Genome-wide identification and expression profiling of the *bZIP* gene family in *Chrysanthemum indicum* and the functional analysis of *CibZIP29* under cadmium stress

Shengyan Chen[1#] 🆔, Yin Zhang[1#], Kaiyuan Zhang[1], Shuguang Liu[1], Yujia Yang[1] and Xingnong Xu[2*]

[1] *College of Landscape Architecture, Northeast Forestry University, Haerbin 150040, Heilongjiang, China*
[2] *Yancheng Third People's Hospital, Yancheng 224008, Jiangsu, China*
[#] Authors contributed equally: Shengyan Chen, Yin Zhang
[*] Corresponding author, E-mail: xingnongx@ntu.edu.cn

Abstract

Recent industrial activities have increased cadmium (Cd) contamination in soils, negatively affecting crop growth and agricultural development. *Chrysanthemum indicum* represents a significant wild species within the *Chrysanthemum* genus. Its distinctive resistance and adaptability serve as valuable references for the enhancement of ornamental chrysanthemum breeding programs. However, the role of *bZIP* family of Cd stress in *C. indicum* remains unclear. This study conducted a genome-wide identification and expression analysis of *bZIP* genes in *C. indicum*, identifying 65 *CibZIP* genes distributed across nine chromosomes. Phylogenetic analysis grouped them into 12 subgroups, supported by conserved motifs, gene, and protein structure analyses. Tandem and segmental duplications were found to be key drivers of *CibZIP* gene family expansion, with Ka/Ks analysis indicating purifying selection of duplicated genes. Promoter analysis revealed *cis*-acting elements involved in plant development and stress response. Transcriptomic data showed that *CibZIP* genes are highly expressed in different tissues and play significant roles in Cd stress tolerance. RT-qPCR analysis further confirmed that specific *CibZIP* genes, especially CibZIP29, are upregulated in response to Cd stress. Overexpression of *CibZIP29* in transgenic *Arabidopsis* enhanced Cd tolerance, providing new insights into the molecular mechanisms underlying Cd stress response in *C. indicum*. These findings contribute to understanding the role of *CibZIP* genes in Cd tolerance and will support future research on Cd stress mechanisms in *Chrysanthemum*.

Introduction

Cadmium (Cd) pollution represents a significant adverse effect of industrialization, primarily originating from activities such as mining, smelting, sewage irrigation, and the application of fertilizers[1]. Cd is known for its high toxicity to plants, as it is readily absorbed by roots and accumulates within plant tissues. This accumulation disrupts several physiological processes, including water and mineral uptake, respiration, and photosynthesis, ultimately resulting in inhibited growth and, in severe cases, plant mortality[2]. In reaction to Cd toxicity, plants have developed a range of defensive strategies, including the extrusion of Cd across the plasma membrane, chelation within the cytosol, and sequestration within vacuoles[3]. To date, the identified Cd transporter proteins primarily encompass the natural resistance-associated macrophage protein (*NRAMP*) family, the heavy metal transporter ATPase (*HMA*) family, members of the zinc- and iron-regulated transport proteins (*ZIP*) family, the ATP-binding cassette (*ABC*) family, and the yellow stripe-like (*YSL*) family, among others.

Transcription factors (TFs) are proteins that interact with specific short sequences of DNA to modulate gene expression, which can interact with the promoter regions of specific target genes, thereby modulating their transcriptional activity either by activation or repression[4,5]. Among these, the basic leucine zipper (*bZIP*) constitutes a superfamily of transcription factors (TF) encoding genes that have garnered significant interest within the eukaryotic domain. The bZIP protein sequence features two conserved bZIP domains, which together measure approximately 60 to 80 amino acids in length. The

basic region consists of approximately 16 amino acids, responsible for recognizing and binding to specific DNA sequences, typically the ACGT motif[6,7]. The leucine zipper, located adjacent to the basic region, is composed of leucine residues at every seventh position, forming an amphipathic alpha-helix that promotes homo- or hetero-dimerization, essential for DNA binding and transcriptional regulation[8,9]. The structural characteristics of plant bZIP transcription factors significantly influence their functional roles. bZIP proteins exhibit a preferential binding affinity for DNA sequences that encompass the ACGT core, including motifs such as the A-box (TACGTA), C-box (GACGTC), and G-box (CACGTG). This binding capability allows bZIP proteins to either enhance or inhibit the expression of downstream target genes, thereby facilitating the ability of plants to adapt to intricate environmental conditions[10].

In recent years, considerable progress has been made in understanding the diverse roles of *bZIP* transcription factors in plants. These factors have been implicated in various biological processes, including stress responses, seed maturation, flower development, and hormone signaling. For example, overexpression of *ZmbZIP4* has been shown to lead to an augmentation in the number of lateral roots, an elongation of primary roots, and an enhancement of the overall root system architecture. Furthermore, the overexpression of *ZmbZIP4* facilitates an increase in the synthesis of abscisic acid, which subsequently enhances the plant's capacity to withstand abiotic stressors[11]. In Rice, *OsbZIP23*, *OsbZIP66*, and *OsbZIP72* can interact with *OsMFT2* to positively regulate ABA-responsive gene expression, thereby mediating rice seed germination[12]. *TaABI5* exhibited high expression levels in wheat seeds during both

maturity and late maturity stages, however, following germination, there was a notable decrease in its expression levels[13]. The *bZIP* transcription factors are not only involved in the regulation of plant growth and development, but they also play significant roles in the plant's response to various adverse stress conditions, encompassing both biotic and abiotic stresses. *OsbZIP52* functions as a negative regulator of the drought response in rice. Overexpression of *OsbZIP52* in transgenic lines resulted in increased sensitivity to drought stress, accompanied by a down-regulation of genes associated with stress response[14]. Transgenic plants that overexpress the wheat gene *TabZIP15* demonstrated markedly improved growth under high salinity conditions compared to wild-type plants[15]. *ZmbZIP68* has been identified as a negative regulator of cold tolerance in maize. It interacts with another negative regulator of cold stress, *ZmMPK8*, which facilitates the phosphorylation of *bZIP68*. This phosphorylation event subsequently enhances the stability and transcriptional activity of *ZmbZIP68*[16]. Plants experience a variety of stressors throughout their growth, including drought, salinity, extreme temperatures, as well as exposure to both light and heavy metal stressors. In *Arabidopsis thaliana*, *bZIP16*, *bZIP68*, and *GBF1* regulate the expression of photosynthesis-related genes by responding to blue light, thereby modulating the plant light response process[17]. The overexpression of *SpbZIP60* in transgenic Arabidopsis plants resulted in increased tolerance to Cd by mitigating the accumulation of reactive oxygen species (ROS), safeguarding the photosynthetic machinery, and reducing the overall Cd concentration within the plants[18]. The overexpression of *GmbZIP152* in Arabidopsis significantly enhances tolerance to heavy metal stress when compared to wild-type (WT) plants[19]. The heterologous expression of *ZmbZIP107* has been shown to improve rice tolerance to lead (Pb) stress while simultaneously reducing Pb absorption in the root system[20]. This indicates that *bZIP* possesses the capability to mitigate the accumulation of reactive oxygen species (ROS), decrease intracellular heavy metal concentrations, and alleviate the detrimental effects induced by heavy metals in plants subjected to heavy metal stress.

Chrysanthemum, notable for its vibrant floral colors and varied flower forms, is widely employed in landscaping and holds considerable economic importance. One of its progenitors, the diploid *C. indicum*, is recognized for its robust stress resistance and is an essential resource for the genetic enhancement of ornamental chrysanthemums[21]. Studies have indicated that *C. indicum* is a predominant species in areas affected by tailings pollution[22] positioning it as a potential candidate for the restoration of environments compromised by heavy metal contamination. Identifying the resistance genes in *C. indicum* are crucial for improving the stress tolerance of ornamental chrysanthemums and for their application in the environmental remediation of urban areas contaminated with Cd. Recent studies on *C. indicum* have predominantly concentrated on the isolation and identification of chemical constituents, the quantification of these compounds, and the investigation of their pharmacological effects. However, there is a notable lack of research addressing the relationship between chemical components and environmental factors, as well as the impact of heavy metal stress on the physiology and quality of *C. indicum*. To date, the *bZIP* gene family has been recognized in a variety of plant species, with 78 *bZIPs* identified in *A. thaliana*, 89 in rice, 160 in soybean, 55 in grape, 86 in populus, and 56 in potato, 57 in pineapple, and so on[23]. Nevertheless, there is a paucity of research concerning the *bZIP* family in *C. indicum*. Recently, the complete genome of *C. indicum* was sequenced[24], which provides a basis for the identification of *bZIP* genes within this species. In this research, the *bZIP* gene family was comprehensively identified across the *C. indicum* genome. The

study examined the phylogenetic relationships, structural features, physicochemical properties, chromosomal localization, and tissue-specific expression of the identified gene members. Additionally, the expression levels of selected members in response to Cd treatment were assessed using quantitative real-time polymerase chain reaction (qRT-PCR). Consequently, *CibZIP29* was chosen for additional investigation regarding its expression and functional characteristics. The findings of this study establish a foundational framework for the excavation and functional analysis of Cd stress-related genes in *C. indicum*. Additionally, these results pave the way for the selection and breeding of *C. indicum* varieties with enhanced tolerance to Cd. This research elucidates the molecular mechanisms underlying the increased stress tolerance in ornamental chrysanthemum varieties by examining stress tolerance-related genes in *C. indicum*. The findings may facilitate the development of transgenic breeding strategies or molecular marker-assisted breeding for chrysanthemums, ultimately contributing to the cultivation of novel ornamental chrysanthemum varieties with improved stress resilience.

Materials and methods

Identification of *bZIP* genes in *C. indicum*

The genome sequence and annotation files for *C. indicum* were obtained from the website (https://figshare.com/projects/Chrysanthemum_indicum_genome_diploid/197683). The AtbZIP protein sequences were downloaded from the PlantTFDB database (http://planttfdb.gao-lab.org/). Two methods were used to identify all the potential bZIP proteins. Firstly, the Pfam database (http://pfam.xfam.org/) was utilized to acquire the profile hidden Markov models (HMMs) of the bZIP domain (PF00170, PF07716, PF12498 and PF03131), which served as a query for identifying and searching the *C. indicum* proteome using HMMER software (version 3.4; http://hmmer.org/), applying an E-value threshold of less than 1×10^{-5}. Subsequently, all the known AtbZIP proteins were utilized as query sequences to conduct a comprehensive search of the entire genome protein sequences of *C. indicum* using the BLAST algorithm for proteins (BLASP), with a significance threshold set at an e-value of less than $1e^{-5}$. The adversarial outcomes were submitted to the SMART (http://smart.embl-heidelberg.de/), Pfam, and NCBI Conserved Domain Database (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for domain identification, with the exclusion of disqualified sequences. The identified amino acid sequence was analyzed using the ProtParam tool available on the ExPASy website (http://web.expasy.org/protparam/) to predict its physicochemical properties. Additionally, the sequence was evaluated for protein subcellular localization predictions using the WoLF PSORT II (https://www.genscript.com/tools/wolf-psort).

Chromosomal location and gene duplication analysis

We conducted a mapping of the *CibZIP* gene members to their respective chromosomes by examining the genome annotation data of *C. indicum* utilizing MapGene2Chrom online tool (http://mg2c.iask.in/mg2c_v2.1/) for analysis. We employed the local BLASTP search to investigate gene duplication events. The examination of tandem duplication events involving bZIP genes in *C. indicum* was conducted according to two specific criteria: (a) a region within 200 kb contained at least two genes; (b) shared more than 70% identity. Segmental duplication events were detected utilizing the Multiple Collinearity Scan toolkit (MCScanX) (https://github.com/wyp1125/MCScanX) with the default settings. The Circos software was employed to generate collinearity maps that illustrate the relationships between duplicated gene pairs. The homology of *bZIP* genes across *C. indicum*, *Oryza sativa*, *A. thaliana*,

Glycine max, *Populus trichocarpa*, and *Vitis vinifera* was analyzed and characterized using the software MCScan (Python version) (https://github.com/tanghaibao/jcvi/wiki/MCScan-(Python-version)). The Ka/Ks calculator version 2.0 (https://sourceforge.net/projects/kakscalculator2/) was employed to determine the ratio of the non-synonymous substitution rate (Ka) to the synonymous substitution rate (Ks) in duplicated genes.

Phylogenetic analysis and classification of *CibZIP* genes

A multiple sequence alignment of putative CibZIPs and previously reported AtbZIPs was conducted utilizing the MAFFT version 7 (https://mafft.cbrc.jp/alignment/software/)[25], employing the full-length amino acid sequences under default parameters. Subsequently, an unrooted phylogenetic tree was generated from the full-length protein sequence alignments using IQ-Tree2 version 2.3.6 (www.iqtree.org)[26], applying the maximum likelihood (ML) method with optimal substitution models and 1,000 bootstrap replications. The resulting phylogenetic tree was further refined using the online platform iTOL (https://itol.embl.de/)[27].

Gene structure, conserved motif, protein domains, and *cis*-element analysis

The corresponding coding DNA sequences (CDS) and DNA sequences of the candidate bZIP genes in *C. indicum* were obtained from the genome database. The distributions, positions, and phases of the introns within the *CibZIP* genes were analyzed. The candidate *CibZIP* protein sequences were examined using MEME version 5.5.5 software (https://meme-suite.org/meme/) with the following parameters: a maximum of 15 motifs, a minimum motif width of 6, a maximum motif width of 50, and a motif occurrence distribution of zero or one per sequence. Protein domains were predicted by NCBI Conserved Domain Database. TBtools was employed to integrate and visualize the phylogenetic tree, gene structures, conserved motifs, and domains[28]. Additionally, regions extending 2,000 kb upstream were utilized to identify potential regulatory elements.

GO annotation, KEGG pathway enrichment, and Protein-Protein Interaction (PPI) analysis

Gene Ontology (GO) analysis and KEGG pathway enrichment were conducted for the *CibZIP* genes utilizing the EggNOG database (http://eggnogdb.embl.de/#/app/home). The complete set of genes from *C. indicum* was employed as the reference cohort. Subsequently, GO and KEGG enrichment analysis were executed, with TBtools facilitating the extraction of GO enrichment terms and KEGG pathways that exhibited corrected *p*-values of 0.05 or lower. Statistical analyses and data mapping were performed using the Biozeron Cloud Platform (accessible at www.cloud.biomicroclass.com/Cloud-Platform). Additionally, the bZIP protein sequences were submitted to the STRING database (https://string-db.org/) for comparative analysis of nodes, and the interactions among key proteins were predicted based on known protein interactions in *Arabidopsis*.

Analysis of *CibZIP* genes expression profiles

RNA sequencing (RNA-Seq) data were downloaded and employed to investigate the expression profiles of *CibZIP* genes across various tissues, including roots, buds, tongue flowers, leaves, tubular flowers, and stems, as well as in response to Cd^{2+} stress. RNA-Seq data of different tissues were downloaded from *C. nankingense* genome database (www.amwayabrc.com/zh-cn/download.htm) and Cd stress was obtained from our unpublished data. The expression data, quantified as fragments per kilobase of transcript per million mapped reads (FPKM), were obtained from genome-wide RNA-Seq databases. Heat maps illustrating gene expression values (FPKM) were generated utilizing TBtools software.

Plant materials, growth conditions, and stress treatments

The wild type of *C. indicum* used in this study was obtained from the Chrysanthemum Research Center in Cold Land, Northeast Forestry University, Harbin, China (126°63' E, 45°72' N). The wild-type control used in this study was *Arabidopsis thaliana* Columbia-0 (Col), which were grown in a chamber at 25 °C for 16 h of light and 15 °C for 8 h of darkness, with a light intensity of 300 μmol·m^{-2}·s^{-1}. Two-month-old wild-type chrysanthemum strains were washed in purified water and pre-cultivated in 1/2 Hoagland nutrient solution for 3 d. Then the 1/2 Hoagland nutrient solution was replaced and 200 μmol·L^{-1} of CdCl$_2$·2.5H$_2$O was added to the treatment group and no Cd was added to the control (CK) group. Leaves were sampled respectively at 1, 2, 6, 12, and 24 h after the treatment, and immediately placed in liquid nitrogen. Subsequently, RNA was extracted to determine the expression of *CibZIPs*. Each treatment encompassed 12 biological replicates. The samples of leaves and roots were separately collected from the Cd-treated group for 24 h and the control group, with three replicates for each group. A total of 12 cDNA libraries were dispatched to Baimike Company for transcriptome sequencing. The wild type and three T3 purist overexpression lines were cultured on 1/2 Murashige and Skoog (MS) solid medium for 5 d and then transferred to 1/2 MS solid medium supplemented with CdCl$_2$·2.5H$_2$O at concentrations of 100 μmol·L^{-1} for 7 d of vertical incubation before photographs were taken to record root length. Seedlings under normal conditions were used as the control. The 1-month-old wild-type and transgenic seedlings were washed in sterile water, transplanted into 1/2 Hoagland nutrient solution containing 100 μmol·L^{-1} CdCl$_2$·2.5H$_2$O, the sample without added Cd^{2+} serves as the control, and mixed for samples after 3 d of treatment, and the samples were stored in a −80 °C refrigerator for subsequent assays, with 12 biological replicates per treatment.

RNA isolation, cDNA synthesis, and quantitative PCR analysis

Total RNA was extracted using a Plant RNA Kit (OMEGA, USA). ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO, Japan) was used for cDNA synthesis. The quantitative real-time PCR reaction was performed to analyze the relative transcript levels of selected genes with the Light-Cycle 96 instrument using the SYBR Green Premix *Pro Taq* HS qPCR Kit III (Low Rox Plus). The reaction was performed as follows: 95 °C for 30 s, followed by 40 cycles of 95 °C, for 5 s, and 55 °C for 30 s, 72 °C for 30 s. Each reaction was performed with three biological replicates, and relative expression was calculated using the 2$^{-\Delta\Delta C_T}$ method. The results were analyzed as mean ± SE. The primers used in this study were designed with Primer5.0 and are listed in Supplementary Table S1. The *CmEF1α* (KF305681) served as an internal control[29].

Generation of *CibZIP29*-overexpressing Arabidopsis

The PCR reaction consisted of 0.5 μL of DNA, 0.75 μL of forward primer, 0.75 μL of reverse primer, 7 μL of mix, and 16 μL of RNase-free water, according to the specifications. The PCR program used was 94 °C for 2 min; 35 cycles of 98 °C for 10 s, 60 °C for 30 s, and 68 °C for 1 min; and a final hold at 4 °C until further analysis. The PCR products were sequenced and inserted into a modified pBI-121-GFP vector[30]. The constructed vector was transformed into the *Agrobacterium tumefaciens* strain GV3101 using the freeze-thaw method. Subsequently, the floral dip method was used to get the transgenic *Arabidopsis* of *CibZIP*. Positively transformed lines were identified using PCR and quantitative PCR. The primers used are listed in Supplementary Table S1. *AtEF1a* (At5g60390) was employed as the internal reference gene[31].

Subcellular localization analysis in tobacco and transcriptional activation assay

Protein localization was conducted utilizing a method previously outlined by Li et al.[32]. The complete coding sequence (CDS) of *CibZIP29* was cloned and subsequently inserted into the pBI-121-GFP vector, facilitating the fusion with a green fluorescent protein (GFP) tag. The resulting recombinant plasmid, along with the empty vector, was introduced into the *Agrobacterium tumefaciens* strain GV1301. Tobacco plants (*Nicotiana benthamiana*) were cultivated in a greenhouse under a 16-h light and 8-h dark photo-period at 25 °C. Following this, *Agrobacterium* was infiltrated into the leaves of 4-week-old *N. benthamiana* plants. The infected plants were maintained under standard growth conditions for 2 d after an initial 24-h period in darkness. GFP fluorescence was subsequently visualized using a laser confocal scanning microscope (Zeiss LSM800, Germany). The pGBKT7 vector underwent a double digestion process utilizing the restriction enzymes *EcoRI* and *BamHI*. Subsequently, the complete coding sequence of *CibZIP29* was ligated into the pGBKT7 vector. The recombinant vectors were transformed into yeast cells Y2HGold. The transformed yeast cells underwent cultivation on SD/-Trp (Coolaber, Beijing, China) and SD/-Trp/-His/-Ade+X-α-gal media at 28 °C for 3 d before being photographed.

Physiological index measurement of transgenic plants and statistical analysis

Three days after Cd exposure, the plants were rinsed three times with deionized water, moisture dried off, and the fresh weight of the entire plant then weighed. Subsequently, the plants were dried to a constant weight and the dry weight then measured (80 °C, 72 h). Each treatment encompassed three biological replicates, and each plant was weighed three times. The chlorophyll content was determined by the method previously published by Arnon[33]. Chlrorphyll was extracted using 80% acetone and subsequently measured with a UV spectrophotometer (Agilent Cary 60 UV-Vis, USA). The activities of antioxidant enzymes, including superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT), in the leaves, were measured using analytical kits (Keming, Suzhou, China). The MDA, REL, and proline contents in the leaves were estimated as previously described[34,35]. Cd content was measured according to a previously described method, with some modifications[36]. Briefly, approximately 0.5 g of dry sample was ground and digested with 1 mL of HCl and 4 mL of concentrated HNO_3 for 6 h at 160 °C. The samples were cooled and diluted to 25 mL with 5% HNO_3. An Atomic Absorption Spectrometer (Agilent 4210 MP-AES, USA) was used to determine the Cd concentration. The transfer coefficient was calculated from the presence of Cd in the leaves and stems compared to that in the roots[37]. All data are expressed as the mean of three technical replicates ± standard deviation. Statistical analyses were performed using GraphPad Prism 10.1.2 software to determine significance levels. The data were analysed using One-way ANOVA with Tukey's post-hoc test, and comparisons between mean values under each treatment were considered statistically significant at $p \leq 0.05–0.001$.

Results

Characterization of *bZIP* genes in *C. indicum*

In the present study, we combined the results of HMM and BLASTP methods, and a total of 65 *bZIP* genes were identified in the *C. indicum* genome. These recently identified members have been assigned nomenclature that includes the prefix 'Ci', which denotes *C. indicum*, followed by numerical designations that correspond to their chromosomal distribution. The comprehensive overview of the physical and chemical properties associated with the *CibZIP* genes is presented in Supplementary Table S2. The results showed significant variation in protein lengths, with the longest CibZIP protein, CibZIP36, consisting of 613 amino acid residues, while the shortest, CibZIP34, comprised merely 118 amino acid residues, corresponding to protein molecular weights of 13.844 kDa (CibZIP34) to 67.464 kDa (CibZIP36). The theoretical isoelectric point (pI) varied from 4.68 (CibZIP21) to 10.05 (CibZIP56). Predictions regarding subcellular localization suggest that a significant proportion of *CibZIP* genes were located in the nucleus, while a smaller number were found in the chloroplasts (CibZIP27) and mitochondria (CibZIP44).
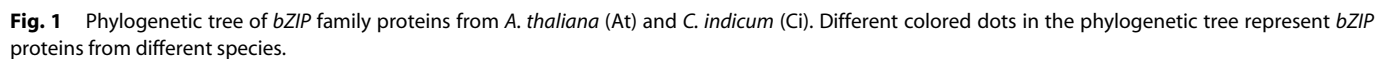
Phylogenetic analysis of *CibZIPs*

To examine the evolutionary relationships and classification of the *bZIP* family in *C. indicum*, an unrooted phylogenetic Maximum Likelihood (ML) tree was constructed with the 65 *CibZIPs* and the known *bZIPs* from *Arabidopsis* using IQtree2 software based on the sequence alignment of the full-length sequences (Fig. 1). The protein sequences of *C. indicum* that were classified alongside *AtbZIPs* have been identified as constituents of the respective subclade within *Arabidopsis*. The results revealed that the *CibZIPs*, together with the members from Arabidopsis, were grouped into 12 subgroups. Overall, the quantity of *CibZIP* genes was comparable to the number of corresponding members in *Arabidopsis* within the same subgroups. The *CibZIP* genes categorized within subgroup S exhibited the highest membership, comprising 14 members in *C. indicum*. In contrast, subgroup B, H, J, and K were characterized by the lowest number of members, with each containing only one member. We also constructed the evolutionary tree for *bZIP* members of the *C. indicum* (Supplementary Fig. S1) and found that the results of classification were consistent with our results above (Fig. 1), demonstrating that the evolutionary trees we constructed were reliable.

Chromosome distribution and collinearity analysis of the *CibZIP* gene family

The expansion of gene families represents a significant mechanism contributing to the evolutionary adaptability of plants. To identify gene duplication events and examine these occurrences, we initially conducted a mapping of the physical distributions of the *CibZIP* genes. As shown in Fig. 2, all the *CibZIP* genes were unevenly distributed across nine chromosomes. Chromosome 08 was the most prominent *bZIP* and contained 12 members (18.5%). Chromosome 01 occupied the second largest number of *CibZIPs*, with a total of nine. Chr 05 and Chr 07 contained eight members, respectively. Chr 04 and Chr 09 occupied seven members, respectively. Chr 02, Chr03, and Chr06 harbored five, six, and three *CibZIP* genes, respectively.

Gene duplication serves as a significant mechanism contributing to the expansion of gene numbers and the emergence of evolutionary innovations. To clarify the mechanisms responsible for the expansion of the *CibZIP* gene family in *C. indicum*, the tools BLASTP and MCScanX were utilized to detect gene duplication events. As shown in Fig. 2 and Supplementary Table S3, a total of three pairs of tandem duplication events were identified and the three duplicated genes (*CibZIP33*, *CibZIP34*, *CibZIP35*) were located on the same chromosome 05 from the same subgroup C. We also conducted a comprehensive analysis of the segmental duplication events associated with the *CibZIP* genes through covariance analysis utilizing the MCScanX tool. According to the results (Fig. 3; Supplementary Table S3), we detected nine putative paralogous gene pairs in the *C. indicum* *bZIP* gene family. Among these segmental duplication events, subgroup S contained the majority *CibZIP* genes, with a number of 6 (9.2%) involving six pairs (66.7%). In general, roughly 21.5% (14/65) of *CibZIP* genes might have participated in tandem or segmental duplication, suggesting that these gene duplication

Fig. 1 Phylogenetic tree of *bZIP* family proteins from *A. thaliana* (At) and *C. indicum* (Ci). Different colored dots in the phylogenetic tree represent *bZIP* proteins from different species.

occurrences could be pivotal for the expansion of this gene family. The findings indicated that all Ka/Ks ratios were below 1.0, implying that the gene pairs experienced significant purifying selection following the duplication of deleterious mutations in the evolutionary development of *bZIPs*.

To enhance our comprehension of the expansion mechanisms within the *bZIP* family of *C. indicum*, we developed comparative syntenic maps that include *C. indicum* in relation to five representative species (*O. sativa, A. thaliana, G. max, P. trichocarpa,* and *V. vinifera*) (Fig. 4; Supplementary Table S4). As the results showed, the *bZIP* genes in *C. indicum* exhibited the highest number of homologous gene pairs with *Glycine max*, totaling 46 orthologous gene pairs. This was followed by *V. vinifera* and *P. trichocarpa*, which displayed 42 and 40 orthologous gene pairs, respectively. In contrast, the number of homologous gene pairs identified between *C. indicum* and *A. thaliana* was significantly lower, with only seven orthologous gene pairs and even fewer were found with *O. sativa*, which had one orthologous gene pair.

Exon–intron organization, conserved motifs, and domain architectures

Through the comparative analysis of the cDNA sequence and the genomic sequence of the *CibZIP* gene, alongside the data derived

from phylogenetic evolutionary trees, the structural characteristics of exons and introns within the *CibZIP* gene were elucidated in accordance with the classifications established by the evolutionary tree (Fig. 5). Fourteen of the 65 *CibZIP* genes had no introns, and the rest had introns ranging from 0 to 12, with more *CibZIP* genes in groups D and G than in the other groups. Most of the genes in each clustering group have similar intron and exon lengths and distributions, and the gene structures were highly similar. However, there were also cases where genes in the same cluster differed in exon/intron arrangement. For example, the vast majority of members in group I contain three introns, and only one gene contains four introns (*CibZIP25*). All members had one or more UTRs in untranslated regions except for *CibZIP17*.

An analysis of potential conserved motif compositions within the *bZIP* gene family was conducted to enhance our understanding of the diversity and functional characteristics of *bZIP* genes in *C. indicum*. The MEME program was employed to identify and annotate the sequences of these motifs, with comprehensive details provided in Supplementary Table S5. The results (Fig. 5) showed that among the 10 conserved motifs retrieved, only *bZIP* (motif 1) was the core motif, which was shared by all members of the *bZIP* family. However, its position on the gene varies among family
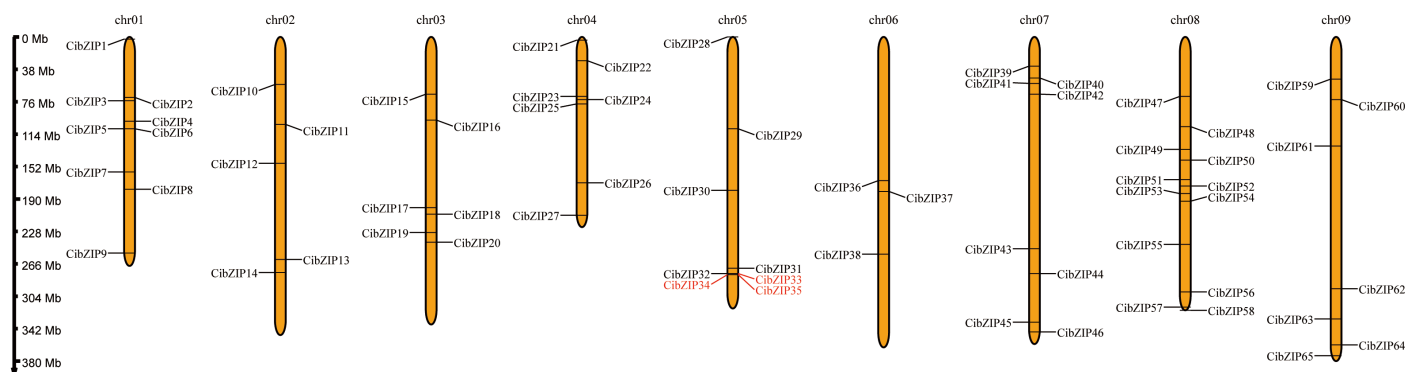
Fig. 2 Chromosome localization of *CibZIPs*. Each brown bar represents a chromosome, the chromosome number is marked at the top and the ratio on the left represents the chromosome length with the unit for the scale is mega bases (Mb). The red marked genes represent the collinear relationship within the species.
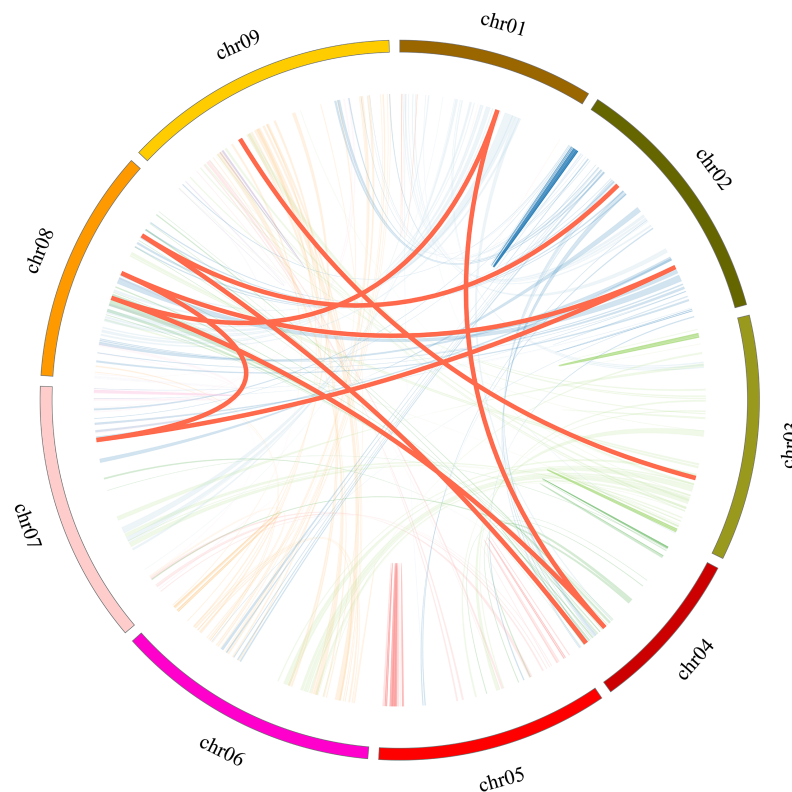

Fig. 3 Collinear relationship of *CibZIP* genes in *C. indicum*. Red curves represent homologous gene pairs. Gray lines in the background indicate the synteny blocks within the whole *C. indicum*.

members. Motif 4 regions were detected in most *bZIP* subgroups, except subgroups F, H, K, and D. Furthermore, a few motifs existed in one or more groups and subgroups. For example, motif 12 was only detected in subgroup I, motifs 2, 3, 5, and 7 only existed in subgroup D. Consistent with the conserved motif prediction, we found that all members contain a conserved structural domain of *bZIP* family based on the results of CDD (NCBI Conserved Domain Database) (Fig. 5), including MFMR, BRLZ, bZIP, COG4372, and so on.

Different subgroups have different distributions of protein conserved motifs, which is the basis for the functional diversity of *CibZIP* genes and provides a reference for the study of functional differentiation among subgroups. In the same subgroup, the distribution of protein conserved motifs and domains were the same, indicating that genes in the same subfamily have similar functions and closer homology relationships, which also proved the accuracy of the phylogenetic tree constructed in this study.

Enrichment analysis of GO annotation, KEGG pathway, and *Cis*-elements analysis in *CibZIPs*

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted to further elucidate the functional roles of the *CibZIP* gene family (Fig. 6; Supplementary Table S6). The results of the GO analysis revealed that these 65 CibZIP proteins were involved in 15 biological processes (Bp) and one cell position (Cc). In the Bp ontology, the genes were primarily involved in biosynthetic process, cellular processes, metabolic processes, biological regulation, and so on. As for the cellular components, the main category was obsolete cell parts. KEGG pathway analysis provides a broader understanding of the metabolic and signaling pathways in which *CibZIP* genes were involved. The analysis showed significant enrichment in pathways related to 'Circadian rhythm', 'Plant hormone signal transduction', and 'MAPK signaling
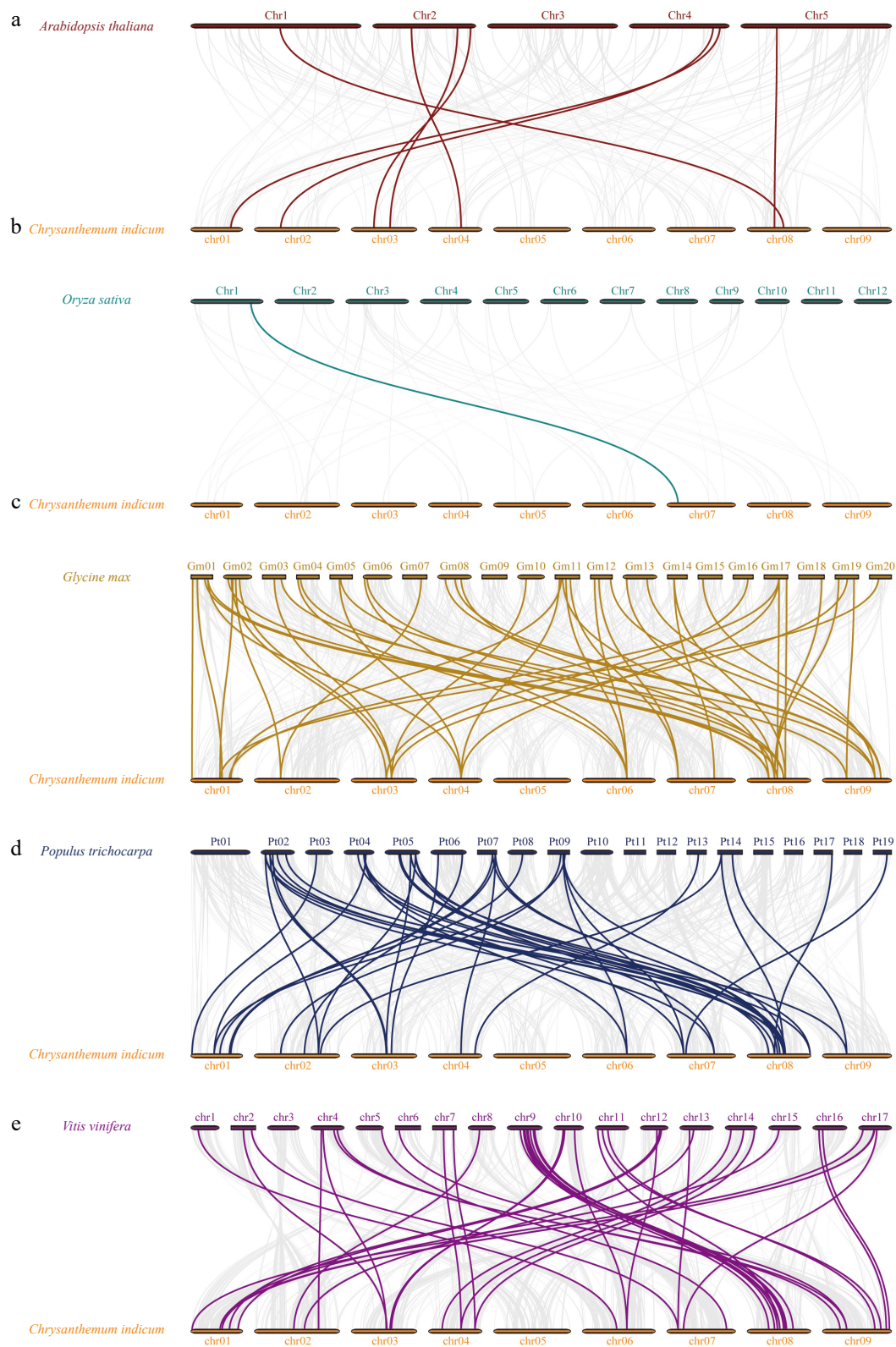
Fig. 4 Collinear analysis of *CibZIP* genes from *C. indicum* with five representative species. Synteny relationships of *bZIPs* between *C. indicun* with (a) *A. thaliana*, (b) *O. sativa*, (c) *G. max*, (d) *P. trichocarpa*, and (e) *V. vinifera*.

pathway'. These pathways are crucial for plant adaptation to environmental stressors and developmental regulation.

Cis-acting regulatory elements, located in the promoter regions of genes, are essential for the precise regulation of gene expression. To gain insights into the regulation of *bZIP* genes, an analysis of their promoter regions was performed, focusing on identifying and characterizing *cis*-acting elements. The promoter regions situated within 2.0 kb upstream of *CibZIP* genes were analyzed using the PlantCARE database to predict *cis*-regulatory elements, aiming to clarify the mechanisms that regulate *CibZIPs* in response to abiotic
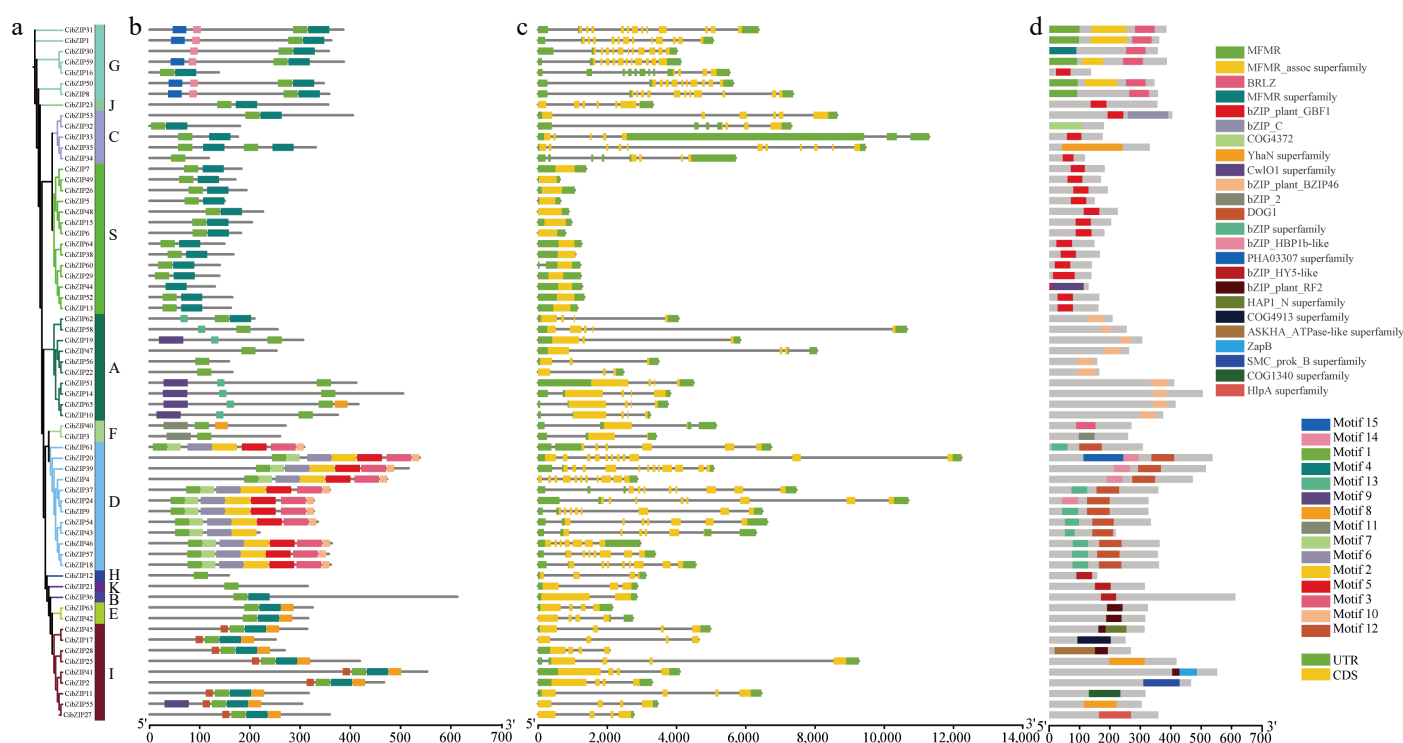
Fig. 5 Phylogenetic relationships, gene structures, conserved motifs and domains of *CibZIPs*. (a) The phylogenetic tree of 65 *CibZIP* family members. (b) Composition and distributions of conserved motifs in *CibZIP* proteins. (c) Exon-intron organization of *CibZIP* genes. Exons are shown as yellow rectangles; introns are shown as black lines, and the untranslated regions (UTR) are shown as green rectangles. (d) Domains distribution. The conserved domains are represented by different rectangle colors.

and biotic stresses. Detailed information on putative *cis*-elements are listed in Supplementary Table S7. In summary, *cis*-acting regulatory elements can be primarily categorized into many distinct groups. However, we selected particularly abundant and more important *cis*-elements to analyze, which can be grouped into three major categories: hormone-responsive, development regulation, and environmental stress response (Fig. 7). The primary category of elements linked to hormone responsiveness predominantly exhibited correlations with abscisic acid (ABA) response elements (ABRE), auxin response elements (AuxRR core, TGA-element), jasmonic acid (JA) motifs (CGTCA-motif, TGACG-motif), ethylene response elements (ERE), gibberellin (GA) motifs (GARE-motif, P-box, TATC-box), and salicylic acid response elements (TCA-element). These elements were extensively identified within the promoter regions of the *CibZIP* genes. After the identification of hormone-responsive *cis*-acting regulatory elements (CAREs), it was determined that environmental stress response elements represent the second most prevalent category within the *CibZIP* transcription factor family. Among these elements, the anaerobic response element (ARE) is deemed critical for the induction of anaerobic conditions, while the dehydration-responsive element (DRE) core is implicated in responses to dehydration, low temperature, and salinity stresses. Additionally, the GC-motif is associated with specific inducibility under anoxic conditions, the low-temperature responsive element (LTR) pertains to responses to low temperatures, and the MYB binding site (MBS) is linked to drought-induced responses. Furthermore, TC-rich repeats are involved in defense mechanisms and stress responses, the wound-responsive element (WUN-motif) is recognized for its role in wound response, and the W box is associated with sugar metabolism and signaling in plant defense. Specifically, ARE, TC-rich repeats, and W-box were identified as the most prevalent, each exhibiting over 11 copies across *CibZIPs*. Furthermore, various

elements linked to developmental regulation were identified. Specifically, the CAT-box, CCGTCC-box, circadian elements, GCN4 motif, O_2-site, and RY-element were observed. These elements are typically associated with plant growth and development, either in relation to the cell cycle and cellular proliferation responses or as necessary components for tissue-specific gene expression.

Expression patterns of *CibZIP* genes in various tissues and under Cd stress based on transcriptome data

To thoroughly elucidate the expression characteristics of the *CibZIP* genes, we analyzed of expression patterns utilizing RNA-seq datasets, including various tissues and under Cd stress. As illustrated in Fig. 8a and Supplementary Table S8, the expression profiles of *CibZIP* genes exhibit variability across distinct tissues. The varying expression patterns observed for each gene across different organs suggest that *CibZIPs* serve distinct functions within *C. indicum*. The expression of all *CibZIP* genes was identified in at least one tissue type, except CibZIP33. Specifically, 47, 38, 46, 50, 47, and 45 genes exhibited high transcript abundance (FPKM > 5) in the bud, leaf, tubular flowers, tongue flowers, root, and stem tissues, respectively. A total of 33 *CibZIP* genes were generally highly expressed in all tissues. Conversely, certain *CibZIPs* demonstrated transcript accumulation patterns that were specific to particular tissues, indicating a potential functional diversification of *CibZIP* genes throughout the processes of growth and development. For example, 10 *CibZIPs* were expressed at a very low level in all tested tissues (*CibZIP1*, *CibZIP2*, *CibZIP4*, *CibZIP20*, *CibZIP23*, *CibZIP28*, *CibZIP29*, *CibZIP30*, *CibZIP41*, and *CibZIP61*). Furthermore, specific *CibZIP* genes exhibited elevated expression levels in certain tissues, as indicated by FPKM values exceeding 20. For instance, *CibZIP11*, *CibZIP25*, and *CibZIP27* exhibited elevated expression levels in buds compared to other tissues, while *CiTCP3* and *CiTCP64* showed greater transcript abundance in tubular flowers relative to other tissues. These
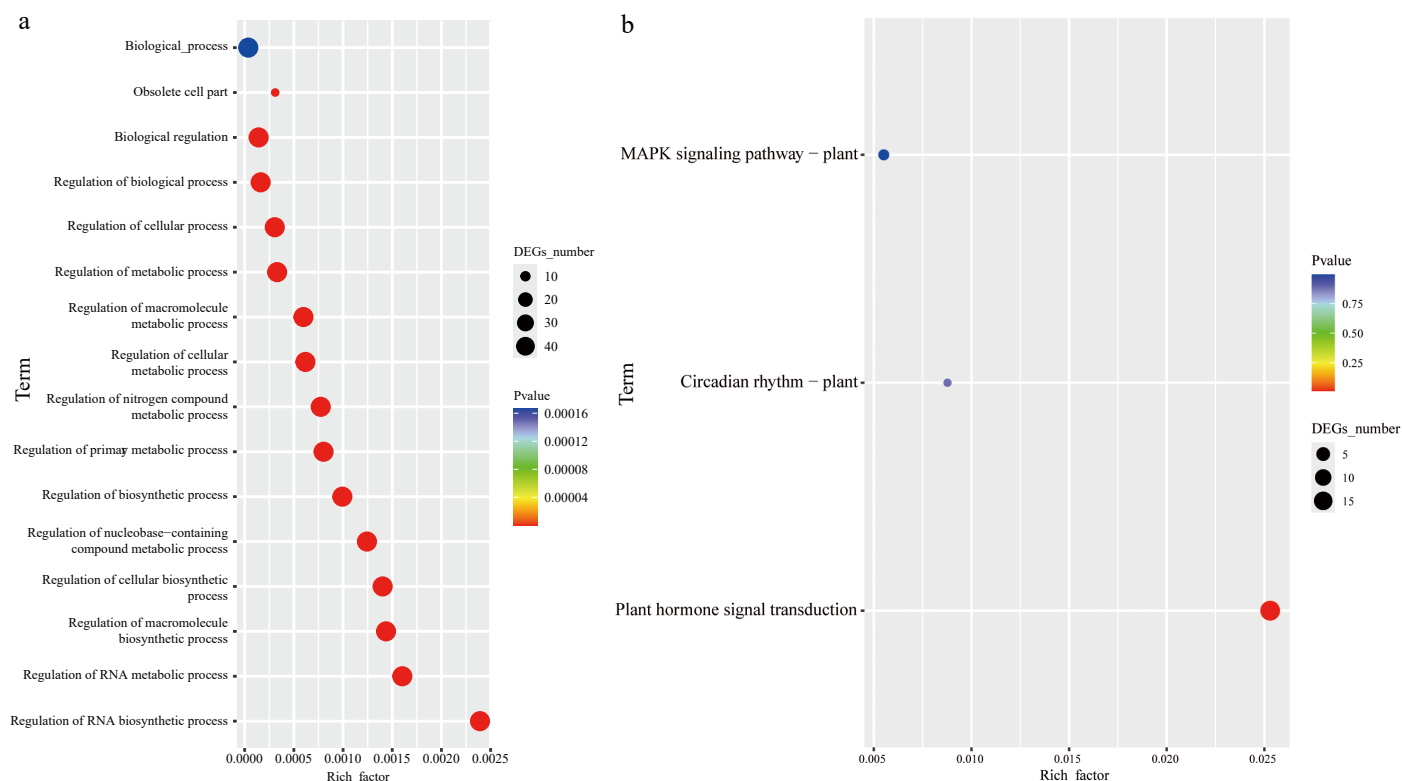
Fig. 6 Functional annotation of *CibZIPs*. (a) Highly enriched GO terms in *CibZIPs*; (b) highly enriched KEGG pathways in *CibZIPs*.

findings indicate the significant roles of these genes in tissue development. It is important to highlight that certain duplicated genes demonstrate analogous expression patterns. For instance, *CiTCP7* and *CiTCP26* exhibited relatively high expression levels in tongue flowers, *CiTCP27* and *CiTCP55* exhibited relatively high expression levels in the bud, while their expression was markedly low in other tissues.

Under Cd stress (Fig. 8b & c; Supplementary Table S8), many genes showed significant up-regulated or down-regulated compared with the control group in leaves, for example, *CibZIP22*, *CibZIP33*, *CibZIP34*, *CibZIP35*, and *CibZIP37* were induced throughout the Cd stress treatment. In contrast, seven *CibZIP* genes (*CibZIP5*, *CibZIP6*, *CibZIP15*, *CibZIP20*, *CibZIP23*, *CibZIP38*, and *CibZIP64*) were significantly down-regulated. In roots, *CibZIP12* and *CibZIP29* were dramatically up-regulated, with more than a 3-fold increase after treatment compared with the control. We also found that some genes showed different expression trends in roots and leaves, for example, *CibZIP7*, *CibZIP28*, *CibZIP38*, *CibZIP48*, and *CibZIP49* were found to be down-regulated in leaf tissues while exhibiting up-regulation in root tissues. Also, *CibZIP19*, *CibZIP22*, *CibZIP39*, *CibZIP47*, *CibZIP56*, and *CibZIP61* were down-regulated in roots but up-regulated in leaves after Cd treatment. Under Cd stress, these genes exhibit contrasting expression patterns across various tissues, indicating that they may play different roles in several biological regulatory processes.

The qRT-PCR analysis of *CibZIP* genes under Cd stress

To clarify the expression profiles of *CibZIP* gene members, we performed qRT–PCR analysis to evaluate the transcript levels of 15 selected candidate genes at various time intervals under Cd stress (Fig. 9). The qRT-PCR analysis indicated that the majority of the *CibZIP* genes examined exhibited up-regulation in response to Cd stress. Furthermore, the expression levels of these genes demonstrated a significant correlation with the RNA sequencing data.

Under Cd treatment, there was a potential induction in the expression of all examined *CibZIP* genes. The findings indicated that all the chosen genes exhibited significant up-regulation or down-regulation following Cd treatment at the 1-h mark, suggesting that these genes are modulated to facilitate a swift response to Cd-induced stress. Four members (*CibZIP23*, *CibZIP30*, *CibZIP37*, and *CibZIP63*) showed similar expression patterns, characterized by an initial increase followed by a subsequent decrease, they reached their peak value at 2, 12, 6, and 24 h, respectively. Ten members demonstrated a similar pattern, characterized by an initial decrease followed by an increase in expression levels. Among these, five members (*CibZIP3*, *CibZIP5*, *CibZIP29*, *CibZIP33*, and *CibZIP62*) achieved their maximum expression at 24 h, two members (*CibZIP12*, *CibZIP56*) peaked at 12 h, while three members (*CibZIP21*, *CibZIP36*, and *CibZIP45*) exhibited lower expression levels at 24 h compared to their untreated state. Furthermore, the expression levels of *CibZIP6* were initially reduced after Cd treatment and subsequently aligned with those observed at the 1-h mark.

Subcellular localization and transcriptional activity of CibZIP29

To ascertain the subcellular localization of CibZIP29, we engineered a 35S::CibZIP29-GFP vector and subsequently transiently expressed it in the leaves of *N. benthamiana* via *Agrobacterium*-mediated transformation. The fluorescence emitted by the 35S::CibZIP29-GFP vector was exclusively observed in the nucleus, in contrast to the unfused GFP vector, which exhibited fluorescence throughout the entire cell (Fig. 10a). These findings suggest that CibZIP29 is localized within the nucleus. To evaluate the transcriptional activity of *CibZIP29*, both the full-length and truncated sequences of *CibZIP29* were cloned into the pGBKT7 vector and subsequently transformed into Y2HGold yeast cells. The yeast strains harboring the full-length sequence or the N-terminal fragment exhibited normal growth and developed a blue coloration on the SD/-Trp/-His/-Ade/+X-α-gal
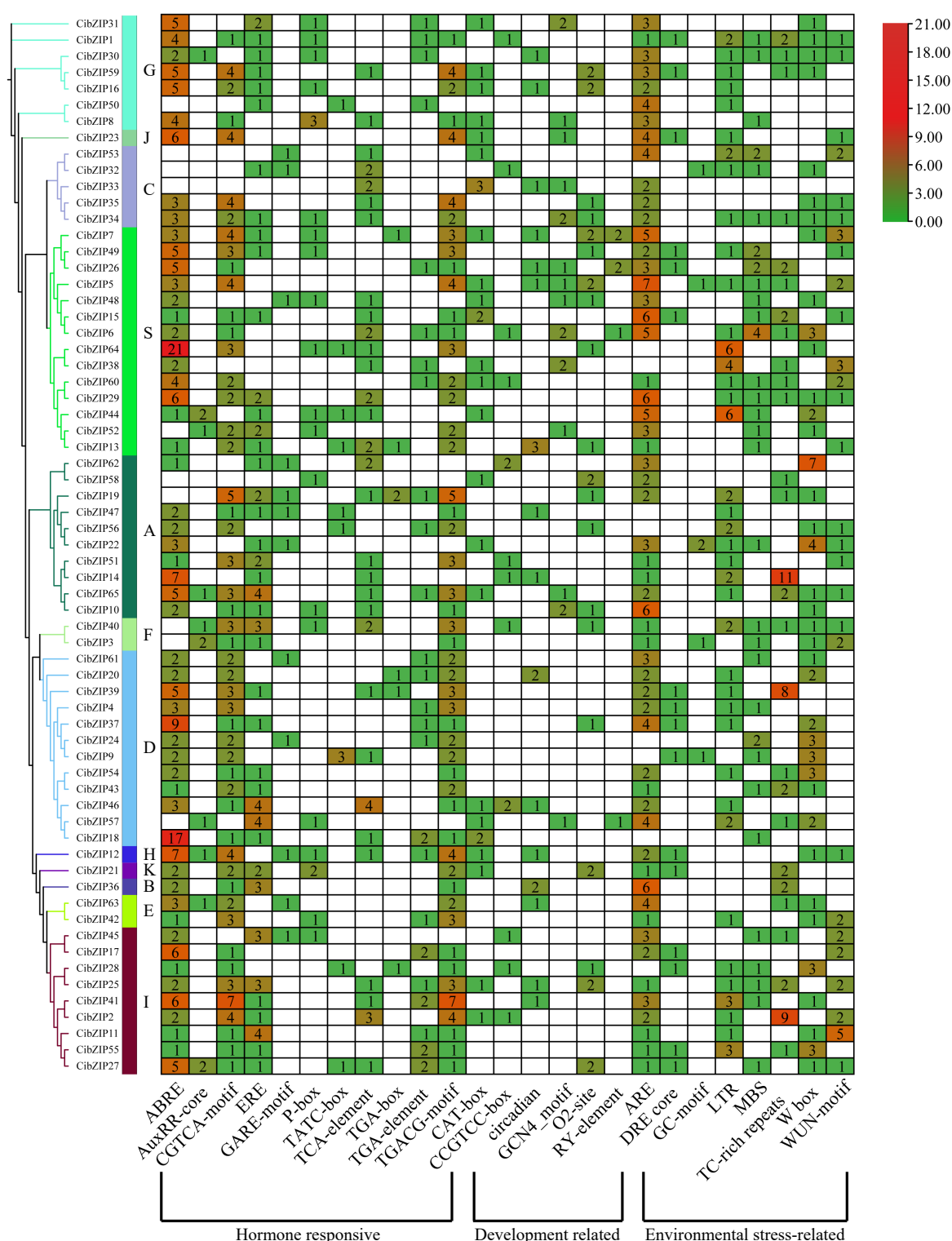
Fig. 7 Analysis of the putative *cis*-elements in promoters of *CibZIP* genes. The phylogenetic tree of 65 *CibZIPs* is presented on the left. The different colors and numbers in the heatmap represent the quantity of *cis*-acting elements in the promoter region of each *CibZIP* gene.

medium. In contrast, the yeast strains containing the C-terminal fragment failed to grow (Fig. 10b). These findings indicate that *CibZIP29* possesses transcriptional activity in yeast, with the N-terminal region being critical for this function.

Overexpression of *CibZIP29* conferred Cd stress in transgenic *A. thaliana*

To investigate the potential roles of *CibZIP29* under Cd stress, we conducted three transgenic *Arabidopsis* lines, designated

CibZIP29-OE1, CibZIP29-OE2, and *CibZIP29-OE3,* which overexpress *CibZIP29*, were developed through the construction of overexpression vectors for *CibZIP29* and subsequent genetic transformation of *Arabidopsis*. According to the transcript data and qRT-PCR results, *CibZIP29* was induced by Cd stress. Hence, we subjected wild-type (WT) and various *CibZIP29* overexpression (OE) seedlings to Cd treatment. We assessed root length and chlorophyll concentration as indicators of plant growth. As shown in Fig. 11, no discernible
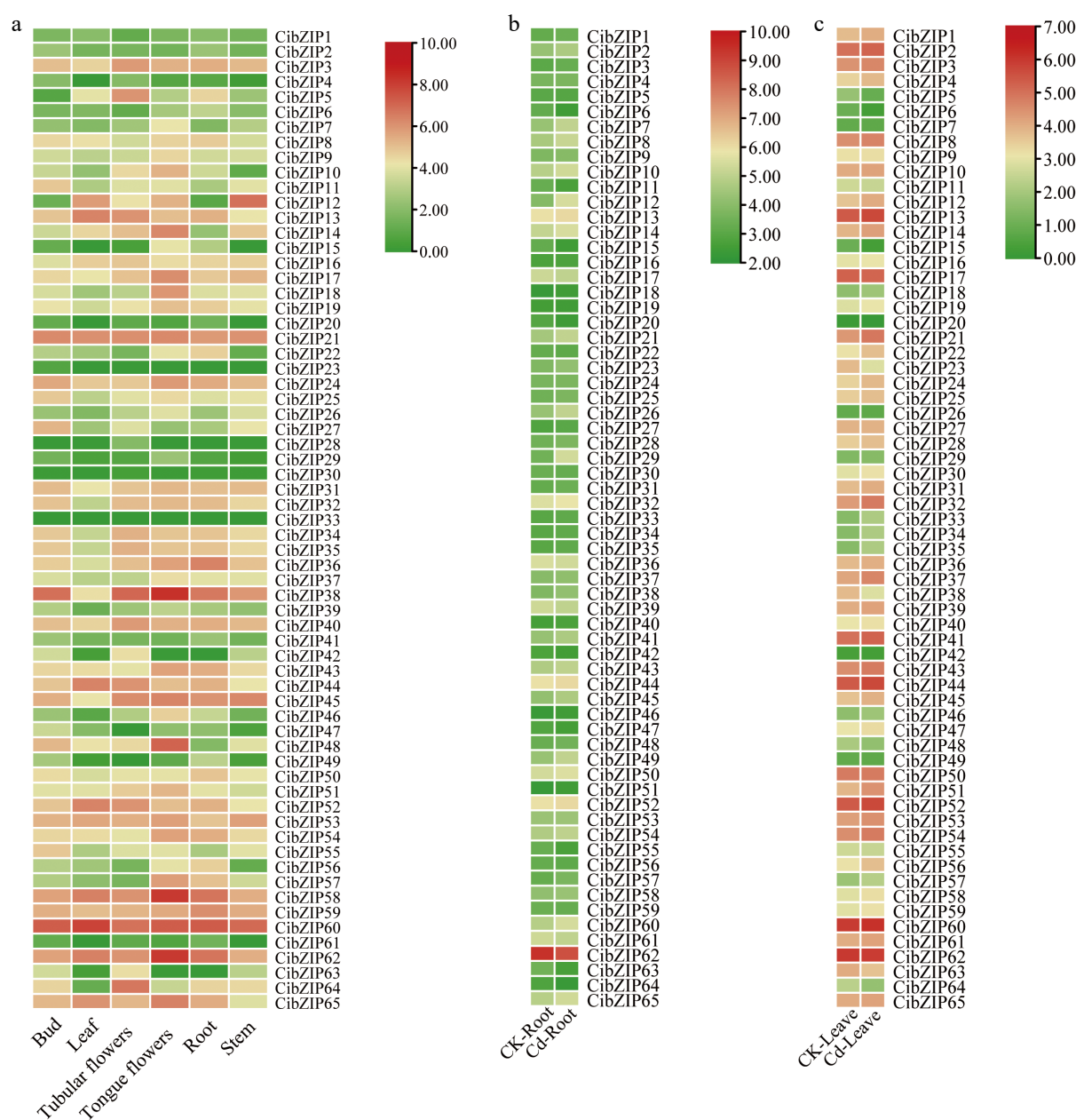
Fig. 8 Expression profiles of *CibZIPs* in different tissues under (a) normal conditions, and (b), (c) in response to Cd stress from RNA-seq data. Color scale at the right of the heatmap shows the expression level, red indicates high transcript abundance while green indicates low abundance.

phenotypic differences were observed between WT and transgenic lines under standard conditions (without Cd treatment). When subjected to Cd treatment, wild-type *Arabidopsis* plants displayed a Cd-sensitive phenotype in contrast to the two lines exhibiting over-expression of *CibZIP29*. In addition, Cd treatment resulted in a reduction in the growth of all plants (Fig. 11a). The root length of transgenic lines was notably greater in comparison to wild-type (WT) plants. Cd treatment also resulted in a reduction of chlorophyll content across all plants. Nevertheless, the transgenic lines exhibited a markedly higher chlorophyll content in comparison to the wild type (WT). The findings indicate that the overexpression of *CibZIP29* mitigated the growth inhibition associated with Cd stress. Furthermore, the Cd content in the roots and leaves of *CibZIP29*-OE plants was significantly decreased, but the transport capacity of Cd to the leaves was enhanced (Fig. 11).

To determine whether the ROS levels in transgenic plants underwent alterations, we conducted measurements of the relative electrical conductivity, MDA content, H_2O_2 content, and superoxide anion content of *CibZIP29*-OE and WT plants under Cd stress. We discovered that under normal circumstances, no conspicuous differences were observed between the WT and transgenic plants. Nevertheless, under Cd stress, in contrast to the WT, the transgenic plants accumulated a lesser amount of ROS (Fig. 12). The results indicated that ROS homeostasis was better maintained in transgenic lines compared with WT under Cd stress.

Additionally, *CibZIP29*-OE seedings exhibited higher activities of superoxide dismutase (SOD), peroxidase (POD), and catalase (CAT), as well as increased proline content (Fig. 12). Proline and these three enzyme substances function as osmotic regulators and ROS scavengers in plants. These results indicated that *CibZIP29* positively regulates *Arabidopsis* resistance to Cd stress.
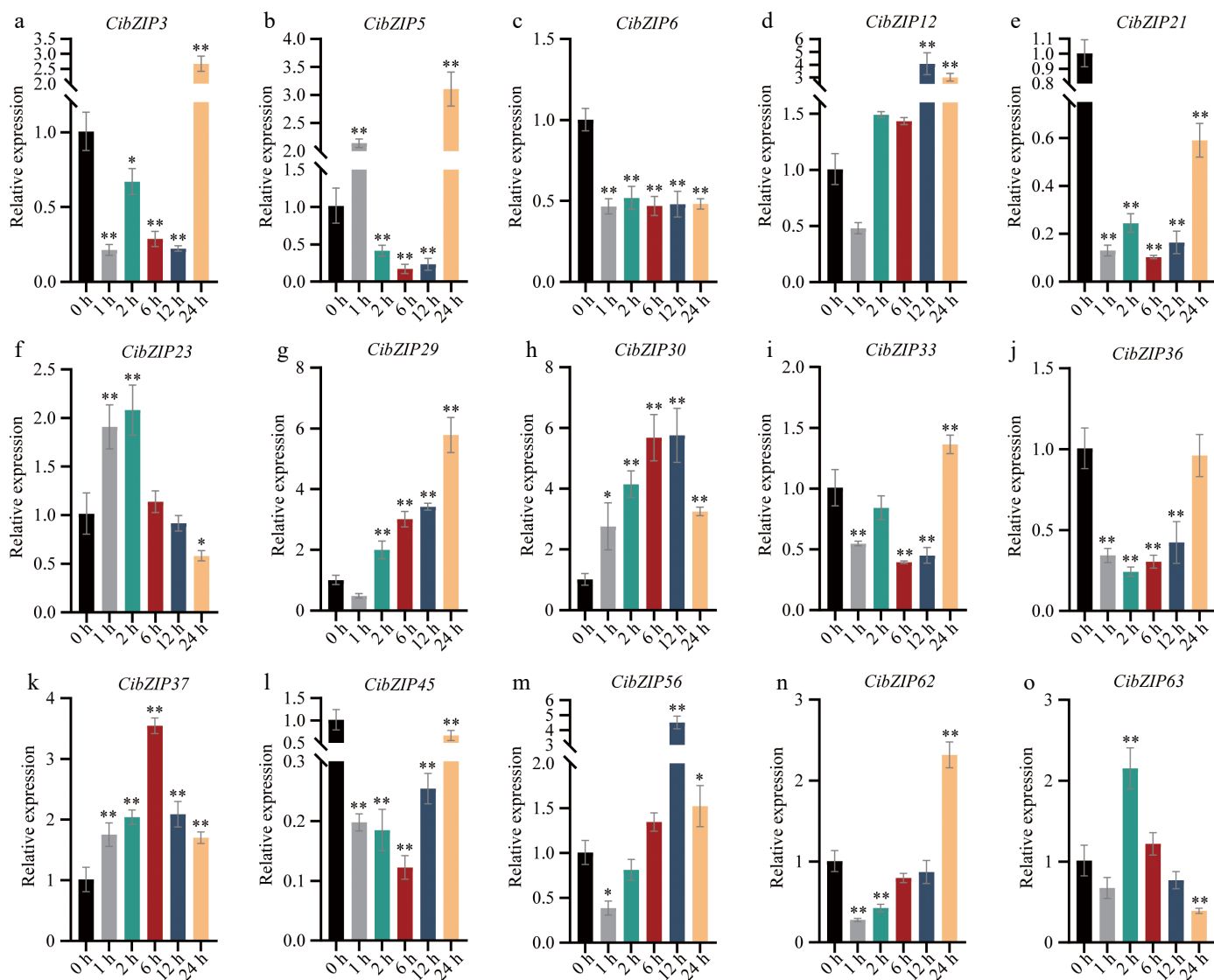
Fig. 9 (a)–(o) Expression pattern of *C. indicum* bZIPs in response to Cd stress determined by qRT-PCR. RNA from leaves at different points. The Y-axis indicates the relative expression level and the X-axis represents different time points after stress treatment taken for expression analysis. The data presented are the average of three biological replicates, the bar represents the standard deviation (T-test, *$p < 0.05$, **$p < 0.01$).
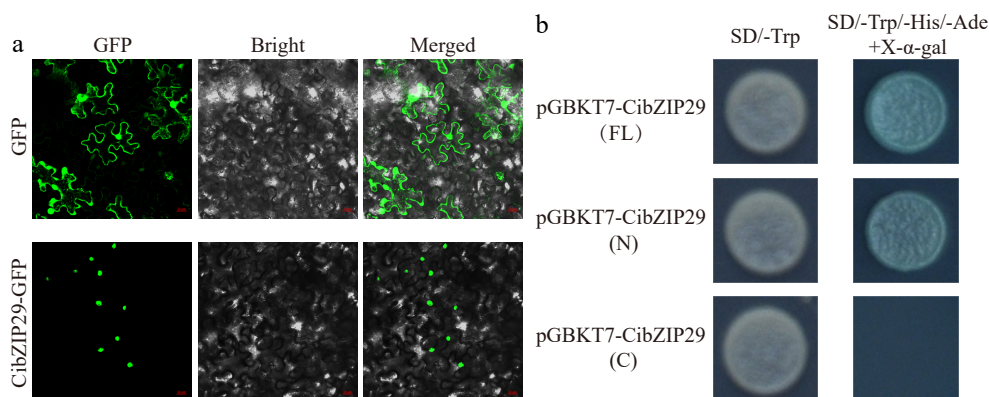

Fig. 10 Subcellular localization and transcriptional activation analysis of CibZIP29 protein. (a) Subcellular localization assay of the CibZIP29 protein. Scale bar: 20 μm; (b) Transcription activation domain identification of CibZIP29 protein in yeast cells.

Discussion

C. indicum possesses a high concentration of bioactive compounds and exhibits a diverse array of pharmacological properties.

Additionally, this plant demonstrates significant resilience to abiotic stressors, making it a valuable candidate for phytoremediation applications. The advancement of sequencing technology has facilitated the availability of genomic data pertaining to *C. indicum*,
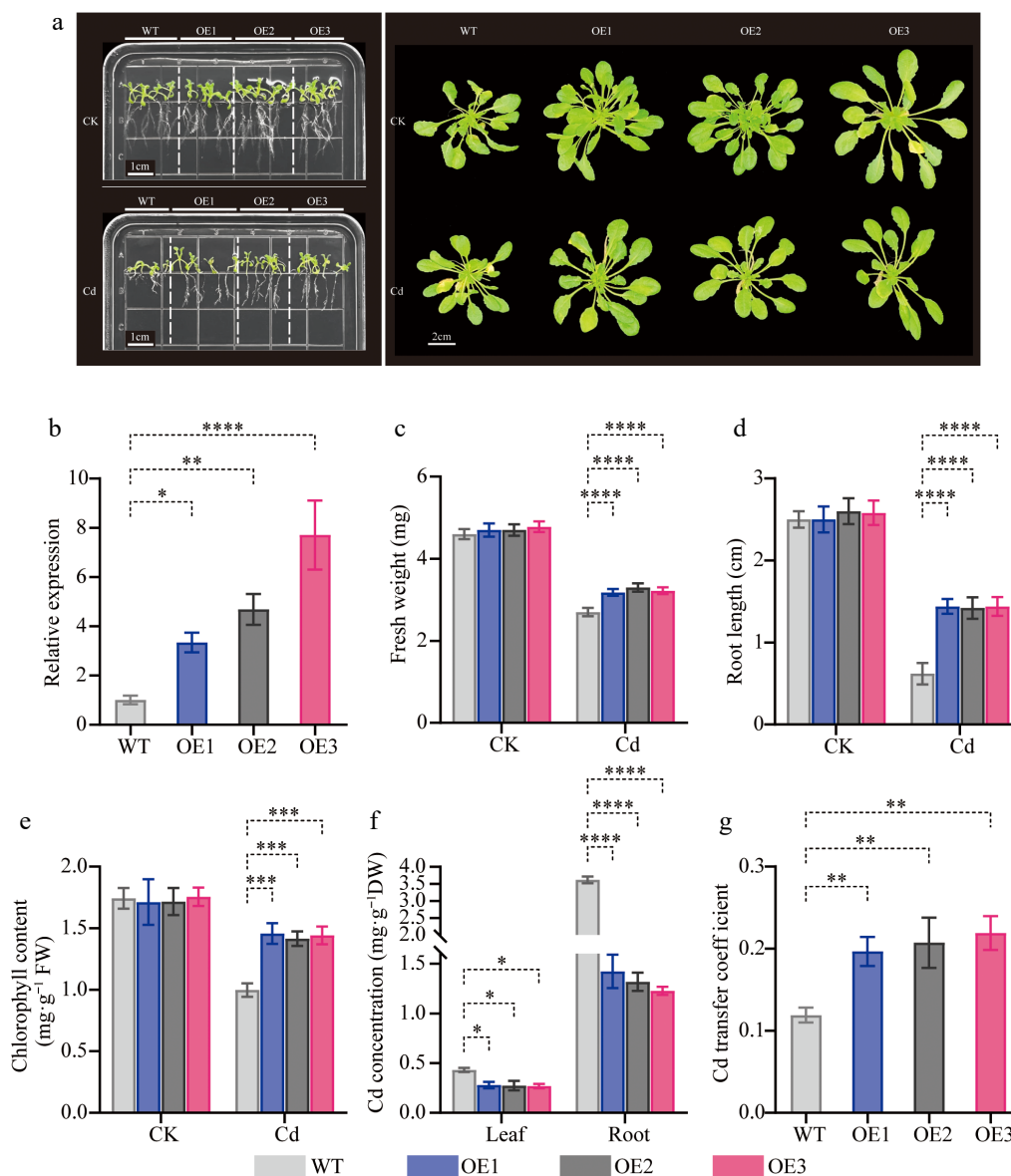
Fig. 11 Phenotypic analysis of transgenic *CibZIP29*-overexpressing *Arabidopsis* seedlings under Cd treatment. (a) Phenotype of transgenic *Arabidopsis* lines and WT plants under Cd stress; (b) The expression level of transgenic lines; (c) Fresh weight; (d) Root length; (e) Total chlorophyll content; (f) Cd concentration; (g) Cd transfer coefficient (T-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

thereby enabling the exploration of genes associated with abiotic stress resistance in *C. indicum*. The *bZIP* transcription factor family, recognized as one of the most extensive families of plant transcription factors, plays a significant role in various developmental processes across plant species. Numerous members of the *bZIP* transcription factor family have been empirically validated in other plant species. Nevertheless, the *bZIP* family has not been documented in *C. indicum*, and the mechanisms underlying the resistance of its members to Cd stress remain unclear.

Identification and characterization of *bZIP* gene family in *C. indicum*

In the current investigation, a comprehensive analysis led to the identification of 65 *bZIP* genes within the genome of *C. indicum*, utilizing bioinformatics techniques grounded in published genomic data. The identified *bZIP* genes, designated from *CibZIP1* to *CibZIP65*, were named according to their respective locations on the chromosome. 63/65 *CibZIPs* were anticipated to be localized within the nucleus, which aligns with the characteristics of transcription factors

and is supported by experimental research conducted in other species, including rice and *Wheel Wingnut*[38]. An examination of the physicochemical characteristics of the proteins encoded by the *bZIP* gene in *C. indicum* revealed significant variability in sequence length, relative molecular weight, and isoelectric point distribution (Supplementary Table S1), which may be due to the fact that *C. indicum* genome has undergone several large-scale replications and the number of *bZIP* genes is large. Based on cluster analysis and the comparative similarity of protein sequences with *Arabidopsis thaliana*, the 65 *CibZIP* proteins have been categorized into 12 subgroups, indicating their close evolutionary relationships. The quantity of subgroups observed aligns with those found in popular and *Arabidopsis*[23], while being lower than that of *Chinese jujube*[39] and exceeding the counts found in grape, tobacco, and buckwheat[23,40]. Among the subgroups, subgroup S exhibited the highest number of *bZIPs*, a pattern that is consistent with observations in *A. thaliana* and tobacco. An examination of the gene structure within the *CibZIP* family revealed notable variations among the

CibZIPs. Our findings indicated that the number of introns ranged from 0 to 11, with 20 % (13/65) of the genes exhibiting an absence of introns and these genes were all from the same subgroup S. This pattern is consistent with observations in other species. We also found that the gain and loss of introns and exons in *CibZIP* gene family, which maybe serve as mechanisms for diversification. In subgroup D, the segmental genes *CibZIP20* and *CibZIP61* exhibit distinct structural characteristics. Specifically, *CibZIP61* comprises six introns and six exons, whereas *CibZIP20* is characterized by 10 introns and 11 exons, indicating the loss of an exon during evolution. A comparable phenomenon has been observed in other species *bZIP* gene families[41,42]. The proliferation of introns appears to facilitate the adaptation of plants to their environments, thereby promoting evolutionary processes. Conversely, genes characterized by a lower intron count may enhance the capacity for rapid environmental responses. The conserved motifs varied in number from 1 to 12, with analogous motifs identified in other plant species belonging to the *bZIP* family. Consequently, the majority of *CibZIP* genes within the same subgroups exhibited comparable gene structures and conserved motifs, suggesting that the functional divergence of *bZIP* genes may be influenced by motifs that are specific to particular groups. Also, these results thereby reinforcing their close evolutionary association and classification within subgroups.

Mechanisms underlying the expansion of the *C. indicum bZIP* gene family

In the current investigation, a total of 65 putative *bZIP* genes were identified within the genome of the *C. indicum*, this number is close to those of *Cucumis sativus* (64), *Pyrus bretschneideri* (62), *Dendrobium catenatum* (62), *Solanum lycopersicum* (69), and *Capsicum annuum* (60), larger than those in *Fragaria vesca* (50), *Beta vulgaris* (48), *Prunus persica* (50), but significantly lower than those in *Gossypium hirsutum* (197), *Phyllostachys edulis* (154), *Triticum*

aestivum (227), *G. max* (160), and *Malus domestica* (112)[43]. In general, factors that can directly affect the number of gene family members may be genome size, duplication events, and polyploidy. For example, 78 *bZIPs* in *A. thaliana* (117 Mb), 89 in rice (466 Mb), 160 in soybean (915 Mb), 125 in maize (2,182 Mb)[44], there exists no correlation between the scale of *bZIP* transcription factors present in a species and the size of its genome. Also, we found that cotton, a 4-fold species, identified 197 *bZIP* members, but strawberry, an 8-fold species, had only 50 *bZIP* members[43]. Therefore, we supposed that the number of *bZIP* family members is also not directly related to the chromosomal ploidy of the species. Taking together with the above discussion and conjecture, it is posited that duplication events may significantly contribute to the diversification of the *bZIP* family.

Prior research has demonstrated that gene replication, encompassing tandem, and segmental duplication as well as whole genome replication, along with polyploidy, significantly contribute to the evolutionary processes of organisms and the diversification of gene families. *C. indicum* has encountered at most two WGD events. The examination of chromosomal distribution and gene duplication revealed that 21.5% (14/65) of the genes were involved in tandem or segmental duplication events (Figs 2 & 3). A total of three tandem duplication events and nine segmental duplication events were detected, suggesting that segmental duplications occur with greater frequency than tandem duplications and both forms of gene duplication have been instrumental in the expansion of the *bZIP* gene family within the *C. indicum* genome. This phenomenon is also observed within the *bZIP* family across various species, for example, there were 15 pairs of segmental genes and five pairs of tandem genes in *Cyclocarya paliurus*, one tandem duplication event and 11 segmental duplication events in *Platycodon grandiflorus*, no tandem duplication and 16 pairs of segmental duplication were observed in
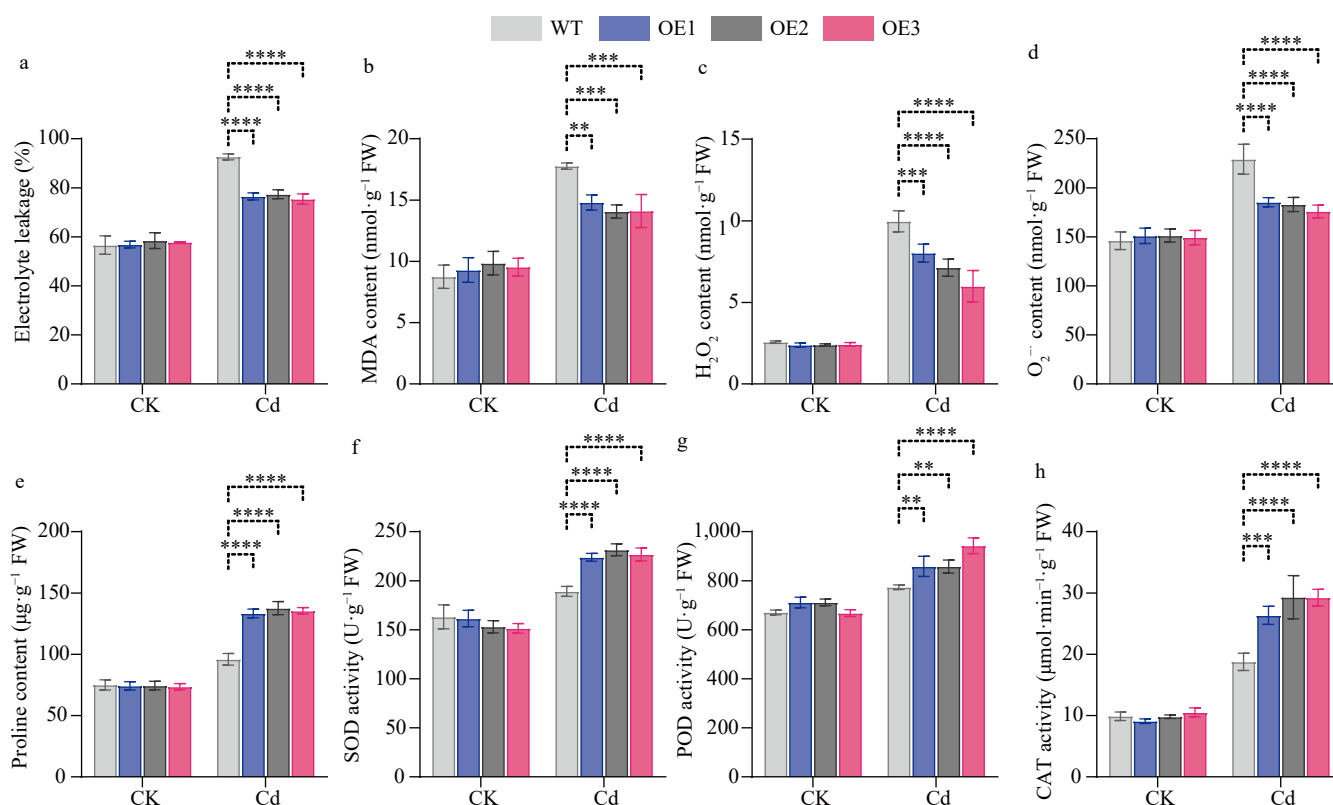

Fig. 12 Physiological indices of WT and transgenic *CibZIP29* lines in *Arabidopsis*. (a) REL; (b) MDA content; (c) H_2O_2 content; (d) O_2^{-} content; (e) Proline content; (f) SOD activity; (g) POD activity; (h) CAT activity (T-test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$).

the *NtbZIP*s[38,44,45]. Conversely, there are instances of contrasting behavior observed in other species. For example, nine pairs of segmental duplicates and 10 pairs of tandem duplicates were detected in *Cymbidium ensifolium*[46]. The findings indicated that the expansion of the *bZIP* gene family varies across species, suggesting that distinct types of duplication events have contributed differently to the evolutionary history of these organisms. Additionally, we also detected the synteny relationship between *C. indicum* with other dicot and monocot plants, the results showed that *C. indicum* had strong synteny in *C. indicum* and soybean, followed by *V. vinifera* and *P. trichocarpa*. The Ka/Ks ratios of the 12 gene pairs revealed that strong purifying selection may be a significant factor in preserving the functional integrity of *CibZIP* proteins, similar to the observations in other plants. The results indicated that the *bZIP* family in *C. indicum* exhibited a significant level of evolutionary divergence and was characterized by a lack of conservation.

Potential function of *CibZIP* genes in response to Cd treatment

The *cis*-regulatory elements within promoter regions are crucial for the initiation of gene expression. Variations in these *cis*-regulatory elements among the promoter sequences of genes can lead to distinct expression patterns in *C. indicum*. According to the findings presented in Fig. 7, a total of 25 distinct types of *cis*-regulatory elements were identified. Among these, 11 were associated with hormone responsiveness, six were linked to developmental processes, and eight were involved in responses to environmental stress. In terms of *cis*-elements related to plant abiotic stress response, TC-rich repeats and W-box may play an important role in the defense of *bZIP* members against Cd stress processes. Also, environmental stress affects phytohormone signaling, and plants respond to Cd stress through hormone signaling and changes in their levels. Therefore, phytohormones can serve as important signaling molecules to regulate plant responses to Cd stress[47]. The results showed that most of the *cis*-regulatory elements in the studied *CibZIP* genes were associated with hormone responsiveness, suggesting that *CibZIPs* may also be involved in Cd stress response through hormone signaling. The distribution and frequency of these elements suggest that *CibZIP* genes are tightly regulated by complex networks involving multiple signaling pathways.

The KEGG pathway enrichment analysis performed on the *CibZIP* transcription factor family genes indicated that this gene family is involved in three specific pathways: Circadian rhythm, the MAPK signaling pathway, and plant hormone signal transduction. The effects of Cd on plant photosynthesis are mainly manifested in the interference of electron transfer during photosynthesis, and the inhibition of photosynthesis-related enzymes and the destruction of chloroplast structure[48]. Plant circadian rhythms are interconnected processes, and a significant number of genes identified through KEGG analysis exhibited altered expression in response to Cd stress (Figs 6 & 8). This alteration likely impacts the photosynthetic pathway, indicating that Cd stress may disrupt the photosynthetic response in marigold plants. Research has demonstrated that the antioxidant enzyme system in plants is modulated by secondary metabolites, including phytohormones and various signaling molecules, in response to Cd stress. Furthermore, this regulatory mechanism is closely associated with the activation of mitogen-activated protein kinases (MAPKs) during mitosis[49]. In experiments involving Cd-exposed seedlings of Arabidopsis thaliana, a notable elevation in the activity of *MPK3* and *MPK6* was recorded, with this enhancement becoming more pronounced as the concentration of Cd treatment increased[50]. Cd stress produces excessive reactive oxygen species leading to plant damage, inducing mitogen-activated protein kinase (MAPK) response and engaging it in defense against

Cd stress[51]. Additionally, previous findings indicated that the high Cd-tolerant early morning glory (*Poa pratensis* L.) influenced signaling pathways by modulating the expression of genes associated with growth hormones, ethylene, oleuropein steroids, and abscisic acid signaling. This regulation subsequently facilitated the coordinated expression of genes pertinent to Cd tolerance[52]. Consequently, in light of the aforementioned predicted outcomes, we hypothesize that the signaling mechanisms of phytohormones, in conjunction with the collaborative effects of various other transduction pathways, may contribute to the increased tolerance to Cd stress observed in *C. indicum*.

CibZIP29 is a positive regulator of Cd tolerance in *C. indicum*

CibZIP29 was chosen for functional analysis due to the observation that its expression is significantly elevated in response to Cd stress. In addition, we found significant differences in the expression of *CibZIP29* in roots and leaves, which was significantly elevated in roots based on the transcriptome results under Cd stress. This observation may be associated with the mechanism by which plants initially detect Cd stress in their roots, which are responsible for the uptake of Cd from the soil. Subsequently, the Cd is translocated to the stems and leaves[53]. Transgenic *Arabidopsis* plants that overexpress *CibZIP29* were developed utilizing the 35S promoter to investigate the biological role of *CibZIP29*. The extent of chlorosis observed in the leaves three days following the initiation of Cd stress treatment was greater in the wild-type (WT) plants compared to the plants that overexpress *CibZIP29*. We also found that the root length of transgenic plants was significantly higher than that of WT under Cd stress, suggesting that the response to Cd stress is anticipated to be more pronounced in the root tissues compared to other plant tissues. The characteristics of chlorophyll fluorescence, which provide insights into the photochemical activities of Photosystem II (PSII), serve as a valuable metric for assessing the impact of heavy metal stress, particularly Cd stress, on the photosynthetic machinery. The total chlorophyll content in the *CibZIP29*-overexpressing plants was significantly greater than that observed in the wild-type (WT) plants. Meanwhile, the Cd concentration in roots were significantly higher than that in leaves. However, the *CibZIP29*-overexpression lines showed significant decrease in roots compared to WT plants. The Cd transfer coefficient of overexpressing lines exhibited an upward trend, compared with WT. The findings suggest that *CibZIP29* may influence the absorption or translocation of Cd.

Conclusions

In our investigation, a total of 65 *bZIP* genes were identified within the genome of *C. indicum*, which were subsequently categorized into 12 distinct subgroups. The *CibZIP* genes exhibited a random distribution across the nine chromosomes. Notably, three pairs of *CibZIP* genes were determined to have arisen from tandem duplication, while nine pairs were found to originate from segmental duplication. The *cis*-regulatory elements present in the promoters of *CibZIP* genes were implicated in various biological processes, including cellular development, responses to phytohormones, environmental stress. Furthermore, the expression patterns of the *CibZIP* gene members varied significantly in different tissues of *C. indicum*, and most members were widely involved in response to Cd stress. Differential responses were observed among 15 *CibZIP* genes when subjected to Cd treatment using qRT-PCR. Additionally, the transgenic characterization of *CibZIP29* indicated that *CibZIP29* played a positive role in response to Cd tolerance. These findings help to deepen the understanding of the molecular mechanism of *CibZIP* gene function and provide a new entry point for the study of

resistance-related genes in Chrysanthemum as an important relative species of ornamental chrysanthemums for the discovery of genetic resources and molecular breeding of resistance in ornamental chrysanthemums. The improvement of stress tolerance is of great significance for the promotion of ornamental chrysanthemum varieties to a wider range of ecological environments.

Author contributions

The authors confirm contribution to the paper as follows: study conceptualization: Xu X, Chen S; methodology: Chen S; software: Zhang Y; validation: Zhang Y, Zhang K, Liu S; formal analysis: Chen S; investigation: Xu X; resources: Chen S; data curation: Yang Y; writing – original draft preparation: Chen S; writing – review and editing: Xu X; visualization: Xu X; supervision: Zhang K; project administration: Zhang Y; funding acquisition, Xu X. All authors reviewed the results and approved the final version of the manuscript.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Pérez-Chaca MV, Rodríguez-Serrano M, Molina AS, Pedranzani HE, Zirulnik F, et al. 2014. Cadmium induces two waves of reactive oxygen species in *Glycine max* (L.) roots. *Plant, Cell & Environment* 37:1672−87
2. Sanità di Toppi L, Gabbrielli R. 1999. Response to cadmium in higher plants. *Environmental and Experimental Botany* 41:105−30
3. Hall JL. 2002. Cellular mechanisms for heavy metal detoxification and tolerance. *Journal of Experimental Botany* 53:1−11
4. Singh K, Foley RC, Oñate-Sánchez L. 2002. Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology* 5:430−36
5. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, et al. 2010. MYB transcription factors in *Arabidopsis*. *Trends in Plant Science* 15:573−81
6. Suckow M, Schwamborn K, Kisters-Woike B, von Wilcken-Bergmann B, Müller-Hill B. 1994. Replacement of invariant bZip residues within the basic region of the yeast transcriptional activator GCN4 can change its DNA binding specificity. *Nucleic Acids Research* 22:4395−404
7. Vinson CR, Sigler PB, McKnight SL. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* 246:911−16
8. Yu Y, Qian Y, Jiang M, Xu J, Yang J, et al. 2020. Regulation mechanisms of plant basic leucine zippers to various abiotic stresses. *Frontiers in Plant Science* 11:1258
9. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, et al. 2002. bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* 7:106−11
10. Izawa T, Foster R, Chua NH. 1993. Plant bZIP protein DNA binding specificity. *Journal of Molecular Biology* 230:1131−44
11. Ma H, Liu C, Li Z, Ran Q, Xie G, et al. 2018. ZmbZIP4 contributes to stress resistance in maize by regulating ABA synthesis and root development. *Plant Physiology* 178:753−70
12. Song S, Wang G, Wu H, Fan X, Liang L, et al. 2020. OsMFT2 is involved in the regulation of ABA signaling-mediated seed germination through interacting with OsbZIP23/66/72 in rice. *The Plant Journal* 103:532−46
13. Utsugi S, Ashikawa I, Nakamura S, Shibasaka M. 2020. TaABI5, a wheat homolog of *Arabidopsis thaliana* ABA insensitive 5, controls seed germination. *Journal of Plant Research* 133:245−56
14. Liu C, Wu Y, Wang X. 2012. bZIP transcription factor *OsbZIP52/RISBZ5*: a potential negative regulator of cold and drought stress response in rice. *Planta* 235:1157−69
15. Bi C, Yu Y, Dong C, Yang Y, Zhai Y, et al. 2021. The bZIP transcription factor TabZIP15 improves salt stress tolerance in wheat. *Plant Biotechnology Journal* 19:209−11
16. Li Z, Fu D, Wang X, Zeng R, Zhang X, et al. 2022. The transcription factor *bZIP68* negatively regulates cold tolerance in maize. *The Plant Cell* 34:2833−51
17. Norén Lindbäck L, Ji Y, Cervela-Cardona L, Jin X, Pedmale UV, et al. 2023. An interplay between bZIP16, bZIP68, and GBF1 regulates nuclear photosynthetic genes during photomorphogenesis in Arabidopsis. *New Phytologist* 240:1082−96
18. Lu Z, Qiu W, Jin K, Yu M, Han X, et al. 2022. Identification and analysis of bZIP family genes in *Sedum plumbizincicola* and their potential roles in response to cadmium stress. *Frontiers in Plant Science* 13:859386
19. Chai M, Fan R, Huang Y, Jiang X, Wai MH, et al. 2022. *GmbZIP152*, a soybean bZIP transcription factor, confers multiple biotic and abiotic stress responses in plant. *International Journal of Molecular Sciences* 23:10935
20. Hou F, Liu K, Zhang N, Zou C, Yuan G, et al. 2022. Association mapping uncovers maize *ZmbZIP107* regulating root system architecture and lead absorption under lead stress. *Frontiers in Plant Science* 13:1015151
21. Wu B, Peng H, Sheng M, Luo H, Wang X, et al. 2021. Evaluation of phytoremediation potential of native dominant plants and spatial distribution of heavy metals in abandoned mining area in Southwest China. *Ecotoxicology and environmental safety* 220:112368
22. Yang W, Glover BJ, Rao GY, Yang J. 2006. Molecular evidence for multiple polyploidization and lineage recombination in the *Chrysanthemum indicum* polyploid complex (Asteraceae). *New Phytologist* 171:875−86
23. Liu Y, Chai M, Zhang M, He Q, Su Z, et al. 2020. Genome-wide analysis, characterization, and expression profile of the basic leucine zipper transcription factor family in pineapple. *International Journal of Genomics* 2020:3165958
24. Hou Z, Yang S, He W, Lu T, Feng X, et al. 2024. The haplotype-resolved genome of diploid *Chrysanthemum indicum* unveils new acacetin synthases genes and their evolutionary history. *The Plant Journal* 119:1336−52
25. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772−80
26. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530−34
27. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* 52:W78−W82
28. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, et al. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194−202
29. Zhu L, Guan Y, Liu Y, Zhang Z, Jaffar MA, et al. 2020. Regulation of flowering time in chrysanthemum by the R2R3 MYB transcription factor *CmMYB2* is associated with changes in gibberellin metabolism. *Horticulture Research* 7:96
30. Gao W, Meng Q, Wang X, Chen F, Zhou Y, et al. 2023. Overexpression of *CiMYC2* transcription factor from *Chrysanthemum indicum* var. *aromaticum* resulted in modifed trichome formation and terpenoid

biosynthesis in transgenic tobacco. *Journal of Plant Growth Regulation* 42:4161−75

31. Uraguchi S, Ohshiro Y, Otsuka Y, Wada E, Naruse F, et al. 2022. Phytochelatin-mediated metal detoxification pathway is crucial for an organomercurial phenylmercury tolerance in *Arabidopsis*. *Plant Molecular Biology* 10:563−77

32. Li GZ, Zheng YX, Liu HT, Liu J, Kang GZ. 2022. WRKY74 regulates cadmium tolerance through glutathione-dependent pathway in wheat. *Environmental Science and Pollution Research* 29:68191−201

33. Arnon DI. 1949. Copper enzymes in isolated chloroplasts. polyphenoloxidase in *Beta vulgaris*. *Plant Physiology* 24:1−15

34. Dong Q, Tian Y, Zhang X, Duan D, Zhang H, et al. 2024. Overexpression of the transcription factor *MdWRKY115* improves drought and osmotic stress tolerance by directly binding to the *MdRD22* promoter in apple. *Horticultural Plant Journal* 10:629−40

35. Gulzar F, Fu J, Zhu C, Yan J, Li X, et al. 2021. Maize WRKY transcription factor *ZmWRKY79* positively regulates drought tolerance through elevating ABA biosynthesis. *International Journal of Molecular Sciences* 22:10080

36. Jin D, Zhang Q, Liu Y, Liang M, Li A, et al. 2022. Overexpression of the maize phytochelatin synthase gene (*ZmPCS1*) enhances Cd tolerance in plants. *Acta Physiologiae Plantarum* 44:114

37. Zhu S, Shi W, Jie Y. 2021. Overexpression of *BnPCS1*, a novel phytochelatin synthase gene from ramie (*Boehmeria nivea*), enhanced Cd tolerance, accumulation, and translocation in *Arabidopsis thaliana*. *Frontiers in Plant Science* 12:639189

38. Tao YT, Chen LX, Jin J, Du ZK, Li JM. 2022. Genome-wide identification and analysis of *bZIP* gene family reveal their roles during development and drought stress in Wheel Wingnut (*Cyclocarya paliurus*). *BMC Genomics* 23:743

39. Zhang Y, Gao W, Li H, Wang Y, Li D, et al. 2020. Genome-wide analysis of the bZIP gene family in Chinese jujube (*Ziziphus jujuba Mill.*). *BMC Genomics* 21:483

40. Liu M, Wen Y, Sun W, Ma Z, Huang L, et al. 2019. Genome-wide identification, phylogeny, evolutionary expansion and expression analyses of *bZIP* transcription factor family in tartaty buckwheat. *BMC Genomics* 20:483

41. Zhang M, Liu Y, Shi H, Guo M, Chai M, et al. 2018. Evolutionary and expression analyses of soybean basic Leucine zipper transcription factor family. *BMC Genomics* 19:159

42. Zhao J, Guo R, Guo C, Hou H, Wang X, et al. 2016. Evolutionary and expression analyses of the apple basic leucine zipper transcription factor family. *Frontiers in Plant Science* 7:376

43. Ma BT, Wu GQ, Wei M. 2024. Roles of bZIP transcription factor in the response to stresses, and growth and development in plants. *Biotechnology Bulletin* 40:148−60

44. Duan L, Mo Z, Fan Y, Li K, Yang M, et al. 2022. Genome-wide identification and expression analysis of the bZIP transcription factor family genes in response to abiotic stress in *Nicotiana tabacum* L. *BMC Genomics* 23:318

45. Fan J, Chen N, Rao W, Ding W, Wang Y, et al. 2024. Genome-wide analysis of bZIP transcription factors and their expression patterns in response to methyl jasmonate and low-temperature stresses in *Platycodon grandiflorus*. *PeerJ* 12:e17371

46. Lai H, Wang M, Yan L, Feng C, Tian Y, et al. 2024. Genome-wide identification of bZIP transcription factors in *Cymbidium ensifolium* and analysis of their expression under low-temperature stress. *Plants* 13:219

47. Monteiro CC, Rolão MB, Franco MR, Peters LP, Cia MC, et al. 2012. Biochemical and histological characterization of tomato mutants. *Anais da Academia Brasileira de Ciencias* 84:573−85

48. Komárková M, Chromý J, Pokorná E, Soudek P, Máchová P. 2020. Physiological and transcriptomic response of grey poplar (*Populus × canescens* Aiton Sm.) to cadmium stress. *Plants* 9:1485

49. Quan M, Liu X, Xiao L, Chen P, Song F, et al. 2021. Transcriptome analysis and association mapping reveal the genetic regulatory network response to cadmium stress in *Populus tomentosa*. *Journal of Experimental Botany* 72:576−91

50. Agrawal GK, Rakwal R, Iwahashi H. 2002. Isolation of novel rice (*Oryza sativa* L.) multiple stress responsive MAP kinase gene, *OsMSRMK2*, whose mRNA accumulates rapidly in response to environmental cues. *Biochemical and Biophysical Research Communications* 294:1009−16

51. Genchi G, Sinicropi MS, Lauria G, Carocci A, Catalano A. 2020. The effects of cadmium toxicity. *International Journal of Environmental Research and Public Health* 17:3782

52. Xian J, Wang Y, Niu K, Ma H, Ma X. 2020. Transcriptional regulation and expression network responding to cadmium stress in a Cd-tolerant perennial grass *Poa Pratensis*. *Chemosphere* 250:126158

53. Pan C, Lu H, Yu J, Liu J, Liu Y, et al. 2019. Identification of cadmium-responsive *Kandelia obovata* SOD family genes and response to Cd toxicity. *Environmental and Experimental Botany* 162:230−38