIP I gene family in legume genomes. *Gene* **2014**, *533*, 218–228. [[CrossRef](#)] [[PubMed](#)]
53. Arce, A.L.; Raineri, J.; Capella, M.; Cabello, J.V.; Chan, R.L. Uncharacterized conserved motifs outside the HD-Zip domain in HD-Zip subfamily I transcription factors; a potential source of functional diversity. *BMC Plant Biol.* **2011**, *11*, 42. [[CrossRef](#)] [[PubMed](#)]
54. Song, A.; Li, P.; Xin, J.; Chen, S.; Zhao, K.; Wu, D.; Fan, Q.; Gao, T.; Chen, F.; Guan, Z. Transcriptome-Wide Survey and Expression Profile Analysis of Putative Chrysanthemum HD-Zip I and II Genes. *Genes* **2016**, *7*, 19. [[CrossRef](#)] [[PubMed](#)]
55. Shen, W.; Li, H.; Teng, R.; Wang, Y.; Wang, W.; Zhuang, J. Genomic and transcriptomic analyses of HD-Zip family transcription factors and their responses to abiotic stress in tea plant (*Camellia sinensis*). *Genomics* **2018**. [[CrossRef](#)]
56. Cabello, J.V.; Arce, A.L.; Chan, R.L. The homologous HD-Zip I transcription factors *HaHB1* and *AtHB13* confer cold tolerance via the induction of pathogenesis-related and glucanase proteins. *Plant J.* **2012**, *69*, 141–153. [[CrossRef](#)]
57. He, X.; Wang, T.; Xu, Z.; Liu, N.; Wang, L.; Hu, Q.; Luo, X.; Zhang, X.; Zhu, L. The cotton HD-Zip transcription factor *GhHB12* regulates flowering time and plant architecture via the GhmiR157-GhSPL pathway. *Commun. Biol.* **2018**, *1*, 229. [[CrossRef](#)]
58. Moreno-Piovan, G.S.; Moreno, J.E.; Cabello, J.V.; Arce, A.L.; Otegui, M.E.; Chan, R.L. A role for LAX2 in regulating xylem development and lateral-vein symmetry in the leaf. *Ann. Bot.* **2017**, *120*, 577–590. [[CrossRef](#)]
59. Cabello, J.V.; Chan, R.L. The homologous homeodomain-leucine zipper transcription factors *HaHB1* and *AtHB13* confer tolerance to drought and salinity stresses via the induction of proteins that stabilize membranes. *Plant Biotechnol. J.* **2012**, *10*, 815–825. [[CrossRef](#)]
60. Hu, J.; Chen, G.; Yin, W.; Cui, B.; Yu, X.; Lu, Y.; Hu, Z. Silencing of *SlHB2* Improves Drought, Salt Stress Tolerance, and Induces Stress-Related Gene Expression in Tomato. *J. Plant Growth Regul.* **2017**, *36*, 578–589. [[CrossRef](#)]
61. Kovalchuk, N.; Chew, W.; Sornaraj, P.; Borisjuk, N.; Yang, N.; Singh, R.; Bazanova, N.; Shavruk, Y.; Guendel, A.; Munz, E.; et al. The homeodomain transcription factor *TaHDZip1-2* from wheat regulates frost tolerance, flowering time and spike development in transgenic barley. *New Phytol.* **2016**, *211*, 671–687. [[CrossRef](#)] [[PubMed](#)]

