# Genome-wide identification of *bHLH* gene family and its response to cadmium stress in *Populus × canescens*

Yuneng Yao[1, 2, 3#], Zhengquan He[1#], Xinmeng Li[1, 2, 3], Jing Xu[2, 3], Xiaojiao Han[2, 3], Hongwei Liang[1], Renying Zhuo[2, 3], Wenmin Qiu[2, 3*]

1 Key Laboratory of Three Gorges Regional Plant Genetic & Germplasm Enhancement (CTGU)/ Biotechnology Research Center, China Three Gorges University, Yichang, 443002, Hubei, China.

2 State Key Laboratory of Tree Genetic and Breeding, Chinese Academy of Forestry, Beijing, China.

3 Key Laboratory of Tree Breeding of Zhejiang Province, The Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou, 311400, P.R. China.

#These authors contributed equally to this work

*Correspondence and requests for materials should be addressed to Wenmin Qiu at the following address, phone number, and email address:

Wenmin Qiu

Address: Research Institute of Subtropical Forestry, Chinese Academy of Forestry.

Fuyang, Hangzhou 311400, China

Tel: +086-571-63311860

E-mail: qiuwm05@163.com

**Abstract**

The basic helix-loop-helix (bHLH) gene family is integral to various aspects of plant development and the orchestration of stress response. This study focuses on the bHLH genes within *Populus × canescens*, a poplar species noted for its significant tolerance to cadmium (Cd) stress. Through our comprehensive genomic analysis, we have identified and characterized 170 *bHLH* genes within *P. canescens* genome. These genes have been systematically classified into 22 distant subfamilies based on their evolutionary relationships. A notable conservation in gene structure and motif compositions were conserved across these subfamilies. Further analysis of the promoter regions of these genes revealed an abundance of essential cis-acting element, which are associated with plant hormonal regulation, development processes, and stress response pathway. Utilizing quantitative PCR (qPCR), we have documented the differential regulation of PcbHLHs in response to elevated Cd concentrations, with distinct expression patterns observed across various tissues. This study is poised to unravel the molecular mechanism underpinning Cd tolerance in *P. canescens*, offering valuable insights for the development of new cultivars with enhanced Cd accumulation capacity and tolerance. Such advancements are crucial for implementing effective phytoremediation strategies to mitigate soil pollution caused by Cd.

**INTRODUCTION**

The eukaryotic domain is replete with basic helix-loop-helix (bHLH) proteins, which constitute one of the largest transcription factor family in plants, second only to the MYB family(Ledent and Vervoort, 2001; Riechmann and Ratcliffe, 2000). The bHLH proteins are characterized by a conserved 60 amino acids domain, which is divisible into two distinct functional regions(Jones, 2004). The N-terminal basic region is critical for DNA binding(Atchley et al., 1999), while the C-terminal helix-loop-helix facilitates protein-protein interactions, essential for the formation of homodimeric or heterodimeric complex(Ferre-D'Amare et al., 1994; Murre et al., 1989).

In plants, bHLH transcription factors are implicated in a myriad of biological processes, including growth, development, and stress responses(Liu et al., 2014; Meraj et al., 2020). Notably, certain bHLH proteins have been identified as key mediators of heavy metal, particularly cadmium (Cd), tolerance(Wu et al., 2012; Xu et al., 2017; Yuan et al., 2008). *Populus × canescens*, a poplar species with remarkable Cd accumulation capacity, is of particular interest for its potential in

2

60 phytoremediation-a strategy for mitigating soil contamination using plants(He et al., 2011).

61 This study marks the first comprehensive characterization of the bHLH gene family in *P.*

62 *canescens*, revealing 170 genes distributed across the genome and classified into 22 subfamilies.

63 Our analysis encompassed gene structures, conserved motifs, DNA-binding capability and cis-

64 acting elements in promoters, alongside gene duplications, synteny, and phylogenetic relationships.

65 Furthermore, we conducted a detailed analysis of tissue-specific and Cd-induced expression

66 profiles of selected *PcbHLH* genes, laying the groundwork for future research into the functional

67 and regulatory mechanisms of these genes in response to Cd stress. This work not only contributes

68 to our understanding of Cd tolerance in *P. canescens* but also identifies potential candidate genes

69 for breeding new germplasm with enhanced Cd pollution resistance.

70

71 **MATERIALS AND METHODS**

72 **Identification of the basic/helix-loop-helix family genes in *P. canescens***

73 The genomic sequence data of *P. canescens* were sourced from the Aspen database

74 (https://www.aspendb.org/downloads and the annotation version was sPta717alba_v2). bHLH

75 transcription factors of *Arabidopsis thaliana* (AtbHLHs) were obtained from Lorenzo Carretero-

76 Paulet(Carretero-Paulet et al., 2010), and those of *Populus trichocarpa* (PtbHLHs) were retrieved

77 from NCBI's database. Hidden Markov Model (HMM) files for the bHLH domain (PF0010) were

78 acquired from interProScan (https://www.ebi.ac.uk/interpro/entry/pfam/PF00010/), which were

79 then utilized to identify bHLH proteins in *P. canescens* using the SPDE software(Xu et al., 2021).

80 BLASTp searches were also performing using AtbHLHs against the *P. canescens* amino acid

81 sequence data. To confirm the presence of the bHLH domain (E-value $< 1e^{-5}$), CD-search

82 (https://www.ncbi.nlm.nih.gov/) and SMART (http://smart.embl-heidelberg.de/) were employed to

83 analyze the identified PcbHLH sequences. The molecular weights (kDa) and isoelectric points (pI)

84 of the PcbHLHs were calculated using the SPDE software.

85 **Chromosomal locations of *PcbHLHs***

86 Chromosomal locations of the *PcbHLH* genes were determined using the *P. canescens* database

87 (https://www.aspendb.org/downloads), and a distribution map was created with TBtools(Chen et al.,

88 2023).

89 **Multiple sequence alignments of PcbHLHs**

3

**Deleted:** B

Multiple sequence alignments were conducted using ClustalX in MEGA7, followed by visualization with Jalview (version 1.8.3). The variable sequences at the N- and C-terminal regions were excised, preserving the conserved domains in the central region. Sequence logos for the bHLHs were created by submitting the multiple alignment sequences to a specialized online platform (https://weblogo.berkeley.edu/logo.cgi). Employing the criteria delineated by Massari and Murre (2000), we classified the PcbHLH proteins into two principal categories based on the sequence information within the N-terminal region of the bHLH domains: DNA binding and Non-DNA binding (containing fewer than four basic amino acids). The DNA binding bHLHs were further categorized into two groups: E-box binding and Non-E-box binding (based on the presence of only Glu12 or Arg15). Consequently, E-box binding were further subdivided into two subgroups: G-box binding (characterized by His/Lys8, Glu12 and Arg16) and Non-G-box binding (defined by the presence of only Glu12 and Arg16).

**Phylogenetic analysis of PcbHLHs**

To elucidate the evolutionary relationships between *P. canescens* and *Arabidopsis* bHLHs, we performed multiple sequences alignment of 170 complete PcbHLH sequences and 167 complete AtbHLH sequences using MEGA7. Subsequently, a Maximum Likelihood phylogenetic tree was constructed in MEGA7, with a bootstrap value set to 1,000 replicates for reliability. The phylogenetic tree was visualized using Itol (https://itol.embl.de/), which provided a clear depiction of the bHLHs' evolutionary history.

**Gene structure and protein motifs analysis**

The exon/intron organization and splicing phase of the predicted *PcbHLHs* were analyzed using the Gff3 annotation files from *P. canescens* genome. This data were then graphically represented using TBtools. To identify conserved motifs within the PcbHLH proteins, we employed MEME (https://meme-suite.org/meme/tools/meme, Version 5.5.4) to identify a maximum of ten motifs with an optimal width ranging from 10 to 100 amino acids. The resulting phylogenetic trees, gene structures, and conserved motifs were integrated and visualized within TBtools for comprehensive analysis.

**Cis-Acting element analysis in PcbHLH promoters**

Promoter regions of *PcbHLH* genes, defined as 2000bp upstream sequences, were analyzed for cis-acting elements using PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

121  TBtools facilitated the visualization of these elements, and a heat map was generated using Rstudio

122  (Version 4.2.2) to represents the distribution of various cis-acting elements, providing insight into

123  the regulatory mechanism of these genes.

124  **Gene duplication and synteny analysis of *PcbHLH* genes**

125  Gene duplication and synteny analysis provided insights into the evolutionary dynamics of *bHLH*

126  genes in poplar and other plants. This analysis was performed using the Multiple Collinearity Scan

127  toolkit (MCScanX), with results visualized in TBtools. Additionally, the nonsynonymous (ka) and

128  synonymous (ks) substitution rates of *bHLH* gene pairs were calculated, offering further evidence

129  of the evolutionary pressures acting on these genes.

130  **Plant growth conditions and and Cd treatment**

131  *P. canescens* seedlings, derived from micropropagation(Li et al., 2023), were cultured

132  hydroponically at 25 °C under a photoperiod of 16 h light and 8 h dark within an artificial climate

133  chamber. After acclimatization in half-strength Hoagland nutrient medium for one month, seedlings

134  were treated with 10 μM $CdCl_2$ to induce Cd stress for 168h. A concentration of 10 μM $CdCl_2$ was

135  sufficient to elicit a transcription factor response in *P. cansecens* without causing harm. Samples of

136  root, stems and leaves from individual *P. canescens* plants were collected at 0, 6, 12, 24, 48, 96, and

137  168 h post-treatment, with all samples collected simultaneously to minimize the impact of

138  diurnal rhythms. All samples were collected in triplicate to negate the impact of biological

139  variability.

140  **RNA extraction and qRT-PCR analysis**

141  Total RNA was extracted from root, stem, and leaf samples of *P. canescens* seedlings using the

142  RNAprep Pure Plant Plus Kit (TIANGEN, Beijing, China). First-strand cDNA synthesis was

143  performed using the PrimeScript™ RT Master Mix (TaKaRa, Dalian, China). Quantitative primers

144  were designed using TBtools and qPCR was conducted with the 2 × Q3 SYBR qPCR Master Mix

145  (TOLOBIO, Shanghai, China) on a 7300 Real-Time PCR System (Applied Biosystems, CA, United

146  States). Technical replicates were conducted three times for each sample. Relative expression levels

147  were quantified using the $2^{-\Delta\Delta CT}$ method with *EF1B* gene serving as a reference(Wildhagen et al.,

148  2010).

149

150  **RESULTS**

**Identification and classification of bHLH protein in *P. canescens***

The initial bioinformatic screening of *P. canescens* using HMMER with Hidden Markov Models and BLASTp yielded a total of 179 predicted bHLH proteins. Subsequent filtering, employing CD-Search and SMART to exclude proteins with incomplete domains, culminated in the identification of 170 putative bHLH proteins (Table S1). These proteins were designated as *PcbHLH1* to *PcbHLH170*, based on their chromosome locations. An analysis of their biochemical properties (Table S2) indicated a range of amino acid lengths from 90 for *PcbHLH43* and *PcbHLH154* to 740 for *PcbHLH46*. The molecular weights spanned from 10.29 kDa to 71.95 kDa, and the predicted isoelectric points ranged from 4.63 to 10.18. All proteins exhibited a grand average of hydropathy values below zero, indicative of their hydrophilic nature.

**Chromosomal localization of *PcbHLH* genes**

Distribution mapping revealed that 169 of the 170 *PcbHLH* genes are unevenly distributed across the *P. canescens* genome, with Chr2 harboring the highest number (17 genes). In contrast, only four *PcbHLH* genes were mapped to Chr3 and Chr7 (Figure 1). Notably, *PcbHLH170* remained unassigned to any chromosomes, which may be attributed to incomplete genome assembly.

**Multiple sequence alignments of PcbHLHs**

A comprehensive multiple sequence alignment of the 170 PcbHLH protein sequences was performed, revealing that the bHLH domain encompasses four conserved regions: a basic region, two helical regions, and a loop region. The basic region is characterized by specific residues, including His-5, Glu-9, Arg-10, Arg-12, and Arg-13, while the first helix region is defined by Ile-16, Leu-23, Leu-27, Val-28, and Pro-29. The loop region is predominantly composed of Lys-35 and Asp-41, and the second helix region comprises Ala-43, Leu-46, Glu-48, Ala-49, Ile-50, Tyr-52, and Leu-56. Notably, Figure 2 demonstrated the high conservation of Leu-23 across the 170 PcbHLH amino acid sequences, underscoring its pivotal role in facilitating dimerization among PcbHLH proteins.

The basic region of the PcbHLH proteins is crucial for DNA-binding. As per Massari's classification, these proteins are divided into two main groups based on sequence information within the basic region of the bHLH domains. The majority, comprising 165 PcbHLH proteins, were identified as DNA binders, while a smaller group of five was classified as non-DNA binders(Massari and Murre, 2000) (Table 1). The DNA-binding proteins are further classified into E-box binders

186 (including G-box binder) and non-E-box binders. Residues conservation suggested that 144

187 PcbHLHs are likely to bind E-boxs, with 109 of these identified as G-box binders, contingent on

188 the presence of His/Lys-5, Glu-9, and Arg-13 in the basic region.

189 **Phylogenetic analysis of *PcbHLH* genes**

190 To elucidate the evolutionary relationships between PcbHLHs and the AtbHLHs, a maximum-

191 likelihood phylogenetic tree was constructed. The analysis of clade support values and tree topology

192 led to the identification of 22 subfamilies. As shown in Figure 3, each subfamily includes

193 representatives from both *A. thaliana* and *P. canescens*, indicating a high degree of conservation in

194 the bHLH domains through their evolution. The largest subfamily, Clade III, contains 37 members,

195 whereas the smallest, Clade XIV, consists of only 5 members.

196 **Analysis of gene structure and conserved motif of bHLH family**

197 Motif analysis using MEME identified ten types of putative conserved protein motifs within the

198 PcbHLH family, furthering our understanding of their conservation and diversity. As depicted in

199 Figure 4, members within the same subfamily exhibit similar motif structure, suggesting shared

200 structural and functional characteristics. Gene structure analysis provided insights into the

201 evolutionary relationships among PcbHLH family members. Of the 170 family members, 11 were

202 found to lack introns and were clustered within two subfamilies. The remaining 159 members, which

203 contained introns, showed significant structural similarities, reinforcing the notion of close

204 evolutionary ties within their respective subfamilies.

205 **Analysis of cis-acting elements in PcbHLH promoters**

206 Cis-acting elements within the promoter regions are crucial for classifying subfamilies and

207 functional characterizing members of the PcbHLH family. A 2000 base pair (bp) upstream of the

208 transcriptional start sites were identified as the promoter regions for this analysis. Several

209 functionally significant cis-acting elements were identified and categorized into three groups: TC-

210 rich repeats (cis-acting element involved in defense and stress responsiveness), LTR (cis-acting

211 element involved in low-temperature responsiveness), MBS (MYB binding site involved in

212 drought-inducibility) belong to stress-responsive. GCN4_motif (cis-regulatory element involved in

213 endosperm expression), CAT-box (cis-acting regulatory element related to meristem expression),

214 O2-site (cis-acting regulatory element involved in zein metabolism regulation) belong to plant

215 development-related. AuxRR-core (cis-acting regulatory element involved in auxin responsiveness),

7

P-box (gibberellin-responsive element), CGTCA/TGACG-motif (cis-acting regulatory element involved in the MeJA-responsiveness), ABRE (cis-acting element involved in the abscisic acid responsiveness), SARE (cis-acting element involved in salicylic acid responsiveness) belong to phytohormone responsive. Figure 5 illustrates that the majority of the PcbHLH members contain phytohormone-related cis-acting elements, implying their potential roles in various abiotic stress responses.

**Gene duplication and synteny analysis of PcbHLHs**

To discern the primary evolutionary forces shaping the *PcbHLH* gene family, gene duplication events within *P. canescens* were analyzed using TBtools. Figure 6A displays 92 segmental duplication gene pairs and eight tandem duplication gene pair (*PcbHLH65/PcbHLH66*, *PcbHLH66/PcbHLH67*, *PcbHLH119/PcbHLH120*, *PcbHLH144/PcbHLH145*, *PcbHLH159/PcbHLH160*, *PcbHLH160/PcbHLH161*, *PcbHLH161/PcbHLH163*, *PcbHLH167/PcbHLH168*). These findings suggest that gene duplication has been a significant mechanism in the evolution of *PcbHLH* genes. The Ka/Ks ratio for these duplicated genes, as detailed in Table S3, was consistently below 0.8, indicating a predominant purifying selection acting on the *bHLH* genes.

**Comparative synteny analysis across species**

To further elucidate the phylogenetic mechanisms underlying the PcbHLH family, we constructed syntenic maps comparing *P. canescens* with *P. trichocarpa* and *Arabidopsis*. A total of 144 orthologous *bHLH* gene pairs were identified between *P. canescens* and P. *trichocarpa*, and 93 orthologs between *P. canescens* and *A. thaliana* (Figure 6B). The higher number of orthologous pairs between *P. canescens* and *P. trichocarpa* suggests a closer phylogenetic relationship compared to *A. thaliana*.

*PcbHLH* **genes expression patterns in response to Cd stress**

Prior research has established the role of certain clade IVc and Ib members as key regulators of Cd stress response in *Arabidopsis*(Hao et al., 2021). Leveraging these findings, we conducted a phylogenetic analysis of AtbHLHs and PcbHLHs to explore the functions of PcbHLH proteins. We meticulously selected 14 *PcbHLH* genes based on their functional homologs within the same subfamily for further analysis. Following Cd treatment, we observed transcriptional changes in these genes (Figure 7A). In roots, genes such as *PcbHLH148* and *PcbHLH98* exhibited upregulated

248 expression, while others like *PcbHLH19* and *PcbHLH50* showed downregulated expression. In

249 stems, *PcbHLH162* displayed decreased expression initially, with other genes upregulated up to 6h.

250 In leaves, all but *PcbHLH61* exhibited upregulated expression, with a notable positive regulatory

251 response observed for most *PcbHLH* genes, particularly in leaf tissues (Figure 7B); *PcbHLH96* was

252 an exception, showing high expression levels in roots.

253 **DISCUSSION**

254 Numerous plant species have been thoroughly investigated to characterize their bHLH families,

255 which has highlighted the bHLH transcription factor family as one of the largest in

256 eukaryotes(Ledent and Vervoort, 2001; Riechmann and Ratcliffe, 2000), and thus, it is of particular

257 significance. The proliferation of plant genomic data has been instrumental, offering essential

258 sequence resources that enable comprehensive identification of *bHLH* genes across plant species.

259 This study identified and characterized 170 *bHLH* genes in *P. canescens* (Figure 3). Surpassing the

260 number found in *A. thaliana* (167) and tomato (159), but fewer than those in rice (177) and *P.*

261 *trichocarpa* (183)(Carretero-Paulet et al., 2010). Relative to genome size, *P. canescens* possesses a

262 higher ratio of *PcbHLH* genes compared to tomato but a lower ratio than *Arabidopsis*. These results

263 indicate that the number of *bHLH* genes is variable across plant species and does not strictly

264 correlate with genome size. Through phylogenetic analysis, we categorized the 170 *PcbHLH* genes

265 into 22 distant subfamilies. The presence of both AtbHLHs and PcbHLHs in each subfamily

266 suggests that the *PcbHLH* genes diversified prior to the evolution of two species(Atchley and Fitch,

267 1997). In contrast, animal genomes harbor only six subfamilies. But in plants, the *bHLH* gene family

268 was divided into 24 subfamilies in tomato(Wang et al., 2015), 17 in *Ginkgo biloba*(Zhou et al.,

269 2020), 20 in *Camellia sinensis*(Cui et al., 2018), 21 in *A. thaliana*(Toledo-Ortiz et al., 2003). This

270 highlights a significant divergence in the classification of the bHLH gene family between plants and

271 animals.

272 Analysis of the conserved protein motifs and gene structures, as illustrated in Figure 4, reveals

273 that subfamily members likely share a common evolutionary origin and are involved in analogous

274 physiological processes(Ke et al., 2020). Our study identified a common conserved motif, motif 2,

275 across 170 *PcbHLH* genes. This suggests that motif 2 could represent the consensus motif within

276 the bHLH domain. Furthermore, alterations in gene structure are identified as key factors

277 contributing to functional diversity among genes. A total of 11 *PcbHLH* genes lack introns, which

278  may be indicative of their evolution within *P. canescens* in response to environmental selective

279  pressures(Lin et al., 2006; Yang et al., 2013). This observation suggests a potential link between

280  intron loss and adaptive evolution in this species. These variations are driven by three principal

281  mechanisms: exon/intron gain or loss, exonization or pseudo-exonization, and insertion or deletion.

282  Additionally, a striking uniformity in the number and arrangement of exons/introns was observed

283  within subfamily members. Notably, prior research has documented a range of 0 to 10 exons/introns

284  in sesame *bHLH* genes(Kazemitabar et al., 2020). Similarly, *Andrographis paniculata* presents a

285  variation of 0 to 14 exons/introns(Xu et al., 2022), and *P. canescens* exhibits a similar range from 0

286  to 14. These observatiobs are consistent with the aforementioned results.

287      The expansion of the *PcbHLH* gene family in *P. canescen*s is likely attributed to various gene

288  duplication events, including tandem, fragment, whole-genome duplication, and

289  transposition(Flagel and Wendel, 2009; Zhang, 2003). Tandem duplication typically results in two

290  or more genes on the same chromosome, whereas segmental duplication occurs across different

291  chromosomes(Schlueter et al., 2007). Segmental and tandem duplications are the primary drivers of

292  the plant gene family expansion throughout evolution. *P. canescens* exhibited eight tandem repeats

293  and 92 segmental duplications (Figure 6A), indicating that these duplications have significantly

294  contributed to the gene family's expansion. Comparative syntenic maps between *P. canescens* and

295  both *P. trichocarpa* and *A. thaliana* further elucidate the origin and evolution of the *PcbHLH* gene

296  family (Figure 6B). A total of 144 and 93 syntenic gene pairs were identified between *P. canescens*

297  and *P. trichocarpa* and *A. thaliana*, respectively, suggesting a closer phylogenetic relationship with

298  *P. trichocarpa*. The Ka/Ks ratio analysis (Table S3) indicates that *PcbHLH* gene pairs have

299  predominantly undergone purifying selection. The basic region's key residues are critical for

300  distinguishing variations in the hexanucleotide core sequence at the promoters of target genes,

301  enabling the classification of plant bHLHs into distinct DNA-binding categories. As per the criteria

302  by Massari and Murre (2000), the PcbHLH proteins are categorized into several categories (Table

303  1) with G-box binding proteins constituting the majority (64%). In contrast to *A. thaliana*, which

304  has 18% Non-DNA binding proteins(Toledo-Ortiz et al., 2003). PcbHLHs exhibit a lower

305  percentage (3%), although the DNA-binding activity of these non-DNA binding sequences warrants

306  further investigation.

307      Cadmium, a toxic heavy metal, poses a significant threat to human health and living organisms.

308    Prolonged exposure to Cd can lead to severe health issues, such as kidney disorders, neurotoxicity,

309    and osteoporosis(Jarup and Akesson, 2009; Satarug et al., 2010). The accumulation of heavy metals

310    in the human body through the consumption of contaminated crops is a concerning issue(Shimbo et

311    al., 2001; Zahir et al., 2005). Phytoremediation, an emerging technology, utilizes hyperaccumulator

312    plants to remediate contaminated soils. While most hyperaccumulators are herbaceous with limited

313    biomass, *P. canescens*, a fast-growing woody species with an extensive root system, is deemed

314    suitable for this purpose. The role of bHLH proteins in plants' response to heavy metal stress is well-

315    documented. For example, the heterologous expression of the soybean GmORG3 gene, a member

316    of the bHLH family, has been shown to enhance Cd tolerance in yeast(Xu et al., 2017). In

317    *Arabidopsis*, *bHLH* genes such as *FIT*/*bHLH3*8 and *FIT*/*bHLH39* have been identified as key

318    regulators of the Cd stress response(Wu et al., 2012; Yuan et al., 2008). Our phylogenetic analysis

319    revealed that PcbHLH proteins from specific clades are closely related to known heavy metal stress-

320    responsive AtbHLH proteins. We selected 14 *PcbHLH* genes from Clade XV (FIT/bHLH38 and

321    FIT/bHLH39), Clade XVI, Clade XIV and Clade XIII (neighboring subfamily) for further study and

322    used qPCR to assess their transcript levels under Cd stress. The results from the qPCR experiments

323    revealed a diverse range of expression patterns among the *PcbHLH* genes (Figure 7B), highlighting

324    their potential roles in the regulatory mechanisms underpinning cadmium stress responses in *P.*

325    *canescens*. Some *PcbHLH* genes exhibited increased expression levels under cadmium stress,

326    suggesting their possible involvement in the direct response to heavy metal toxicity or in the

327    activation of detoxification processes. Conversely, the downregulation of other *PcbHLH* genes may

328    indicate their participation in maintaining cellular homeostasis or in the adaptation mechanisms that

329    allow the plant to tolerate or accumulate cadmium. The variation in the expression profiles of the

330    *PcbHLH* genes across different tissues and time points further underscores the complexity of the

331    plant's response to cadmium. This temporal and spatial regulation of gene expression could be a

332    strategic adaptation by the plant to optimize its resource allocation and stress response. The qPCR

333    data, therefore, not only contribute to our understanding of the molecular basis of heavy metal stress

334    in plants but also pave the way for future research aimed at identifying key regulators and potential

335    targets for the development of plants with improved phytoremediation capabilities.

336    To examine the interact between transcription factors and cis-acting elements in regulating the

337    expression of downstream genes, we analyzed the cis-elements from the transcriptional start site

11

extending 2000 bps upstream. The analysis of cis-acting element indicated that PcbHLH family genes participate in numerous physiological processes, including plant growth and development, hormone responses, and stress responses. Furthermore, a multitude of stress and hormone response-related elements are prevalent in the promoter regions of *PcbHLH* gene (Figure 5), underscoring the crucial role of *PcbHLHs* in the mediation of plant response to abiotic stress(Yamaguchi-Shinozaki and Shinozaki, 2005). Prior research has established the involvement of these elements in modulating plant responses to various abiotic stressors, including drought, salinity, and temperature fluctuations(Saidi and Hajibarat, 2019). For instance, both ABA-dependent and ABA-independent regulatory mechanisms contribute to stress-responsive gene expression(Shinozaki et al., 2003; Thomashow, 1999; Zhu, 2002). The cis-acting element analysis of the PcbHLH promoters reveal a wide range of physiological processes and stress responses, highlighting the pivotal role of PcbHLHs in mediating abiotic stress in plants.

The bHLH family, one of the largest transcription factor families in plants, remains poorly understood, with many members' functions yet to be explored. The comprehensive analysis of *PcbHLH* genes presented herein aims to enhance our fundamental knowledge and provide a theoretical foundation for developing new germplasms resistant to Cd pollution. The ultimate goal is to cultivate *P. canescens* materials with high Cd concentration and tolerance, offering a potential solution for soil remediation.

**References**

Atchley, W. R., Fitch, W. M., 1997, A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci U S A, 94: 5172-5176.

Atchley, W. R., Terhalle, W., Dress, A., 1999, Positional dependence, cliques, and predictive motifs in the bHLH protein domain. J Mol Evol, 48: 501-516.

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martinez-Garcia, J. F., Bilbao-Castro, J. R., Robertson, D. L., 2010, Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. Plant Physiol, 153: 1398-412.

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y., Feng, J., Chen, H., He, Y., Xia, R., 2023, TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant, 16: 1733-1742.

Cui, X., Wang, Y. X., Liu, Z. W., Wang, W. L., Li, H., Zhuang, J., 2018, Transcriptome-wide identification and expression profile analysis of the bHLH family genes in Camellia sinensis. Funct Integr Genomics, 18: 489-503.

Ferre-D'Amare, A. R., Pognonec, P., Roeder, R. G., Burley, S. K., 1994, Structure and function of the

372          b/HLH/Z domain of USF. EMBO J, 13: 180-189.

373    Flagel, L. E., Wendel, J. F., 2009, Gene duplication and evolutionary novelty in plants. New Phytol, 183:
374          557-564.

375    Hao, Y., Zong, X., Ren, P., Qian, Y., Fu, A., 2021, Basic Helix-Loop-Helix (bHLH) Transcription Factors
376          Regulate a Wide Range of Functions in Arabidopsis. International Journal of Molecular Sciences,
377          22: 7152.

378    He, J., Qin, J., Long, L., Ma, Y., Li, H., Li, K., Jiang, X., Liu, T., Polle, A., Liang, Z., Luo, Z. B., 2011, Net
379          cadmium flux and accumulation reveal tissue-specific oxidative stress and detoxification in
380          Populus x canescens. Physiol Plant, 143: 50-63.

381    Jarup, L., Akesson, A., 2009, Current status of cadmium as an environmental health problem. Toxicol
382          Appl Pharmacol, 238: 201-208.

383    Jones, S., 2004, An overview of the basic helix-loop-helix proteins. Genome Biol, 5: 226.

384    Kazemitabar, S. K., Faraji, S., Najafi-Zarrini, H., 2020, Identification and in silico evaluation of bHLH genes
385          in the Sesamum indicum genome: Growth regulation and stress dealing specially through the
386          metal ions homeostasis and flavonoid biosynthesis - ScienceDirect. Gene Reports, 19.

387    Ke, Y. Z., Wu, Y. W., Zhou, H. J., Chen, P., Wang, M. M., Liu, M. M., Li, P. F., Yang, J., Li, J. N., Du, H., 2020,
388          Genome-wide survey of the bHLH super gene family in Brassica napus. BMC Plant Biol, 20: 115.

389    Ledent, V., Vervoort, M., 2001, The basic helix-loop-helix protein family: comparative genomics and
390          phylogenetic analysis. Genome Res, 11: 754-770.

391    Li, S., Zhuo, R., Yu, M., Lin, X., Xu, J., Qiu, W., Li, H., Han, X., 2023, A novel gene SpCTP3 from the
392          hyperaccumulator Sedum plumbizincicola redistributes cadmium and increases its
393          accumulation in transgenic Populu × canescens. Frontiers in Plant Science, 14.

394    Lin, H., Zhu, W., Silva, J. C., Gu, X., Buell, C. R., 2006, Intron gain and loss in segmentally duplicated genes
395          in rice. Genome Biol, 7: R41.

396    Liu, JH, Peng, Dai, WS, 2014, Critical cis-Acting Elements and Interacting Transcription Factors: Key
397          Players Associated with Abiotic Stress Responses in Plants. PLANT MOL BIOL REP, 2014,32(2):
398          303-317.

399    Massari, M. E., Murre, C., 2000, Helix-loop-helix proteins: regulators of transcription in eucaryotic
400          organisms. Mol Cell Biol, 20: 429-440.

401    Meraj, T. A., Fu, J., Raza, M. A., Zhu, C., Shen, Q., Xu, D., Wang, Q., 2020, Transcriptional Factors Regulate
402          Plant Stress Responses through Mediating Secondary Metabolism. Genes, 11: 346.

403    Murre, C., McCaw, P. S., Baltimore, D., 1989, A new DNA binding and dimerization motif in
404          immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell, 56: 777-783.

405    Riechmann, J. L., Ratcliffe, O. J., 2000, A genomic perspective on plant transcription factors. Curr Opin
406          Plant Biol, 3: 423-434.

407    Saidi, A., Hajibarat, Z., 2019, Characterization of cis-elements in hormonal stress-responsive genes in
408          Oryza sativa. Asia-Pacific Journal of Molecular Biology and Biotechnology: 95-102.

409    Satarug, S., Garrett, S. H., Sens, M. A., Sens, D. A., 2010, Cadmium, environmental exposure, and health
410          outcomes. Environ Health Perspect, 118: 182-190.

411    Schlueter, J. A., Lin, J. Y., Schlueter, S. D., Vasylenko-Sanders, I. F., Deshpande, S., Yi, J., O'Bleness, M.,
412          Roe, B. A., Nelson, R. T., Scheffler, B. E., Jackson, S. A., Shoemaker, R. C., 2007, Gene duplication
413          and paleopolyploidy in soybean and the implications for whole genome sequencing. BMC
414          Genomics, 8: 330.

415    Shimbo, S., Zhang, Z. W., Watanabe, T., Nakatsuka, H., Matsuda-Inoguchi, N., Higashikawa, K., Ikeda, M.,

416      2001, Cadmium and lead contents in rice and other cereal products in Japan in 1998-2000. Sci
417      Total Environ, 281: 165-175.

418 Shinozaki, K., Yamaguchi-Shinozaki, K., Seki, M., 2003, Regulatory network of gene expression in the
419      drought and cold stress responses. Curr Opin Plant Biol, 6: 410-417.

420 Thomashow, M. F., 1999, PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory
421      Mechanisms. Annu Rev Plant Physiol Plant Mol Biol, 50: 571-599.

422 Toledo-Ortiz, G., Huq, E., Quail, P. H., 2003, The Arabidopsis basic/helix-loop-helix transcription factor
423      family. Plant Cell, 15: 1749-1770.

424 Wang, J., Hu, Z., Zhao, T., Yang, Y., Chen, T., Yang, M., Yu, W., Zhang, B., 2015, Genome-wide analysis of
425      bHLH transcription factor and involvement in the infection by yellow leaf curl virus in tomato
426      (Solanum lycopersicum). BMC Genomics, 16: 39.

427 Wildhagen, Durr, Ehlting, PHYSIOL, R. J. T., 2010, Seasonal nitrogen cycling in the bark of field-grown
428      Grey poplar is correlated with meteorological factors and gene expression of bark storage
429      proteins. 2010,30(9): 1096-1110.

430 Wu, H., Chen, C., Du, J., Liu, H., Cui, Y., Zhang, Y., He, Y., Wang, Y., Chu, C., Feng, Z., Li, J., Ling, H. Q., 2012,
431      Co-overexpression FIT with AtbHLH38 or AtbHLH39 in Arabidopsis-enhanced cadmium
432      tolerance via increased cadmium sequestration in roots and improved iron homeostasis of
433      shoots. Plant Physiol, 158: 790-800.

434 Xu, D., Lu, Z., Jin, K., Qiu, W., Qiao, G., Han, X., Zhuo, R., 2021, SPDE: a multi-functional software for
435      sequence processing and data extraction. Bioinformatics, 37: 3686-3687.

436 Xu, J., Xu, H., Zhao, H., Liu, H., Xu, L., Liang, Z., 2022, Genome-wide investigation of bHLH genes and
437      expression analysis under salt and hormonal treatments in Andrographis paniculata. Industrial
438      Crops and Products, 183: 114928.

439 Xu, Z., Liu, X., He, X., Xu, L., Huang, Y., Shao, H., Zhang, D., Tang, B., Ma, H., 2017, The Soybean Basic
440      Helix-Loop-Helix Transcription Factor ORG3-Like Enhances Cadmium Tolerance via Increased
441      Iron and Reduced Cadmium Uptake and Transport from Roots to Shoots. Front Plant Sci, 8:
442      1098.

443 Yamaguchi-Shinozaki, K., Shinozaki, K., 2005, Organization of cis-acting regulatory elements in osmotic-
444      and cold-stress-responsive promoters. Trends Plant Sci, 10: 88-94.

445 Yang, Y. F., Zhu, T., Niu, D. K., 2013, Association of intron loss with high mutation rate in Arabidopsis:
446      implications for genome size evolution. Genome Biol Evol, 5: 723-733.

447 Yuan, Y., Wu, H., Wang, N., Li, J., Zhao, W., Du, J., Wang, D., Ling, H. Q., 2008, FIT interacts with AtbHLH38
448      and AtbHLH39 in regulating iron uptake gene expression for iron homeostasis in Arabidopsis.
449      Cell Res, 18: 385-397.

450 Zahir, F., Rizwi, S. J., Haq, S. K., Khan, R. H., 2005, Low dose mercury toxicity and human health. Environ
451      Toxicol Pharmacol, 20: 351-360.

452 Zhang, J., 2003, Evolution by gene duplication: an update. Trends in Ecology Evolution, 18: 292-298.

453 Zhou, X., Liao, Y., Kim, S. U., Chen, Z., Xu, F., 2020, Genome-wide identification and characterization of
454      bHLH family genes from Ginkgo biloba. Scientific Reports, 10.

455 Zhu, J. K., 2002, Salt and drought stress signal transduction in plants. Annu Rev Plant Biol, 53: 247-273.

456