CCCH Zinc Finger Genes in Barley: Genome-Wide Identification, Evolution, Expression and Haplotype Analysis

Qi Ai
North West Agriculture and Forestry University
Wenqiu Pan
North West Agriculture and Forestry University
Yan Zeng
Jiangxi Agricultural University
Yihan Li
Jiangxi Agricultural University
Licao Cui (cuilicao@jxau.edu.cn)
Chinese Academy of Agricultural Sciences

Research Article

Keywords: Barley, CCCH gene family, Expression pattern, Genetic variation, Haplotype analysis

Posted Date: November 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1114922/v1

Version of Record: A version of this preprint was published at BMC Plant Biology on March 15th, 2022. See the published version at https://doi.org/10.1186/s12870-022-03500-4.

Abstract

Background: CCCH transcription factors are important zinc finger transcription factors involved in the response to biotic and abiotic stress and physiological and developmental processes. Barley (*Hordeum vulgare*) is an agriculturally important cereal crop with multiple uses, such as brewing production, animal feed, and human food. The identification and assessment of new functional genes are important for the molecular breeding of barley.

Results: In this study, a total of 35 protein-encoding *CCCH* genes unevenly dispersed on seven different chromosomes were identified in barley. Phylogenetic analysis categorized the barley *CCCH* genes (*HvC3H*s) into seven subfamilies according to their distinct features, and this classification was supported by intron–exon structure and conserved motif analysis. Despite the large genome size of barley, the lower number of *CCCH* genes in barley might be attributed to the low frequency of segmental and tandem duplication events. Furthermore, the *HvC3H* genes displayed distinct expression profiles for different developmental processes and in response to various types of stresses. The expression of *HvC3H9* was significantly induced by multiple types of abiotic stress and/or phytohormone treatment, which might make it an excellent target for the molecular breeding of barley. Genetic variation of *HvC3H*s was characterized using publicly available exome-capture sequencing datasets. Clear genetic divergence was observed between wild and landrace barley populations in *HvC3H* genes. For most *HvC3H*s, nucleotide diversity and the number of haplotype polymorphisms decreased during barley domestication.

Conclusion: Overall, our study provides a comprehensive characterization of barley *CCCH* transcription factors, their diversity, and their biological functions.

Background

Transcription factors (TFs), also known as trans-acting factors, are important regulatory proteins that activate or inhibit specific target genes by interacting with *cis*-regulatory elements [1]. They play a role in multiple biological processes, such as growth, development, signal transduction, cellular morphogenesis, and responses to environmental stress [2]. Zinc finger TFs are some of the most abundant TFs in plants and play key roles in regulating transcription and various biological functions [3]. They are characterized by the typical zinc finger motif and a compact three-dimensional finger-type structure formed by cysteine and/or histidine residues coordinated with a few zinc atoms, which are critical for their specific roles in binding to DNA, RNA, and proteins [4]. Zinc finger proteins can be further categorized into at least 14 families based on their structural and functional characteristics, such as DOF, ERF, LIM, ring-finger, and WRKY families [5–9].

Most of the zinc finger families are protein- or DNA-binding proteins; a recently identified group of zinc finger proteins referred to as the *CCCH* gene family exhibits RNA-binding and processing activity through their specific motifs in both animals and plants [10, 11]. The CCCH proteins typically contain 1–6 CCCH-type motifs characterized by three cysteine residues and one histidine residue. According to the number of amino acid spacers between cysteine and histidine residues, the consensus sequence of the CCCH motif was originally defined as $C–X_{6-14}–C–X_{4-5}–C–X_3–H$ and later re-defined as $C–X_{4-17}–C–X_{4-6}–C–X_3–H$ (C stands for cysteine, H for histidine, and X for any amino acid) [11–13]. The consensus sequence $C\text{-}X_{7-8}\text{-}C\text{-}X_5\text{-}C\text{-}X_3\text{-}H$ is the most abundant motif of CCCH proteins in *Arabidopsis* and rice, suggesting that other motifs might be derived from these motifs [11].

An increasing number of studies have shown that CCCH-type zinc finger proteins play a role in cell fate specification and developmental processes in plants. For example, *AtKHZ1* and *AtKHZ2* are required for flowering and senescence in *Arabidopsis* [14]. *AtC3H59/ZFWD3* plays an essential role in seedling development, seed germination, and development by interacting with the *PPPDE* gene family protein Desil [15]. In rice, *OsDOS* and *OsTZF1* can act as repressors of leaf senescence [16, 17]. *OsGZF1* affects glutelin accumulation during seed development [18]. *GmZF351* and *GmZF392* in soybean are involved in seed lipid accumulation [19, 20]. Functional studies have also shown that some CCCH proteins play a role in regulating plant growth and development through hormone signal transduction pathways. Specifically, *AtTZF4/5/6* can act as negative regulators of light and gibberellins (GA) and positive regulators of abscisic acid (ABA)-mediated regulation of seed development, dormancy, and germination [21]. The CCCH-type zinc finger gene *OsLIC* is involved in the biosynthesis and/or signal transduction of

brassinosteroid, which affects the architecture of rice plants [22]. In switchgrass, *PvC3H69* can act as a negative regulator of leaf senescence by repressing ABA synthesis and ABA signaling pathways [23].

Several CCCH genes are implicated in the response to biotic and abiotic stress in plants. For example, *OsC3H10*, *OsC3H47*, and *OsTZF5* are involved in the regulation of tolerance to drought stress in rice [24–26]. Overexpression of *GhTZF1* significantly increases drought tolerance and delays drought-induced leaf senescence by negatively regulating the production and accumulation of reactive oxygen species (ROS) in transgenic *Arabidopsis* [27]. *Arabidopsis AtZFP1* has been reported to confer salt tolerance by limiting oxidative and osmotic stress and maintaining an ionic balance [28]. Another non-tandem CCCH-type gene in *Arabidopsis*, *AtC3H17*, has pleiotropic effects in the response to salt stress via the ABA-dependent signaling pathway [29]. *PvC3H72* was the first *CCCH* family gene identified to be involved in plant chilling and freezing tolerance. Overexpression of *PvC3H72* improves cold tolerance in switchgrass, possibly via the ABA-mediated pathway [30]. More recently, *DgC3H1* has been shown to confer cold tolerance in *Chrysanthemum* plants by regulating the osmotic and ROS system, as well as the expression of genes associated with the response to cold stress [31]. In addition, CCCH proteins have been shown to be involved in other adaptive processes, such as bacterial blight disease resistance [32], zinc homeostasis [33], and hydrogen peroxide [17] and oxidative stress [34]. These findings indicate that CCCH zinc finger proteins have various biological functions in model plants. However, few studies have evaluated the roles of *CCCH* gene family members in other plants, especially crop plants.

Genome assemblies have been used for the global identification and characterization of genes in numerous plant species. Genome-wide identification and characterization of *CCCH* genes have been carried out in *Arabidopsis*, rice [11], maize [13], poplar [35], tomato [36], *Medicago truncatula* [37], grape [38], citrus [39], switchgrass [40], *Brassica rapa* [41], and recently in the common bean [42], moso bamboo [43], wheat [44], and soybean [45]. However, *CCCH* genes have not yet been identified in barley (*Hordeum vulgare*), and their biological functions and evolutionary history remain poorly understood. Barley is one of the world's earliest domesticated crops and ranks fourth among all cereal crops in both area and tonnage harvested. Barley is more adaptable to harsh and marginal environments than its wild relative wheat; it is thus widely used for animal feed and beer materials and provides a major source of calories in isolated regions [46]. The draft sequence and reference sequence V1 (Morex V1) [47] and V2 (Morex V2) [48] assemblies of the barley cultivar Morex have been released by the International Barley Sequencing Consortium (IBSC) [46]. The improved barley genome provides valuable information for the isolation, genetic cloning, and functional characterization of unknown genes. The aim of this study was to genomically identify and characterize barley *CCCH* genes (*HvC3H*s). The phylogenetic relationships, distribution of motifs, intron-exon organization, and gene duplication events were comprehensively analyzed. We also conducted RNA-seq and qRT-PCR analysis to determine the possible function of *HvC3H* genes. Finally, genomic variation, genetic diversity, and selection on these genes during barley domestication were also investigated using barley resequencing data (including wild and landrace barley accessions). Our preliminary analysis provides new insight into the evolutionary history of *CCCH* genes and will aid future efforts to functionally characterize and genetically improve barley.

Methods

Identification of CCCH Proteins in Barley

The genomic proteins of barley Morex V2 were downloaded from the IPK database (https://doi.org/10.5447/ipk/2019/8). The CCCH protein sequences of *Arabidopsis* and rice were used as queries to search against the barley proteins with Basic Local Alignment Search Tool (BLAST) software. The Hidden Markov Model (HMM) of CCCH conserved domain (PF00642) was used as a query to search against the barley proteins by HMMER v3.0 with default parameters. The candidate CCCH proteins were further verified by Simple Modular Architecture Research Tool (SMART) (https://www.omicsclass.com/article/681), National Center for Biotechnology Information - Conserved Domains Database (NCBI-CDD) (https://www.omicsclass.com/article/310) and PFAM (http://pfam.xfam.org/) online databases. Putative proteins without CCCH domain were removed. A BLASTN search against barley expressed sequence tags (ESTs) was conducted to detect the existence of CCCH proteins. The computational physical and chemical properties of CCCH family members, including molecular weight (MW), theoretical isoelectric point (pI), instability index (II), and grand average of hydropathicity (GRAVY) were evaluated

using the online tool ExPASy (http://web.expasy.org/protparam/). The subcellular location was predicted using the cello software (http://cello.life.nctu.edu.tw/).

Phylogeny, Gene Structure and Conserved Motif Analysis

The ClustalX v2.1 software was used to perform multiple alignments using the full-length CCCH protein sequences with default parameters. The phylogenetic analysis was carried out based on Neighbor-Joining (NJ) method with bootstrap value of 1000 replications using MEGA X software [99]. The intron-exon organization of Hv*C3H* genes was generated by Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/) based on the gene annotation Gene Transfer Format (GTF) file [100]. The Multiple Expectation Maximization for Motif Elicitation (MEME) (http://meme-suite.org/) was employed to analyze the conserved protein motifs with the following parameters: the number of all primitives was set to 10, the minimum and maximum width were set to 6 and 50, respectively [101]. The upstream 1.5kb genomic sequences of *HvC3H* genes were extracted and then submitted to the PlantCARE online database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to detect the potential *cis*-acting regulatory elements in the promoter region.

Chromosome Localization and Gene Synteny Analysis

The chromosomal locations of *HvC3H* genes were obtained from IPK database (https://doi.org/10.5447/ipk/2019/8), and the chromosome maps were visualized using MapChart v2.32. The MCScanX software was employed to analyze the synteny relationships of *HvC3H*s in rice (*Oryza sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*) soybean (*Glycine max*), tomato (*Solanum lycopersicum*) and *Brassica rapa*[102]. The gene duplication events of *HvC3H*s were identified according to the genomic comparison. Tandem duplicated genes were defined based on the following criteria (1) located within the same chromosome; (2) <1 intervening gene [52]. The syntenic and duplicated gene pairs were visualized by the Circos v0.67 tool. The non-synonymous substitution (Ka) / synonymous substitution (Ks) ratio was calculated to estimate genes evolutionary rate using the PAL2NAL online tools (http://www.bork.embl.de/pal2nal/) [103]. Ka/Ks >1, =1 and <1 represent positive, neutral and purifying selection, respectively. The divergence time of syntenic and duplicated gene pairs was calculated based on the formula T= (Ks / 2λ) × 10^{-6} million years ago (MYA) (λ = 6.5 × 10−9) [104]. The BLAST and orthoVeen2 software employed to analyze the homologous genes between barley and other related species [105].

Expression Patterns of *HvC3H* Gene Members

In order to explore the expression patterns of *HvC3H*s, the publicly available 148 RNA-seq samples were downloaded from the NCBI Sequence Reading Archive (SRA) database, including different developmental stages and tissues and different biotic and abiotic stresses. The accession number and sample information of RNA-seq were listed in Supplementary Table S1.
The fragments per kilobase per million (FPKM) value was calculated by the HISAT2 v2.1.0 and StringTie v1.3.5 pipeline. The R package pheatmap was used to visualize the expression profiles using the log2 transformed FPKM value.

Plant Material, Stress Treatment, RNA Extraction and qRT-PCR Analysis

Seeds of barley Morex were obtained from the College of Agronomy, Northwest A&F University, and were used as the experimental material. the barley seeds were hydroponically grown in growth chamber under controlled conditions (23±1 °C, 16-h light/8-h dark cycle). The seedlings were processed for stress treatments at the three-leaf stage. For salt, drought, cold and ABA treatment, the plants were incubated in 150 mM NaCl, 19.2% (w/v) PEG-6000, 4°C and 100 μM ABA for 0, 1, 3, 6, 12, and 24 hours, respectively. Seedlings without any treatment at the same time point were considered as the control. Leaves of all these samples were randomly collected with three biological replications. The collected samples were immediately frozen in liquid nitrogen and stored at 80°C for RNA extraction.

To further reveal the possible functions of *HvC3H*s, a total of 20 *HvC3H*s were randomly selected to investigate their expression profile under various stresses by quantitative real-time PCR (qRT-PCR) analysis. *HvACTIN* (HORVU.MOREX.r2.5HG0378970) was used as the internal reference gene and the detail information of all the primers was listed in Supplementary Table S2. The total RNA was extracted by Plant RNA Kit reagent (Omega BioTek, USA), and cDNA synthesis was performed using 5X All-in-one RT

MasterMix (ABM, Canada) according to the manufacturer's instructions. The TB-Green ®Premix Ex Taq™ II kit (Takara, Dalian, China) was used to conduct qRT-PCR amplification in the Quant Studio™ Real-Time PCR system (Thermo Fisher, USA). The thermal cycling protocol was as follows: 95°C temperature for 30 seconds, followed by 40 cycles of 95°C for 3 seconds, and 30 seconds at 60°C. The relative expression level was calculated by the $2^{-\Delta\Delta CT}$ method [106]. Three technical replications were applied for each treatment. The T-test was conducted to evaluate the significance of the results using R. One asterisk (*) indicate 0.05 significance level and double asterisk (**) indicate 0.01 significance level, respectively.

Population Genetics Analysis of *HvC3H*-related Variants

The exome-captured resequencing data of 220 geographically-referenced barley accessions were retrieved from the NCBI SRA database (PRJEB8044/ERP009079) [93]. Raw reads were trimmed using Trimmonmatic v0.36 with default parameters [107]. The high-quality reads were mapped to the reference genome of barley Morex V2 using BWA-MEM v0.7.13r1126. The single nucleotide polymorphism (SNP) and insertion-deletion (InDel) was identified using the Picard-GATK pipeline [108]. The following criteria was used for SNPs filtration. (1) minor allele frequency (MAF) >0.05 and <0.95; (2) a maximum missing rate <0.1; (3) biallelic alleles. SNP and InDel were annotated using the SnpEff v4.3 according to the barley genome GTF file [109]. To better reveal the evolutionary relationships, barley accessions with SNP missing rate larger than 0.1 were excluded. Totally, the final collection included 95 landraces and 51 wild barley accessions (Supplementary Table S3). Only SNPs that located within the *HvC3H* genes were extracted for phylogenetic tree, population structure and principal component analysis (PCA). The *HvC3H*-related SNPs were used to construct NJ tree with the MEGA X. The phylogenetic tree was visualized by Interactive Tree of Life (iTOL). The population structure was quantified using ADMIXTURE v1.3.0 with predefined K values ranging from 2 to 4. The PCA was performed using the samrtpca sub-package implemented in EIGENSOFT v4.2. The nucleotide diversity (π) was estimated using vcftools v0.1.16. The DNAsp v6.12.01 was employed to calculate the haplotypes for each *HvC3H* genes. Finally, the media-joining haplotype networks were constructed using the PopART v1.7 [110].

Results

Genome-wide Identification and Characterization of CCCH Proteins in Barley

The most updated barley Morex assembly was used for the identification of barley *CCCH* genes. A multifarious approach was used to obtain the most comprehensive results. First, the AtC3H and OsC3H proteins were used as queries to search against barley proteins as well as the HMM profile of the CCCH domain in Pfam (PF00642). A total of 52 putative *CCCH*-encoding genes were acquired from the barley genome. The protein sequences of *HvC3H* candidates were submitted to the Pfam, NCBI-CDD, and SMART online databases to ensure that the putative genes contained the CCCH domain. Only proteins that were verified by at least two databases were retained (Supplementary Table S4). A total of 17 genes were excluded from our dataset because no complete CCCH domain was verified by these online databases. A total of 35 high-confident *HvC3H* genes with complete open reading frames were identified, accounting for 1.07% of the total annotated protein-coding genes in barley (Table 1). Because there is no standard nomenclature for barley CCCH genes, the candidate *HvC3H*s were designated as *HvC3H1* to *HvC3H35* according to their chromosomal number and location. A BLAST search against the barley ESTs indicated that 32 of the 35 *HvC3H*s possessed EST records, which supported the existence of the *HvC3H*s. Analysis of the physicochemical properties of HvC3H demonstrated that the amino acid length varied from 211 amino acids (HvC3H25) to 1,456 amino acids (HvC3H4), with an average length of 457.2 amino acids. The pI varied from 5.13 to 10.15, and the MW ranged from 4.501 kDa to 160.371 kDa. All of these CCCH proteins possessed negative GRAVY values (average value: -0.7715), indicating the hydropathic nature of HvC3Hs. Subcellular localization prediction showed that most of these proteins were located in the nucleus (29 HvC3Hs; 82.86%). In addition, three HvC3Hs were located in the chloroplast, and the other three were located in both the nucleus and chloroplast, which was consistent with their localization in *Arabidopsis*, rice, and wheat [11, 44].

Table 1
Characteristics of CCCH transcription factor gene family in barley.

Gene Name	Gene ID	Protein Length (aa)	Isoelectric Point	Molecular Weight (kDa)	Subcellular Location	Grand Average of Hydropathicity	ESTs Hit
HvC3H1	HORVU.MOREX.r2.1HG0072390	304	9.15	34.276	Nucleus	-1.163	11
HvC3H2	HORVU.MOREX.r2.1HG0074950	402	9.57	42.999	Nucleus	-0.305	2
HvC3H3	HORVU.MOREX.r2.1HG0074970	327	9.42	35.232	Nucleus	-0.389	6
HvC3H4	HORVU.MOREX.r2.1HG0078680	1456	5.13	160.373	Nucleus	-0.84	14
HvC3H5	HORVU.MOREX.r2.2HG0091490	697	6.26	73.624	Chloroplast	-0.384	25
HvC3H6	HORVU.MOREX.r2.2HG0140180	341	10.15	38.974	Nucleus	-1.009	22
HvC3H7	HORVU.MOREX.r2.2HG0166340	304	9.38	31.487	Nucleus	-0.623	33
HvC3H8	HORVU.MOREX.r2.2HG0176080	695	5.49	77.804	Nucleus	-1.149	11
HvC3H9	HORVU.MOREX.r2.3HG0196310	379	7.51	40.767	Nucleus	-0.53	57
HvC3H10	HORVU.MOREX.r2.3HG0200320	224	9.13	24.591	Nucleus	-0.81	0
HvC3H11	HORVU.MOREX.r2.3HG0200330	232	6.17	24.956	Nucleus	-0.616	1
HvC3H12	HORVU.MOREX.r2.3HG0210880	467	7.85	49.838	Nucleus	-0.493	53
HvC3H13	HORVU.MOREX.r2.3HG0210900	426	8.07	4.502	Nucleus	-0.658	16
HvC3H14	HORVU.MOREX.r2.3HG0225190	500	8.67	54.722	Nucleus	-0.648	32
HvC3H15	HORVU.MOREX.r2.3HG0228250	384	8.6	42.042	Nucleus	-0.419	9
HvC3H16	HORVU.MOREX.r2.3HG0230880	281	9.59	32.735	Chloroplast Nucleus	-1.177	64
HvC3H17	HORVU.MOREX.r2.3HG0258540	435	8.82	47.4	Nucleus	-0.564	61
HvC3H18	HORVU.MOREX.r2.4HG0318770	299	8.3	32.522	Nucleus	-0.595	2

Table 1
Characteristics of CCCH transcription factor gene family in barley (Continued)

Gene Name	Gene ID	Protein Length (aa)	Isoelectric Point	Molecular Weight (kDa)	Subcellular Location	Grand Average of Hydropathicity	ESTs Hit
HvC3H19	HORVU.MOREX.r2.4HG0325540	326	7.14	36.231	Nucleus	-1.003	11
HvC3H20	HORVU.MOREX.r2.5HG0362710	691	9.39	78.976	Nucleus	-1.218	5
HvC3H21	HORVU.MOREX.r2.5HG0374920	509	6.39	55.923	Nucleus	-0.702	23
HvC3H22	HORVU.MOREX.r2.5HG0377520	442	8.78	47.516	Nucleus	-0.504	20
HvC3H23	HORVU.MOREX.r2.5HG0407060	314	9.25	36.792	Chloroplast Nucleus	-1.241	36
HvC3H24	HORVU.MOREX.r2.5HG0429150	752	7.39	85.417	Nucleus	-1.256	5
HvC3H25	HORVU.MOREX.r2.6HG0475520	211	9.17	23.087	Nucleus	-0.813	0
HvC3H26	HORVU.MOREX.r2.6HG0475540	358	8.63	38.145	Chloroplast	-0.34	0
HvC3H27	HORVU.MOREX.r2.6HG0475570	308	9.49	31.997	Chloroplast	-0.541	39
HvC3H28	HORVU.MOREX.r2.6HG0505660	359	6.66	40.211	Nucleus	-0.46	75
HvC3H29	HORVU.MOREX.r2.6HG0515160	1001	8.81	110.273	Nucleus	-1.121	16
HvC3H30	HORVU.MOREX.r2.6HG0526270	647	5.23	71.397	Nucleus	-1.069	45
HvC3H31	HORVU.MOREX.r2.7HG0560290	489	9.46	55.275	Nucleus	-0.79	59
HvC3H32	HORVU.MOREX.r2.7HG0579580	363	6.85	38.791	Nucleus	-0.706	20
HvC3H33	HORVU.MOREX.r2.7HG0600900	297	9.64	30.833	Chloroplast Nucleus	-0.541	46
HvC3H34	HORVU.MOREX.r2.7HG0602740	407	8.56	44.684	Nucleus	-0.954	26
HvC3H35	HORVU.MOREX.r2.7HG0607870	375	9.32	42.524	Nucleus	-1.372	6
Supplementary materials							

Ccch Domain Structure Analysis Of Hvc3hs

To evaluate the degree to which CCCH proteins in barley are evolutionarily conserved, the full-length proteins were used to characterize the domain organization of HvC3Hs. Significant differences in the domain organization of HvC3Hs were observed. A total of five domain organizations of 77 CCCH motifs (C-X_{7-10}-C-X_{4-5}-C-X_3-H) were identified, with an average of 2.2 CCCH motifs per protein. CCCH proteins have been shown to have one to six CCCH motifs (Figure 1) [11, 22, 49, 50], and the same pattern was observed in our study. However, no HvC3H proteins contained four CCCH motifs. Approximately 77.14% (27) of the HvC3H proteins had one or two CCCH motifs, 14.29% contained five copies, and 5.71% had six copies (Supplementary Table S5); similar patterns have also been observed in rice and *Arabidopsis*[51]. Although different frequencies of CCCH motifs have been identified among barley *CCCH* genes, two conventional CCCH motifs C-X_8-C-X_5-C-X_3-H and C-X_7-C-X_5-C-X_3-H were the two most common, suggesting that C-X_{7-8}-C-X_5-C-X_3-H might be ancestral to the other CCCH motifs (Supplementary Figure S1). Additionally, a total of four non-conventional CCCH zinc finger motifs, including 2 C-X_7-C-X_4-C-X_3-H, 1 C-X_9-C-X_5-C-X_3-H, and 1 C-X_{10}-C-X_5-C-X_3-H, were observed, which were previously identified to be abundant non-conventional CCCH motifs in *Arabidopsis* and rice. In contrast to other plants, the uncommon CCCH motifs C-$X_{4-6,11-15}$-C-X_{4-5}-C-X_3-H were not detected; no novel motif types were detected in barley CCCH proteins, suggesting that the CCCHs in barley are highly conserved. Aside from the CCCH zinc finger motifs, some HvC3H proteins also contained several other known functional domains, such as KH, RRM, and RING. Four (HvC3H7, -26, -27, and -33) and five (HvC3H1, -5, -16, -23, and -24) HvC3H members possessed the KH and RRM domains, respectively.

Phylogenetic Relationships, Gene Structure, and Conserved Domain Organization of HvC3H Genes

To determine the evolutionary relationships among *HvC3H*s, a neighbor-joining (NJ) phylogenetic tree was constructed based on the alignment of the full-length CCCH protein sequences of barley. According to the criteria proposed by Wang and Peng et al. with slight modifications [11, 13], the 35 *HvC3H* family members within the phylogenetic tree were classified into seven subfamilies (group I to group VII) (bootstrap values >60%) (Figure 2A). Sixteen *HvC3H*s formed eight sister gene pairs, seven of which possessed high bootstrap support (98%). Most of the phylogenetic clades had high bootstrap support; the relationships of some CCCH proteins (eight HvC3Hs) remained ambiguous because of their low bootstrap values at deep nodes. The number of CCCH proteins varied greatly for different subfamilies; group I and group II had seven and five members, respectively, whereas groups V and VI had only two members. We also constructed another phylogenetic tree based on the alignment of 170 CCCH proteins from *Arabidopsis* (68), rice (67), and barley (35) (Supplementary Figure S2). The phylogenetic tree revealed an alternating distribution of monocot and eudicot *CCCH* genes in most of the subfamilies. Within each subfamily, the *HvC3H* genes tended to cluster with their *Arabidopsis* and/or rice orthologs, suggesting that the orthologous genes might have evolved from a common ancestor when duplication events took place prior to species divergence. However, some groups consist of *Arabidopsis* and rice *CCCH* genes but lack barley genes. This indicates a presumed barley-specific loss of *CCCH* genes following the divergence of monocots and dicots.

The intron–exon gene structure not only provides important insights into the functional diversification of genes but also the evolutionary relationships within a gene family [52]. Unlike other TF family genes, which tend to lack introns, the average intron number of *HvC3H*s was 3.97 (ranging from 0 to 12) (Figure 2B). In general, genes within the same subfamily had a similar structure of introns and exons. For example, genes from subfamily VII tended to be intron-less; subfamilies II, IV, and VI were nearly identical in their intron/exon lengths and structural organization. The intron–exon gene structure of subfamily I was highly variable. For example, *HvC3H14* possessed 12 introns, whereas *HvC3H12*, *HvC3H17*, and *HvC3H22* had six introns. *HvC3H* genes within subfamily I also had variable intron lengths.

To gain additional insights into the functional regions of *HvC3H*s and their evolutionary relationships, the distribution of conserved motifs of HvC3Hs was visualized (Figure 2C). Consistent with the patterns in intron–exon gene structure, HvC3H proteins within the same subfamily tended to have a similar organization of motifs, and the patterns were highly variable among different phylogenetic clades. For example, the HvC3Hs in subfamily VII possessed two CCCH motifs and one RRM motif. Subfamilies IV, V, and VI had only one CCCH motif. Although the gene structure varied greatly among subfamily I members, the protein motif composition was conserved; there were five to seven CCCH motifs for each protein. The variation in gene structure and motif composition among subfamilies suggests prior sub-functionalization or neofunctionalization of these *HvC3H*s.

Chromosomal Distribution And Gene Duplication

Chromosome location analysis revealed that the 35 *HvC3H* genes were unevenly located on the seven barley chromosomes, and chromosome 3H possessed the most abundant *CCCH* genes (nine *HvC3H*s) (Supplementary Figure S3). By contrast, chromosomes 1H, 2H, and 4H had four, four, and two *CCH* genes, respectively. Chromosome 6H contained six *CCCH* genes, and chromosome 5H and 7H both had five. There was no significant correlation between the number of *HvC3H* genes and chromosome length (Pearson correlation r = 0.0278, p-value = 0.9528), demonstrating that longer chromosomes do not have more *HvC3H* genes.

Gene duplication is considered one of the primary drivers of gene family expansion in plants and plays an important role in the evolution of new gene functions and adaptation [53, 54]. A total of three duplicated gene pairs were identified (Figure 3). Two gene pairs (*HvC3H25*/*HvC3H33* and *HvC3H28*/*HvC3H31*) were segmentally duplicated genes; all segmentally duplicated genes were located between chromosome 6H and 7H. *HvC3H10*/*HvC3H11* were tandem duplicated genes and located on chromosome 3H. These duplicated genes were clustered in the same clade. Specifically, *HvC3H10/HvC3H11* were clustered in subfamily III, *HvC3H25*/*HvC3H33* were assigned to subfamily II, and *HvC3H28*/*HvC3H31* belonged to subfamily VI. The

aforementioned results indicated that the segmental and tandem duplication events might have played a role in the expansion of the *HvC3H* gene family.

To further evaluate the evolutionary constraints acting on *HvC3H*s, the Ka/Ks ratios for each duplicated gene pair were calculated. The Ka/Ks ratios of the segmentally duplicated genes *HvC3H28/HvC3H31* and *HvC3H25*/ *HvC3H33* and the tandem duplicated genes *HvC3H10*/ *HvC3H11* were all lower than 1, indicating purifying selection (Supplementary Table S6) [55]. The timing of duplication events was estimated to be approximately 22.83 – 808.04 MYA.

Syntenic relationships with six other representative species, including three monocots (*Zea mays*, *Oryza sativa*, and *Triticum aestivum*) and three dicots (*Brassica rapa*, *Solanum lycopersicum*, and *Glycine max*), were analyzed to determine the mechanisms underlying the evolution of *HvC3H*s (Figure 4). A total of 63, 24, and 16 orthologous gene pairs between barley and *Triticum aestivum*, *Zea mays*, and *Oryza sativa* were identified, respectively. A total of 16 of 24 (66.7%) of the *HvC3H* genes were orthologous to three copies of *CCCH* genes in wheat, which might stem from the fact that the heterologous hexaploid wheat contained three distinct ancestral genomes, namely A, B, and D [56]. *HvC3H11* on chromosome 3H was orthologous to four *CCCH* genes in wheat, which were located on chromosome 3A (one gene), chromosome 3B (one gene), and chromosome 3D (two genes). In addition, six *HvC3H*s possessed two orthologous genes, and two *HvC3H*s had one orthologous gene in wheat. This inconsistency might stem from gene loss and duplication events during the process of wheat polyploidization. However, the number of orthologous gene pairs between barley and three dicots (*Glycine max*, *Brassica rapa*, and *Solanum lycopersicum*) was ten, five, and one, respectively, which was much lower than those between barley and three monocots. This finding is consistent with the observed phylogenetic relationships between barley and these species. *HvC3H*s are phylogenetically closer to *CCCH*s in *Triticum aestivum*, *Zea mays*, and *Oryza sativa* than *CCCH*s in *Glycine max*, *Brassica rapa*, and *Solanum lycopersicum*. Furthermore, we found that most of the *HvC3H* genes showed syntenic bias towards the chromosomes of these plants. For example, orthologous genes were on chromosome 1 in *Oryza sativa*, whereas the syntenic relationships in *Zea mays* tended to be located on chromosomes 3 and 8, suggesting that chromosomal rearrangements such as inversions or duplications might shape the organization and distribution of *CCCH* genes in these genomes. A total of seven *HvC3H* genes (*HvC3H7*, *-11*, *-14*, *-15*, *-17*, *-18*, and *-31*) were only syntenic with *Triticum aestivum*, *Zea mays*, and *Oryza sativa*, suggesting that these orthologous pairs may have formed after the divergence between monocotyledonous and dicotyledonous plants.

The Ka/Ks ratios were calculated for each orthologous gene pair to characterize selection on orthologous gene pairs. The overall Ka/Ks ratios (*Triticum aestivum*: 0.2733, *Oryza sativa*: 0.1682, *Zea mays*: 0.1873, *Brassica rapa*: 0.0149, *Solanum lycopersicum*: 0.0166, and *Glycine max*: 0.0295) of all the orthologous gene pairs were less than 1, suggesting that these *HvC3H*s might have experienced strong purifying selection (Supplementary Table S7).

Cis **-element Analysis of** HvC3H **Genes**

Cis-elements play important roles in the transcriptional regulation of genes throughout the life cycle of plants. To obtain preliminary insights into the potential function and transcriptional regulation of *HvC3H* genes, the *cis*-regulatory elements in the promoter regions were analyzed. A total of 52 functional *cis*-elements were identified and grouped into five categories. A large number of light-responsive elements were detected in the promoter regions of *HvC3H*s, which accounted for most of the putative *cis*-elements (Supplementary Table S8, Supplementary Figure S4). We also obtained a total of 11 types of hormone-responsive regulatory elements, such as auxin-responsive elements (AuxRR-core, TGA-box, and TGA-element), gibberellin-responsive elements (P-box, GARE-motif, and TATC-box), salicylic acid-responsive elements (TCA-element), MeJA-responsive elements (CGTCA-motif and TGACG-motif), ABA-responsive elements (ABRE), and ethylene-responsive elements (ERE). Several types of biotic and abiotic stress-related regulatory elements were observed in *HvC3H* promoters. Anaerobic induction elements (44 ARE and 30 GC-motif) were detected in 29 *HvC3H*s, which were the most abundant *cis*-regulatory elements involved in the response to environmental stress. A total of 23 low temperature-responsive elements (LTR) and 28 drought-responsive elements (MBS, myeloblastosis ninding site) were detected in 15 and 17 *HvC3H*s, respectively. A total of six *HvC3H* genes possessed wound-responsive elements (WUN-motif), suggesting that these genes might play essential roles in the response to biotic stress in plants. Additionally, eight types of plant organogenesis-related *cis*-elements were identified, such as the meristem expression-

related element CAT-box (11 genes), zein metabolism regulation-related element O2-site (seven genes), endosperm expression-related element GCN4-motif (six genes), and seed-specific regulation element RY-element (four genes). These findings suggest that *HvC3H*s might play an important role in barley plant growth and development, hormone signal transduction, and the response to biotic and abiotic stress.

Temporal-spatial And Stress-induced Expression Pattern Analysis

Analysis of tissue-/stage-specific expression profiles provided valuable insights into the potential functions of genes in plant species. To investigate the expression patterns of *HvC3H*s in different tissues and developmental stages, publicly available RNA-seq data from 16 different tissues were used to estimate the expression levels of *HvC3H*s, and the *HvC3H*s were clustered into three major groups according to their expression levels (Figure 5). Distinct expression patterns were observed for the *HvC3H*s, suggesting that these genes might have undergone significant differentiation and played various roles in particular stages during barley growth and development. The expression levels of *HvC3H*s in group I were lower than those of genes in the other groups; eight genes were not expressed in most of the tissues/stages. By contrast, a total of eight genes in group II were highly expressed in all studied tissues/stages. *HvC3H17* was predominantly expressed in LOD, CAR15, and EPI, whereas *HvC3H1* and *HvC3H12* showed high expression in INF2 and LOD, respectively. Genes in cluster III showed a medium expression level. Within this cluster, *HvC3H5*, *-7*, *-9*, *-27*, and *-35* tended to be expressed in INF1 and INF2. These findings indicate that these *HvC3H*s might be associated with the development of these tissues in barley. These genes might thus be interesting targets for barley breeding. The differentiation of subfamilies observed in the phylogenetic tree was not consistent with the expression clustering results, indicating that sequence similarity was not a strong predictor of expression patterns and functional similarity [57].

We analyzed the expression of *HvC3H*s in response to different types of environmental stress. Under cold treatment, six *HvC3H*s (*HvC3H2*, *-3*, *-10*, *-11*, *-15*, and *-26*) with FPKM values equal to 0 were excluded from the subsequent analysis (Figure 6A). The rest of the *HvC3H*s were clustered into two major groups. Five *HvC3H* genes displayed increased expression patterns (>2.0-fold change). Among these genes, *HvC3H18*, *HvC3H9*, and *HvC3H20* exhibited their highest level of expression under cold treatment, showing fold-changes of 9.21, 3.01, and 2.88, respectively. In addition, the expression of six genes was slightly down-regulated under cold treatment. Specifically, the expression of *HvC3H4* decreased 0.65-fold, and *HvC3H17* decreased 0.74-fold compared with the control, which was not exposed to low temperature. Salt stress induced differential expression patterns of *HvC3H* genes in the three root regions (Figure 6B). Compared with the unstressed control, a total of four, five, and three *HvC3H* genes were highly expressed in the meristematic, elongation, and maturation zones, respectively, especially *HvC3H15*, which exhibited a 23.16 and 32.50-fold increase in expression in the meristematic and elongation zones relative to the unstressed control. *HvC3H17* was up-regulated in all tissues; its expression was increased 8.18-, 8.41-, and 2.11-fold in the meristematic, elongation, and maturation zones, respectively. Similarly, the expression of *HvC3H22* and *HvC3H28* was up-regulated in both the meristematic and elongation zones, and *HvC3H9* was significantly up-regulated in the elongation and maturation zones. Under metal ion stress, the expression of *HvC3H4*, *HvC3H9*, and *HvC3H18* was significantly up-regulated, and the up-regulation of *HvC3H9* was induced by copper treatment (Figure 6C). Under zinc stress, *HvC3H3* was up-regulated 2.67-fold.

Expression of HvC3Hs under Drought, Salt, Cold, and ABA Treatment by qRT-PCR Analysis

To investigate the expression of *HvC3H* genes in response to multiple treatments, 26 *HvC3H*s were randomly subjected to qRT-PCR analysis. Under drought treatment, nine *HvC3H*s were up-regulated at all time points (Supplementary Figure S5), and the expression of six of the nine *HvC3H*s (*HvC3H1*, *-5*, *-6*, *-12*, *-24*, and *-33*) peaked at 24 h. The expression of *HvC3H1* was approximately 54-fold larger than that of the control at 24 h. Several MBS *cis*-acting elements associated with drought inducibility were also identified within their promoter regions. However, no MBS *cis*-acting elements were detected in some genes, such as *HvC3H5* and *HvC3H6*, suggesting that unknown regulatory elements related to drought stress might exist for these genes.

After salt treatment, the expression of *HvC3H13* was suppressed compared with the control at all time points; the expression of 21 genes was significantly up-regulated, and the expression of these genes peaked at different times (Supplementary Figure

S6). For example, the expression of *HvC3H1* peaked at 3 h and was up-regulated 43-fold, whereas the expression of *HvC3H5*, *-6*, and *-12* was initially slightly up-regulated and peaked at 24 h.

The expression levels of *HvC3H*s after cold treatment were analyzed, and the expression of six genes (*HvC3H4*, *-5*, *-7*, *-20*, *-25*, and *-27*) was inhibited compared with the control; the expression of the other *HvC3H*s was up-regulated at specific time points (Supplementary Figure S7). The expression of three *HvC3Hs* was up-regulated at 1 h (*HvC3H23*, *-30*, and *-32*), 3 h (*HvC3H6*, *-28*, and *-33*), and 6 h (*HvC3H1*, *-17*, and *-22*), suggesting that these *HvC3H*s might primarily function in the initial stage in the response to cold injury. The expression of the other *HvC3H* genes peaked at 12 h or 24 h. The promoters of these genes were found to possess abundant LTR *cis*-elements, which might be responsible for their expression profiles under cold stress.

Plant CCCH proteins have been shown to be effective regulators of ABA-mediated stress responses [58]. qRT-PCR analysis showed that ABA treatment had a pronounced effect on the expression patterns of *HvC3H*s, and a complex expression profile was observed (Supplementary Figure S8). For example, the expression of *HvC3H1* was significantly up-regulated at 1 h and 3 h but down-regulated thereafter. By contrast, the expression of *HvC3H24* was down-regulated before 12 h but significantly up-regulated at 24 h. With the exception of *HvC3H5*, *-13*, *-20*, *-21*, and *-30*, whose expression was suppressed relative to the control, the maximum expression levels of the other *HvC3H*s peaked at different time points. Abundant ABRE *cis*-elements, which are important *cis*-acting elements associated with ABA-responsiveness, were detected in the promoter regions, including five ABRE *cis*-elements for *HvC3H23*, three for *HvC3H33*, and two for *HvC3H17*. Overall, our results demonstrated that ABA sensitivity might be a common feature of *HvC3H* genes.

Genetic Variation of CCCH Genes

We analyzed the sequence diversity of *HvC3H* genes at the population level based on exome-captured sequencing datasets. The single nucleotide polymorphism (SNP)-calling pipeline generated 331 high-confidence SNPs, 172 of which were in *HvC3H21*, followed by *HvC3H14* (42) and *HvC3H34* (23) (Supplementary Table S9; Supplementary Table S10). Most *HvC3H*-related SNPs were located within the intron regions (318); the rest of the SNPs were non-synonymous (8) and synonymous (5) SNPs, with a non-synonymous to synonymous ratio of 1.6. There were 277 InDels ranging from 1 to 55 bp in length, and short InDels 1 to 4 bp (76.90%) in length were more common than long InDels (Supplementary Table S11). Similarly, most *HvC3H*-related InDels were located in introns, which might be explained by the fact that the reading frame-independent variants were under weaker negative selection than frame-change variants.

To investigate the relatedness among the landraces and wild barley accessions worldwide, we carried out principal component analysis using *HvC3H*-related SNPs (Figure 7A,B; Supplementary Table S12). The first principal component was correlated with the biological differentiation between landrace from wild barley and explained 25.16% of the total genetic variance; the second and third principal components captured 5.09% and 5.00% of the genetic variance, respectively, and revealed geographical differentiation in barley accessions. These patterns were consistent with the topology of the NJ tree (Figure 7C). The phylogenetic tree revealed genetically divergent clusters associated with the contrast between barley wild accessions versus landraces rather than their geographical origins. ADMIXTURE analysis confirmed these findings (Figure 7D). When K=2, two groups coinciding with landraces and wild barley were observed. Increasing K to 4 provided additional insights. Within barley landraces, we detected two geographically distributed components from Europe and Africa, whereas the rest of the landraces from Mediterranean areas displayed signs of genetic admixture. Within wild barley accessions, accessions from the Southern and Northern Levant regions formed two distinct groups.

Genetic Diversity and Haplotypes of HvC3Hs in Wild and Domesticated Barley Populations

Population-based nucleotide diversity was calculated to assess the occurrence of prior genetic bottlenecks of *HvC3H* genes during barley domestication. The genetic diversity of *HvC3H*s was further analyzed using exome-captured resequencing datasets; a total of 51 barley landraces and 95 wild barley accessions were included in the analysis. The total genetic diversity of *HvC3H* genes decreased by ~34.94% from the wild (π = 0.07622) to domesticated (π = 0.04959) barley population (Supplementary Figure S9).

A haplotype consists of a group of closely linked SNPs located in a specific chromosome region that determine the same trait. We constructed the haplotype network for each *HvC3H* gene using their SNPs. A total of 529 haplotypes belonging to 22 *HvC3H* genes were observed, with an average of 24.05 haplotypes per gene (Figure 8). Specific haplotypes represented in more than half of wild or landrace populations were defined as dominant haplotypes. Four *HvC3H* genes (*HvC3H14*, *-21*, *-22*, and *-30*) had no dominant haplotype, whereas 13 *HvC3H* genes had the same dominant haplotype in both wild and landrace populations. Nevertheless, clear genetic differentiation in haplotypes between wild and domesticated barley accessions was observed. *HvC3H20* in wild barley mainly had the AAAAGGGG**GG**TTTTGGCC haplotype, but domesticated barley mainly had the AAAAGGGG**AA**TTTTGGCC haplotype. The dominant haplotype in wild barley was AAGTTTTC**CC**TTGGGG**AA**, but haplotype AAGTTTTC**TT**TTGGGG**TT** was the most common in domesticated barley. Some rare haplotypes were also observed for specific *HvC3H* genes, such as *HvC3H32*, *HvC3H34*, and *HvC3H35*. The appearance of novel allelic variants greatly increased the degree of haplotype polymorphism of *HvC3H*s. The rare haplotypes were mainly observed in the wild barley group, which was consistent with the results of the genetic diversity analysis indicating that these genes experienced a severe genetic bottleneck during barley domestication and that the haplotype diversity decreased in domesticated barley relative to the wild population. By contrast, the haplotype polymorphism of *HvC3H14* and *HvC3H20* in domesticated barley was higher than that in wild barley. We speculate that the process of artificial breeding has introduced various novel alleles and increased the degree of haplotype polymorphism.

Discussion

Characterization of CCCH Genes in Barley

CCCH genes comprise a distinct zinc finger TF gene family that regulates gene expression through RNA binding and processing. CCCH domain-containing proteins are involved in various biological processes of plant growth, development, and adaptation. Given their important regulatory roles, many CCCH zinc finger family genes have been identified and functionally characterized in model plants and crops [45]. Barley is the fourth most abundant cereal crop worldwide in terms of both acreage and tonnage harvested, and its adaptability to a wide range of cultivation conditions might stem in part from its ability to withstand harsh environments [46]. However, there is limited information on CCCH proteins in agriculturally important cereal crops. The large genome size and high content of transposon elements have inhibited the assembly of the barley genome. With the advent of novel sequencing technologies (e.g., 10X genomics, chromosome conformation capture sequencing (Hi-C), and Bionano optical mapping) and computational algorithms (e.g., TRITEX workflow), barley genome assemblies, from the first draft reference genome to its subsequent revisions (Morex V1 and V2) have undergone multiple rounds of improvement [46–48]. The most recently updated barley reference genome assembly (Morex V3) based on PacBio HiFi sequence has near-complete coverage and high performance in the repeat-rich intergenic regions. Because the same transcriptomic resources were used for the previous genome assembly, the gene models of Morex V3 displayed near-complete alignments (≥95% alignment coverage, ≥99% identity) with the V2 assembly in the gene space [59]. Presently, we have determined the expression profile and characterized genetic variation using Morex V2 as the reference genome. Given the consistency in the gene models and computational calculations, comprehensive identification and characterization of *HvC3H*s herein were performed using the barley Morex V2 assembly, which was considered to be a qualified reference genome for functional gene analysis [60–62].

A BLAST search [11, 41], HMM search [44, 45], or both [13, 35] were used to identify candidate *CCCH* genes in various species. Here, both search tools were adopted in our study based on the following criteria: BLAST e-value<1e-5 and HMMER e-value<0.001. The methodology was further validated by the SMART, NCBI-CDD, and Pfam online databases. Our results enhance our understanding of *HvC3H*s. We identified 35 high-confident *CCCH* zinc finger genes within the barley genome through a comprehensive search. We found that the number CCCH proteins in barley was approximately half of those identified in *Arabidopsis* (67), rice (68), maize (68), and grape (69); a third of those in poplar (91), *Brassica rapa* (103), and switchgrass (103); and a similar number to those in *Medicago truncatula* (34), suggesting that species origin and genome size are not directly associated with the number of CCCH genes.

Plant CCCH proteins usually contain one to six CCCH motifs [11]. Similar to the other species, each barley CCCH protein has at least one CCCH motif, and the most common number of CCCH motifs was one (17 *HvC3H*s) and two (10 *HvC3H*s). Although

plant CCCH proteins displayed a wide spacing pattern similar to C-X_{4-15}-C-X_{4-6}-C-X_3-H, two conventional CCCH motif types, C-X_7-C-X_5-C-X_3-H and C-X_8-C-X_5-C-X_3-H, were the most common in different species, suggesting that the consensus sequence C-X_{7-8}-C-X_5-C-X_3-H might be ancestral to the other CCCH motifs [35]. The CCCH motifs in barley showed similar patterns to the highly conserved consensus sequences, indicating that these domains might be important for the function of the C3H proteins. With the exception of the CCCH motifs, several other functional domains, such as KH, RRM, and RING, were also detected in barley CCCH proteins. Some studies indicate that KH and RRM are common RNA-binding domains in eukaryotes, and KH and RRM domain-containing proteins are involved in various aspects of RNA metabolism, such as pre-mRNA splicing, nucleocytoplasmic transport of mRNAs, polyadenylation, and transcriptional control [63, 64]. These results suggest that the four CCCH proteins harboring these specific domains might also function as RNA-binding proteins and play a role in RNA processing. In addition, *HvC3H31* contained a RING domain, which suggests that it plays a key role in the response to various types of abiotic stress [65]. These additional domains might contribute to the neo-functionalization of *HvC3H* genes.

Expansion Patterns Might Affect the Gene Family Size of HvC3Hs

The 35 barley *CCCH* genes were clustered into seven subfamilies based on the phylogenetic analysis. Genes within the same subfamily tend to be similar in structure and domain composition, which supported the observed topology. The number of clades of *HvC3H*s was slightly lower than the number of clades of *HvC3H*s in *Arabidopsis* (11) and rice (8). Some genes could not be assigned to any groups because of their low sequence similarity. Therefore, a combined NJ tree was constructed to evaluate their evolutionary relationships using *CCCH* proteins from barley, rice and *Arabidopsis*. Most of the *HvC3H* genes displayed closer relationships with their orthologs than their paralogs, suggesting that these CCCH proteins might have originated from the common ancestor of these plants and then expanded independently in different species. Moreover, rice and/or *Arabidopsis*-specific clades without barley genes were also observed in our phylogenetic analysis, suggesting that a presumed barley-specific loss of *CCCH* genes may have occurred after the divergence between barley and other species.

A gene family is a group of similar genes that originated from a common ancestor, and gene duplications, such as segmental, tandem, whole-genome, or polyploidization duplications, contribute greatly to the rapid expansion and evolution of genes in a gene family [66–68]. In barley, duplication events were a dominant force contributing to the expansion of some gene families. For example, ten and seven pairs of *HvGRAS*s have undergone segmental and tandem duplications, accounting for approximately 54.83% of the total number of *HvGRAS* genes [69]. Segmental duplication was the main force driving the expression of *HvbHLH* genes (42.55%) [70], whereas tandem duplication was largely responsible for the expansion of *HvNRT2.1* (80.00%) [71]. Previous studies have indicated that duplication events are responsible for the expansion of the *CCCH* gene family. In wheat, allopolyploidization duplication stems from the combination of three distinct ancestral genomes, viz. A, B, and D, which significantly expands the number of *CCCH* gene members. Synteny analysis identified 63 gene pairs (comprising 24 *HvC3H*s and 63 *TaC3H*s) in the barley and wheat genomes. A total of 15 *HvC3H* genes were orthologous to three *TaC3H* genes corresponding to the three sets of sub-genomes, which confirmed the hexaploidy duplication of the wheat genome. Additionally, some *CCCH* genes of diverse species have undergone segmental and/or tandem duplications [11, 13, 38]. By contrast, only three gene pairs (one tandem duplication and two segmental duplications) were involved in the duplication of *HvC3H*s, which was far less than the number of gene pairs in other species. Therefore, despite its large genome size, barley had fewer *CCCH* gene members, which might be attributed to the fact that its *HvC3H*s have undergone fewer duplication events.

HvC3Hs Might Play a Role in Plant Growth, Abiotic Stress, and Phytohormone Responses

CCCH proteins are an important class of zinc finger TFs that are involved in various plant growth processes and responses to stress, but their potential functions in barley are rather limited. *Cis*-acting regulatory elements within the promoter region are involved in the spatiotemporal regulation of gene expression [72]. In our current study, various *cis*-elements in the promoter regions of *HvC3H*s were identified and divided into five groups based on their functions: structure and composition, hormone induction, plant growth development/tissue-specific, stress, and light response. These results indicated that *HvC3H*s mediate various functions in multiple biological processes. Given that prediction of *HvC3H cis*-elements was carried out based on a bioinformatics approach, experiments need to be conducted to validate the function of predicted *HvC3H* genes.

The biological function of *CCCH*s has been well-documented in various plants, especially model plants [73]. The diversity of expression patterns of *HvC3H*s in specific tissues or at specific stages provides insight into their possible functions at specific developmental stages. For example, *HvC3H7*, a KH domain-containing *CCCH* zinc finger gene, tended to be expressed highly in young inflorescences. In plants, the KH domain-containing genes *FLK* (*Flowering Locus KH Domain*) and *PEP* (*PEPPER*) regulate flowering by negatively and positively modulating the expression of *FLC*, respectively [74, 75]. The homologs of *HvC3H17* in *Arabidopsis* are *KHZ1* and *KHZ2*, which are two non-tandem CCCH zinc finger and K-homolog domain proteins that act as repressors of the splicing efficiency of *FLC* pre-mRNA and redundantly promote flowering and senescence [76, 77], suggesting that *HvC3H17* might be involved in the transition from vegetative to reproductive development in barley. *HvC3H15* was expressed highly in NOD, PAL, and LEM. Its orthologous gene *AtC3H14* is the direct target of the MYB domain TF *MYB46* and a master switch for cell elongation in *Arabidopsis* [78, 79], which suggests that *HvC3H15* might have similar functions in regulating plant organogenesis in barley.

Given their sessile nature, terrestrial plants are often exposed to various adverse challenges such as drought, high salinity, extreme temperatures, and pathogen infection [80]. To adapt to such environmental conditions, plants have evolved complicated signal transduction pathways to perceive stress signals and coordinate plant growth and development [81]. Several regulatory components, including receptors, TFs, protein kinases, and plant hormones that simulate or repress other components [82], are needed to construct interconnected and sophisticated regulatory networks. To obtain preliminary insight into the potential functions of *HvC3H*s in response to environmental stress, the expression patterns of *HvC3H*s under cold, salt, and metal ion treatments were characterized based on publicly available RNA-seq datasets.

Soil salinization affects seed germination and crop growth and development and has become a major environmental barrier for agricultural production globally [83]. The expression of CCCH proteins has been reported to be induced by salt stress, indicating that CCCH proteins are involved in salt stress tolerance in plants [84]. For example, *Arabidopsis AtSZF1* and *AtSZF2*, two closely related *CCCH* zinc finger genes, negatively regulate the expression of salt-responsive genes and modulate tolerance to salt stress [85]. In rice, *OsTZF1* negatively regulates leaf senescence under salt conditions and confers stress tolerance by delaying stress-responsive phenotypes, possibly through post-transcriptional control of the RNA metabolism of salt stress-responsive genes [17]. In barley, *HvC3H9* showed the highest homology among these genes, and its expression was significantly induced under salt stress according to both RNA-seq and qRT-PCR analysis, suggesting that *HvC3H9* might have similar functions in response to salinity stress.

Recently, several lines of evidence from multiple approaches have shown that CCCH proteins directly increase the drought tolerance of plants [84]. In our study, *HvC3H18* was the most up-regulated gene at 1 h, 6 h, and 12 h under drought stress. Homology analysis revealed that its orthologous gene *OsC3H47* is involved in drought tolerance through its elevated sensitivity to ABA [25]. Another CCCH-tandem zinc finger protein *OsTZF5*, whose optimal BLAST hit was *HvC3H9* in barley, promoted both drought avoidance and drought tolerance in rice [26]. Several MBS *cis*-acting elements associated with drought inducibility within the promoter regions of *HvC3H18* and *HvC3H9* genes were predicted, suggesting that these genes might play a potential role in the response to drought stress.

Low temperature greatly limits the geographical distribution, growth, and development, as well as the yield, quality, and storage properties of crops [86]. However, few studies have examined the molecular regulation of *CCCH*s in response to environmental temperatures. In *Chrysanthemum*, *DgC3H1* enhances low temperature tolerance via regulation of the plant ROS system and the expression of downstream cold-related genes. However, the expression of *HvC3H27*, which is orthologous to *DgC3H1*, was not induced by cold stress for both RNA-seq and qRT-PCR analysis. Moreover, there was no *cis*-acting element, such as LTR, involved in low temperature responsiveness within the promoter region of *HvC3H27* [31]. These results indicate that *HvC3H27* might have functionally diversified in barley. Compared with the control, the expression of *HvC3H9* changed approximately 3.01-fold under cold treatment for RNA-seq. qRT-PCR analysis also revealed significant up-regulation at 1 h, 3 h, and 6 h. Its orthologous gene in switchgrass *PvC3H72* acts as an added signaling component by regulating the ICE1-CBF-COR regulon and ABA-responsive genes [30]. Overall, several candidate *CCCH* genes that could provide targets for subsequent genetic isolation and functional investigation in barley as well as in other cereal crops were identified. Experiments are needed to determine the functions of *HvC3H*s.

Genetic Variation and Haplotype Polymorphism of HvC3Hs during the Domestication of Barley

SNPs are the most common type of genomic variants in the genome [87]. They have become an increasingly invaluable and powerful molecular marker for the high-resolution genetic mapping of traits, linkage disequilibrium analysis, and association studies [88, 89]. High densities of SNPs also provide invaluable resources for genome mapping, the generation of ultra-high-density genetic maps, the identification of haplotypes near a locus of interest, and the map-based positional cloning of genes. The application of next-generation sequencing technologies that increase the quantity of DNA and reduce the cost of sequencing has facilitated the discovery of DNA polymorphisms *in silico* [90]. In recent years, SNPs derived from resequencing data have been used to investigate the genetic relationships between wild and domesticated barley [91–94]. Genetic variation in *HvC3H* genes in barley populations was determined using publicly released exome capture datasets [93]. We obtained a large number of high-confident *HvC3H*-related SNPs and InDels in 95 landraces and 51 wild barley accessions. Most of the variants were enriched within the introns, possibly because of weak selection on the non-coding regions. The mutations also displayed significant base bias of C/T (39.27%) and A/G (32.33%). The overall transition/transversion (Ts/Tv) ratio was 2.52, which indicated the greater importance of transitions compared with transversions.

We also characterized the evolutionary history and genetic divergence of *HvC3H* genes in wild and domesticated barley populations. Two genetically divergent populations associated with domestic versus wild barley rather than barley populations with different geographic origins were observed, which was consistent with the deep phylogenetic split between wild and domesticated barley of *HvmTERF*s [60]. The greater effect of human-driven selection on the genomic regions of *HvC3H*s relative to natural selection was confirmed by the genetic diversity analysis. The nucleotide diversity of *HvC3H*s displayed a significant reduction characteristic of a genetic bottleneck between domesticated and wild barley. The estimated reduction by ~34.94% from wild to domesticated barley was higher compared with that estimated for barley adaptive genes (22.5%) [95], housekeeping genes (20%) [86], and disease resistance genes (18.2%) [95]. The estimated reduction was also much higher when compared with that reported at the whole-genome level (reduction of 22%) [93]. We concluded that these genes have undergone a genetic bottleneck from wild barley to cultivated barley and might be domestication-related genes.

Haplotype reconstruction and characterization of the haplotype distribution can shed light on the differentiation of important genes and the processes underlying their evolution [96]. Haplotype networks indicated that the haplotype composition of the *HvC3H* family in wild barley was rich compared with that in cultivated barley, which indicated that initial human selection was focused on the retention of specific haplotypes by screening out a large number of undesirable haplotypes during the process of domestication. Domestication is a plant–animal coevolutionary process driven by the human demands for certain morphological and physical characteristics of crops [97]; this results in a severe genetic bottleneck that reduces the nucleotide diversity of alleles [98]. Wild barley populations possessed specific haplotypes that were absent in domesticated populations, which suggests that the genetic traits controlled by *HvC3H*s in domesticated barley could be enriched. Characterization of *HvC3H* haplotypes will not only provide new insight into the functional divergence between cultivated and wild barley but also help establish associations between genetic variants and important agronomic traits so that they can be used as molecular markers.

Conclusion

Despite the importance of *CCCH* genes in plant growth and development, the response to biotic and abiotic stress, and disease resistance, the precise roles of *CCCH* gene family members in barley have not yet been elucidated. Here, our genome-wide identification and characterization of *HvC3H* genes revealed the physical-chemical properties, phylogeny, intron–exon structure, and expansion patterns of these genes. The results of the expression profiling analysis suggest that *HvC3H* members might be associated with multiple physiological, metabolic, and developmental processes, especially in response to various types of biotic stress. The population structure based on the most recently released exome capture sequencing data revealed a deep phylogenetic split in the *HvC3H*s between wild and domesticated barley. The nucleotide and haplotype diversity of most *HvC3H*s indicated that these genes have undergone a severe genetic bottleneck during the transition from wild relatives to domesticated barley populations. Overall, these findings will aid future studies examining the evolutionary history of *HvC3H*s as well as functional studies of candidate *HvC3H* genes for molecular breeding in barley.

Abbreviations

ABA, Abscisic Acid; ABRE, Abscisic Acid Responsive Elements; ARE, Anaerobic Induction Elements; BLAST, Basic Local Alignment Search Tool; ERE, Ethylene Responsive Element; EST, Expressed Sequence Tag; FPKM, Fragments Per Kilobase per Million; GA, Gibberellins GRAVY, Grand Average of Hydropathicity; GSDS, Gene Structure Display Sever; GTF, Gene Transfer Format; HMM, Hidden Markov Model; HvC3H, Barley CCCH genes; IBSC, International Barley Sequencing Consortium; II, Instability Index; InDel, Insertion-Deletion; iTOL, Interactive Tree of Life; Ka, Non-synonymous substitution; Ks, Synonymous substitution; LTR, Low-Temperature Responsive; MAF, Minor Allele Frequency; MBS, Myeloblastosis Binding Site; MEME, Multiple Expectation Maximization for Motif Elicitation; MW, Molecular Weight; MYA, Million Years Ago; NCBI-CDD, National Center for Biotechnology Information - Conserved Domains Database; NJ, Neighbor-Joining; PCA, Principal Component Analysis; pI, Isoelectric Point; qRT-PCR, Quantitative Real-time PCR; ROS, Reactive Oxygen Species; SMART, Simple Modular Architecture Research Tool; SRA, Sequence Read Archive; SNP, Single Nucleotide Polymorphism; TF, Transcription Factor; π, Nucleotide diversity.

Declarations

Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author Contributions

YL and LC designed and supervised the project. QA and WP performed the data analysis, WP and YZ performed the experiments. QA and LC drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was mainly funded by the National Natural Science Foundation of China (Grant No. 32060458), Jiangxi Natural Science Foundation (Grant No. 20202BAB215002), and partially supported by the Science and Technology Research Project of Jiangxi Provincial Department of Education (Grant No. GJJ180241). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Guo A-Y, Chen X, Gao G, Zhang H, Zhu Q-H, Liu X-C, Zhong Y-F, Gu X, He K, Luo J: **PlantTFDB: a comprehensive plant transcription factor database**. *Nucleic Acids Research* 2008, **36**:D966-D969.
2. Liu LS, White MJ, MacRae TH: **Transcription factors and their genes in higher plants - Functional domains, evolution and regulation**. *European Journal of Biochemistry* 1999, **262**(2):247–257.
3. Hall TMT: **Multiple modes of RNA recognition by zinc finger proteins**. *Current Opinion in Structural Biology* 2005, **15**(3):367–373.
4. Moore M, Ullman C: **Recent developments in the engineering of zinc finger proteins**. *Briefings in Functional Genomics & Proteomics* 2003, **1**(4):342–355.
5. Lijavetzky D, Carbonero P, Vicente-Carbajosa J: **Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families**. *BMC Evol Biol* 2003, **3**:17.

6. Nakano T, Suzuki K, Fujimura T, Shinshi H: **Genome-wide analysis of the ERF gene family in Arabidopsis and rice**. *Plant Physiology* 2006, **140**(2):411–432.
7. Arnaud D, Dejardin A, Leple JC, Lesage-Descauses MC, Pilate G: **Genome-wide analysis of LIM gene family in Populus trichocarpa, Arabidopsis thaliana, and Oryza sativa**. *DNA Research* 2007, **14**(3):103–116.
8. Kosarev P, Mayer KF, Hardtke CS: **Evaluation and classification of RING-finger domains encoded by the Arabidopsis genome**. *Genome Biol* 2002, **3**(4):Research0016.
9. Zhang Y, Wang L: **The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants**. *BMC Evol Biol* 2005, **5**:1.
10. Bai C, Tolias PP: **Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein**. *Mol Cell Biol* 1996, **16**(12):6661–6667.
11. Wang D, Guo YH, Wu CG, Yang GD, Li YY, Zheng CC: **Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice**. *Bmc Genomics* 2008, **9**.
12. Berg JM, Shi YG: **The galvanization of biology: A growing appreciation for the roles of zinc**. *Science* 1996, **271**(5252):1081–1085.
13. Peng XJ, Zhao Y, Cao JG, Zhang W, Jiang HY, Li XY, Ma Q, Zhu SW, Cheng BJ: **CCCH-Type Zinc Finger Family in Maize: Genome-Wide Identification, Classification and Expression Profiling under Abscisic Acid and Drought Treatments**. *Plos One* 2012, **7**(7).
14. Yan Z, Jia J, Yan X, Shi H, Han Y: **Arabidopsis KHZ1 and KHZ2, two novel non-tandem CCCH zinc-finger and K-homolog domain proteins, have redundant roles in the regulation of flowering and senescence**. *Plant molecular biology* 2017, **95**(6):549–565.
15. Seok HY, Bae H, Kim T, Mehdi SMM, Nguyen LV, Lee SY, Moon YH: **Non-TZF Protein AtC3H59/ZFWD3 Is Involved in Seed Germination, Seedling Development, and Seed Development, Interacting with PPPDE Family Protein Desi1 in Arabidopsis**. *International Journal of Molecular Sciences* 2021, **22**(9).
16. Kong ZS, Li MN, Yang WQ, Xu WY, Xue YB: **A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice**. *Plant Physiology* 2006, **141**(4):1376–1388.
17. Jan A, Maruyama K, Todaka D, Kidokoro S, Abo M, Yoshimura E, Shinozaki K, Nakashima K, Yamaguchi-Shinozaki K: **OsTZF1, a CCCH-Tandem Zinc Finger Protein, Confers Delayed Senescence and Stress Tolerance in Rice by Regulating Stress-Related Genes**. *Plant Physiology* 2013, **161**(3):1202–1216.
18. Chen Y, Sun AJ, Wang M, Zhu Z, Ouwerkerk PBF: **Functions of the CCCH type zinc finger protein OsGZF1 in regulation of the seed storage protein GluB-1 from rice**. *Plant Molecular Biology* 2014, **84**(6):621–634.
19. Li QT, Lu X, Song QX, Chen HW, Wei W, Tao JJ, Bian XH, Shen M, Ma BA, Zhang WK *et al*: **Selection for a Zinc-Finger Protein Contributes to Seed Oil Increase during Soybean Domestication**. *Plant Physiology* 2017, **173**(4):2208–2224.
20. Lu L, Wei W, Li QT, Bian XH, Lu X, Hu Y, Cheng T, Wang ZY, Jin M, Tao JJ *et al*: **A transcriptional regulatory module controls lipid accumulation in soybean**. *New Phytologist* 2021, **231**(2):661–678.
21. Bogamuwa S, Jang JC: **The Arabidopsis tandem CCCH zinc finger proteins AtTZF4, 5 and 6 are involved in light-, abscisic acid- and gibberellic acid- mediated regulation of seed germination**. *Plant Cell and Environment* 2013, **36**(8):1507–1519.
22. Wang L, Xu Y, Zhang C, Ma Q, Joo SH, Kim SK, Xu ZH, Chong K: **OsLIC, a Novel CCCH-Type Zinc Finger Protein with Transcription Activation, Mediates Rice Architecture via Brassinosteroids Signaling**. *Plos One* 2008, **3**(10).
23. Xie ZI, Yu GH, Lei SS, Zhang CC, Xu B, Huang BR: **CCCH protein-PvCCCH69 acted as a repressor for leaf senescence through suppressing ABA-signaling pathway**. *Horticulture Research* 2021, **8**(1).
24. Seong SY, Shim JS, Bang SW, Kim JK: **Overexpression of OsC3H10, a CCCH-Zinc Finger, Improves Drought Tolerance in Rice by Regulating Stress-Related Genes**. *Plants-Basel* 2020, **9**(10).
25. Wang WY, Liu BH, Xu MY, Jamil M, Wang GP: **ABA-induced CCCH tandem zinc finger protein OsC3H47 decreases ABA sensitivity and promotes drought tolerance in Oryza sativa**. *Biochemical and Biophysical Research Communications* 2015, **464**(1):33–37.

26. Selvaraj MG, Jan A, Ishizaki T, Valencia M, Dedicova B, Maruyama K, Ogata T, Todaka D, Yamaguchi-Shinozaki K, Nakashima K *et al*: **Expression of the CCCH-tandem zinc finger protein gene OsTZF5 under a stress-inducible promoter mitigates the effect of drought stress on rice grain yield under field conditions**. *Plant Biotechnology Journal* 2020, **18**(8):1711–1721.
27. Zhou T, Yang XY, Wang LC, Xu J, Zhang XL: **GhTZF1 regulates drought stress responses and delays leaf senescence by inhibiting reactive oxygen species accumulation in transgenic Arabidopsis**. *Plant Molecular Biology* 2014, **85**(1-2):163–177.
28. Han GL, Wang MJ, Yuan F, Sui N, Song J, Wang BS: **The CCCH zinc finger protein gene AtZFP1 improves salt resistance in Arabidopsis thaliana**. *Plant Molecular Biology* 2014, **86**(3):237–253.
29. Seok HY, Nguyen LV, Park HY, Tarte VN, Ha J, Lee SY, Moon YH: **Arabidopsis non-TZF gene AtC3H17 functions as a positive regulator in salt stress response**. *Biochemical and Biophysical Research Communications* 2018, **498**(4):954–959.
30. Xie ZN, Lin WJ, Yu GH, Cheng Q, Xu B, Huang BR: **Improved cold tolerance in switchgrass by a novel CCCH-type zinc finger transcription factor gene, PvC3H72, associated with ICE1-CBF-COR regulon and ABA-responsive genes**. *Biotechnology for Biofuels* 2019, **12**(1).
31. Bai HR, Lin P, Li X, Liao XQ, Wan LH, Yang XH, Luo YC, Zhang L, Zhang F, Liu SL *et al*: **DgC3H1, a CCCH zinc finger protein gene, confers cold tolerance in transgenic chrysanthemum**. *Scientia Horticulturae* 2021, **281**.
32. Deng HQ, Liu HB, Li XH, Xiao JH, Wang SP: **A CCCH-Type Zinc Finger Nucleic Acid-Binding Protein Quantitatively Confers Resistance against Rice Bacterial Blight Disease**. *Plant Physiology* 2012, **158**(2):876–889.
33. Yang Z, Wu YR, Li Y, Ling HQ, Chu CC: **OsMT1a, a type 1 metallothionein, plays the pivotal role in zinc homeostasis and drought tolerance in rice**. *Plant Molecular Biology* 2009, **70**(1-2):219–229.
34. Hunt AG, Xu RQ, Addepalli B, Rao S, Forbes KP, Meeks LR, Xing DH, Mo M, Zhao HW, Bandyopadhyay A *et al*: **Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling**. *Bmc Genomics* 2008, **9**.
35. Chai GH, Hu RB, Zhang DY, Qi G, Zuo R, Cao YP, Chen P, Kong YZ, Zhou GK: **Comprehensive analysis of CCCH zinc finger family in poplar (Populus trichocarpa)**. *Bmc Genomics* 2012, **13**.
36. Xu RR: **Genome-wide analysis and identification of stress-responsive genes of the CCCH zinc finger family in Solanum lycopersicum**. *Molecular Genetics and Genomics* 2014, **289**(5):965–979.
37. Zhang CQ, Zhang HM, Zhao Y, Jiang HY, Zhu SW, Cheng BJ, Xiang Y: **Genome-wide analysis of the CCCH zinc finger gene family in Medicago truncatula**. *Plant Cell Reports* 2013, **32**(10):1543–1555.
38. Wang XL, Zhong Y, Cheng ZM: **Evolution and Expression Analysis of the CCCH Zinc Finger Gene Family in Vitis vinifera**. *Plant Genome* 2014, **7**(3).
39. Liu SR, Khan MRG, Li YP, Zhang JZ, Hu CE: **Comprehensive analysis of CCCH-type zinc finger gene family in citrus (Clementine mandarin) by genome-wide characterization**. *Molecular Genetics and Genomics* 2014, **289**(5):855–872.
40. Yuan SX, Xu B, Zhang J, Xie ZN, Cheng Q, Yang ZM, Cai QS, Huang BR: **Comprehensive analysis of CCCH-type zinc finger family genes facilitates functional gene discovery and reflects recent allopolyploidization event in tetraploid switchgrass**. *Bmc Genomics* 2015, **16**.
41. Pi BY, He XH, Ruan Y, Jang JC, Huang Y: **Genome-wide analysis and stress-responsive expression of CCCH zinc finger family genes in Brassica rapa**. *Bmc Plant Biology* 2018, **18**.
42. Zhang Q, Zhang WJ, Yin ZG, Li WJ, Zhao HH, Zhang S, Zhuang L, Wang YX, Zhang WH, Du JD: **Genome- and Transcriptome-Wide Identification of C3Hs in Common Bean (Phaseolus vulgarisL.) and Structural and Expression-Based Analyses of Their Functions During the Sprout Stage Under Salt-Stress Conditions**. *Frontiers in Genetics* 2020, **11**.
43. Chen F, Liu HL, Wang K, Gao YM, Wu M, Xiang Y: **Identification of CCCH Zinc Finger Proteins Family in Moso Bamboo (Phyllostachys edulis), and PeC3H74 Confers Drought Tolerance to Transgenic Plants**. *Frontiers in Plant Science* 2020, **11**.
44. Cheng XR, Cao JJ, Gao C, Gao W, Yan SN, Yao H, Xu KL, Liu X, Xu DM, Pan X *et al*: **Identification of the wheat C3H gene family and expression analysis of candidates associated with seed dormancy and germination**. *Plant Physiology and*

Biochemistry 2020, **156**:524–537.

45. Hu X, Zuo JF: **The CCCH zinc finger family of soybean (Glycine max L.): genome-wide identification, expression, domestication, GWAS and haplotype analysis**. *Bmc Genomics* 2021, **22**(1).
46. Mayer KFX, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, Matsumoto T, Sato K, Schulman A, Muehlbauer GJ *et al*: **A physical, genetic and functional sequence assembly of the barley genome**. *Nature* 2012, **491**(7426):711-+.
47. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J *et al*: **A chromosome conformation capture ordered sequence of the barley genome**. *Nature* 2017, **544**(7651):426-+.
48. Monat C, Padmarasu S, Lux T, Wicker T, Gundlach H, Himmelbach A, Ens J, Li CD, Muehlbauer GJ, Schulman AH *et al*: **TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools**. *Genome Biology* 2019, **20**(1).
49. Hudson BP, Martinez-Yamout MA, Dyson HJ, Wright PE: **Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d**. *Nature Structural & Molecular Biology* 2004, **11**(3):257-264.
50. Kramer S, Kimblin NC, Carrington M: **Genome-wide in silico screen for CCCH-type zinc finger proteins of Trypanosoma brucei, Trypanosoma cruzi and Leishmania major**. *Bmc Genomics* 2010, **11**.
51. Wang D, Guo Y, Wu C, Yang G, Li Y, Zheng C: **Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice**. *BMC Genomics* 2008, **9**:44.
52. Ke YZ, Wu YW, Zhou HJ, Chen P, Wang MM, Liu MM, Li PF, Yang J, Li JN, Du H: **Genome-wide survey of the bHLH super gene family in Brassica napus**. *Bmc Plant Biology* 2020, **20**(1).
53. Moore RC, Purugganan MD: **The early stages of duplicate gene evolution**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(26):15682–15687.
54. Panchy N, Lehti-Shiu M, Shiu SH: **Evolution of Gene Duplication in Plants**. *Plant Physiol* 2016, **171**(4):2294–2316.
55. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**(5494):1151–1155.
56. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J *et al*: **Shifting the limits in wheat research and breeding using a fully annotated reference genome**. *Science* 2018, **361**(6403):661-+.
57. Zhang ZL, Tong T, Fang YX, Zheng JJ, Zhang X, Niu CY, Li J, Zhang XQ, Xue DW: **Genome-Wide Identification of Barley ABC Genes and Their Expression in Response to Abiotic Stress Treatment**. *Plants-Basel* 2020, **9**(10).
58. Bogamuwa SP, Jang JC: **Tandem CCCH Zinc Finger Proteins in Plant Growth, Development and Stress Response**. *Plant and Cell Physiology* 2014, **55**(8):1367–1375.
59. Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, Ens J, Gundlach H, Boston LB, Tulpova Z *et al*: **Long-read sequence assembly: a technical evaluation in barley**. *Plant Cell* 2021, **33**(6):1888–1906.
60. Li T, Pan W, Yuan Y, Liu Y, Li Y, Wu X, Wang F, Cui L: **Identification, Characterization, and Expression Profile Analysis of the mTERF Gene Family and Its Role in the Response to Abiotic Stress in Barley (Hordeum vulgare L.)**. *Frontiers in plant science* 2021, **12**:684619.
61. Cai KF, Zeng FR, Wang JM, Zhang GP: **Identification and characterization of HAK/KUP/KT potassium transporter gene family in barley and their expression under abiotic stress**. *Bmc Genomics* 2021, **22**(1).
62. Zhong X, Feng X, Li Y, Guzman C, Lin N, Xu Q, Zhang Y, Tang H, Qi P, Deng M *et al*: **Genome-wide identification of bZIP transcription factor genes related to starch synthesis in barley (Hordeum vulgare L.)**. *Genome* 2021:1–14.
63. Dai GY, Chen DK, Sun YP, Liang WY, Liu Y, Huang LQ, Li YK, He JF, Yao N: **The Arabidopsis KH-domain protein FLOWERING LOCUS Y delays flowering by upregulating FLOWERING LOCUS C family members**. *Plant Cell Reports* 2020, **39**(12):1705–1717.
64. Lee KC, Jang YH, Kim SK, Park HY, Thu MP, Lee JH, Kim JK: **RRM domain of Arabidopsis splicing factor SF1 is important for pre-mRNA splicing of a specific set of genes**. *Plant Cell Reports* 2017, **36**(7):1083–1095.
65. Guo LQ, Nezames CD, Sheng LX, Deng XW, Wei N: **Cullin-RING Ubiquitin Ligase Family in Plant Abiotic Stress Pathways**. *Journal of Integrative Plant Biology* 2013, **55**(1):21–30.
66. Sankoff D: **Gene and genome duplication**. *Current opinion in genetics & development* 2001, **11**(6):681–684.

67. Soltis PS, Marchant DB, Van de Peer Y, Soltis DE: **Polyploidy and genome evolution in plants**. *Current opinion in genetics & development* 2015, **35**:119–125.
68. Shi Q, Zhang YY, To VT, Shi J, Zhang DB, Cai WG: **Genome-wide characterization and expression analyses of the auxin/indole-3-acetic acid (Aux/IAA) gene family in barley (Hordeum vulgare L.)**. *Scientific Reports* 2020, **10**(1).
69. To VT, Shi Q, Zhang YY, Shi J, Shen CQ, Zhang DB, Cai WG: **Genome-Wide Analysis of the GRAS Gene Family in Barley (Hordeum vulgareL.)**. *Genes* 2020, **11**(5).
70. Ke QL, Tao WJ, Li TT, Pan WQ, Chen XY, Wu XY, Nie XJ, Cui LC: **Genome-wide Identification, Evolution and Expression Analysis of Basic Helix-loop-helix (bHLH) Gene Family in Barley (Hordeum vulgare L.)**. *Current Genomics* 2020, **21**(8):624–644.
71. Guo BJ, Li Y, Wang S, Li DF, Lv C, Xu RG: **Characterization of the Nitrate Transporter gene family and functional identification of HvNRT2.1 in barley (Hordeum vulgare L.)**. *Plos One* 2020, **15**(4).
72. Yu SH, Sun QG, Wu JX, Zhao PC, Sun YM, Guo ZF: **Genome-Wide Identification and Characterization of Short-Chain Dehydrogenase/Reductase (SDR) Gene Family in Medicago truncatula**. *International Journal of Molecular Sciences* 2021, **22**(17).
73. Jang JC: **Arginine-rich motif-tandem CCCH zinc finger proteins in plant stress responses and post-transcriptional regulation of gene expression**. *Plant Science* 2016, **252**:118–124.
74. Mockler TC, Yu XH, Shalitin D, Parikh D, Michael TP, Liou J, Huang J, Smith Z, Alonso JM, Ecker JR *et al*: **Regulation of flowering time in Arabidopsis by K homology domain proteins**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(34):12759–12764.
75. Ripoll JJ, Rodriguez-Cazorla E, Gonzalez-Reig S, Andujar A, Alonso-Cantabrana H, Perez-Amador MA, Carbonell J, Martinez-Laborda A, Vera A: **Antagonistic interactions between Arabidopsis K-homology domain genes uncover PEPPER as a positive regulator of the central floral repressor FLOWERING LOCUS C**. *Developmental Biology* 2009, **333**(2):251–262.
76. Yan ZY, Jia JH, Yan XY, Shi HY, Han YZ: **Arabidopsis KHZ1 and KHZ2, two novel non-tandem CCCH zinc-finger and K-homolog domain proteins, have redundant roles in the regulation of flowering and senescence**. *Plant Molecular Biology* 2017, **95**(6):549–565.
77. Yan Z, Shi H, Liu Y, Jing M, Han Y: **KHZ1 and KHZ2, novel members of the autonomous pathway, repress the splicing efficiency of FLC pre-mRNA in Arabidopsis**. *J Exp Bot* 2020, **71**(4):1375–1386.
78. Kim WC, Kim JY, Ko JH, Kang H, Kim J, Han KH: **AtC3H14, a plant-specific tandem CCCH zinc-finger protein, binds to its target mRNAs in a sequence-specific manner and affects cell elongation in Arabidopsis thaliana**. *Plant Journal* 2014, **80**(5):772–784.
79. Kim WC, Ko JH, Han KH: **Identification of a cis-acting regulatory motif recognized by MYB46, a master transcriptional regulator of secondary wall biosynthesis**. *Plant Molecular Biology* 2012, **78**(4-5):489–501.
80. Zhu JK: **Abiotic Stress Signaling and Responses in Plants**. *Cell* 2016, **167**(2):313–324.
81. Verma V, Ravindran P, Kumar PP: **Plant hormone-mediated regulation of stress responses**. *BMC Plant Biol* 2016, **16**:86.
82. Chen JT, Heidari P: **Cell Signaling in Model Plants 2.0**. *International Journal of Molecular Sciences* 2021, **22**(15).
83. Song J, Wang BS: **Using euhalophytes to understand salt tolerance and to develop saline agriculture: Suaeda salsa as a promising model**. *Annals of Botany* 2015, **115**(3):541–553.
84. Han GL, Qiao ZQ, Li YX, Wang CF, Wang BS: **The Roles of CCCH Zinc-Finger Proteins in Plant Abiotic Stress Tolerance**. *International Journal of Molecular Sciences* 2021, **22**(15).
85. Sun JQ, Jiang HL, Xu YX, Li HM, Wu XY, Xie Q, Li CY: **The CCCH-type zinc finger proteins AtSZF1 and AtSZF2 regulate salt stress responses in Arabidopsis**. *Plant and Cell Physiology* 2007, **48**(8):1148–1158.
86. Ding YL, Shi YT, Yang SH: **Molecular Regulation of Plant Responses to Environmental Temperatures**. *Molecular Plant* 2020, **13**(4):544–564.
87. Lai KT, Duran C, Berkman PJ, Lorenc MT, Stiller J, Manoli S, Hayden MJ, Forrest KL, Fleury D, Baumann U *et al*: **Single nucleotide polymorphism discovery from wheat next-generation sequence data**. *Plant Biotechnology Journal* 2012,

10(6):743–749.

88. Duran C, Eales D, Marshall D, Imelfort M, Stiller J, Berkman PJ, Clark T, McKenzie M, Appleby N, Batley J *et al*: **Future tools for association mapping in crop plants**. *Genome* 2010, **53**(11):1017–1023.
89. Jannink JL, Lorenz AJ, Iwata H: **Genomic selection in plant breeding: from theory to practice**. *Briefings in Functional Genomics* 2010, **9**(2):166–177.
90. Duran C, Appleby N, Vardy M, Imelfort M, Edwards D, Batley J: **Single nucleotide polymorphism discovery in barley using autoSNPdb**. *Plant Biotechnology Journal* 2009, **7**(4):326–333.
91. Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M *et al*: **Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond**. *Plant Journal* 2013, **76**(3):494–505.
92. Pankin A, Altmuller J, Becker C, von Korff M: **Targeted resequencing reveals genomic signatures of barley domestication**. *New Phytologist* 2018, **218**(3):1247–1259.
93. Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, Bayer M, Milne I, Marshall-Griffiths T, Heinen S *et al*: **Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation**. *Nature Genetics* 2016, **48**(9):1024-+.
94. Zeng XQ, Guo Y, Xu QJ, Mascher M, Guo GG, Li SC, Mao LK, Liu QF, Xia ZF, Zhou JH *et al*: **Origin and evolution of qingke barley in Tibet**. *Nature Communications* 2018, **9**.
95. Fu YB: **Population-based resequencing analysis of wild and cultivated barley revealed weak domestication signal of selection and bottleneck in the Rrs2 scald resistance gene region**. *Genome* 2012, **55**(2):93–104.
96. Jan HU, Guan M, Yao M, Liu W, Wei DY, Abbadi A, Zheng M, He X, Chen H, Guan CY *et al*: **Genome-wide haplotype analysis improves trait predictions in Brassica napus hybrids**. *Plant Science* 2019, **283**:157–164.
97. Purugganan MD: **Evolutionary Insights into the Nature of Plant Domestication**. *Current Biology* 2019, **29**(14):R705-R714.
98. Haas M, Himmelbach A, Mascher M: **The contribution of cis- and trans-acting variants to gene regulation in wild and domesticated barley under cold stress and control conditions**. *J Exp Bot* 2020, **71**(9):2573–2584.
99. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: **MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms**. *Molecular Biology and Evolution* 2018, **35**(6):1547–1549.
100. Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2015, **31**(8):1296–1297.
101. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching**. *Nucleic Acids Research* 2009, **37**:W202-W208.
102. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH: **Synteny and collinearity in plant genomes**. *Science* 2008, **320**(5875):486–488.
103. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments**. *Nucleic Acids Research* 2006, **34**:W609-W612.
104. Wang Y, Wang QQ, Zhao Y, Han GM, Zhu SW: **Systematic analysis of maize class III peroxidase gene family reveals a conserved subfamily involved in abiotic stress response**. *Gene* 2015, **566**(1):95–108.
105. Xu L, Dong ZB, Fang L, Luo YJ, Wei ZY, Guo HL, Zhang GQ, Gu YQ, Coleman-Derr D, Xia QY *et al*: **OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species**. *Nucleic Acids Research* 2019, **47**(W1):W52-W58.
106. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method**. *Methods* 2001, **25**(4):402–408.
107. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114–2120.
108. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: **The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome Research*

2010, **20**(9):1297–1303.

109. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3**. *Fly* 2012, **6**(2):80–92.

110. Leigh JW, Bryant D: **POPART: full-feature software for haplotype network construction**. *Methods in Ecology and Evolution* 2015, **6**(9):1110–1116.

Figures

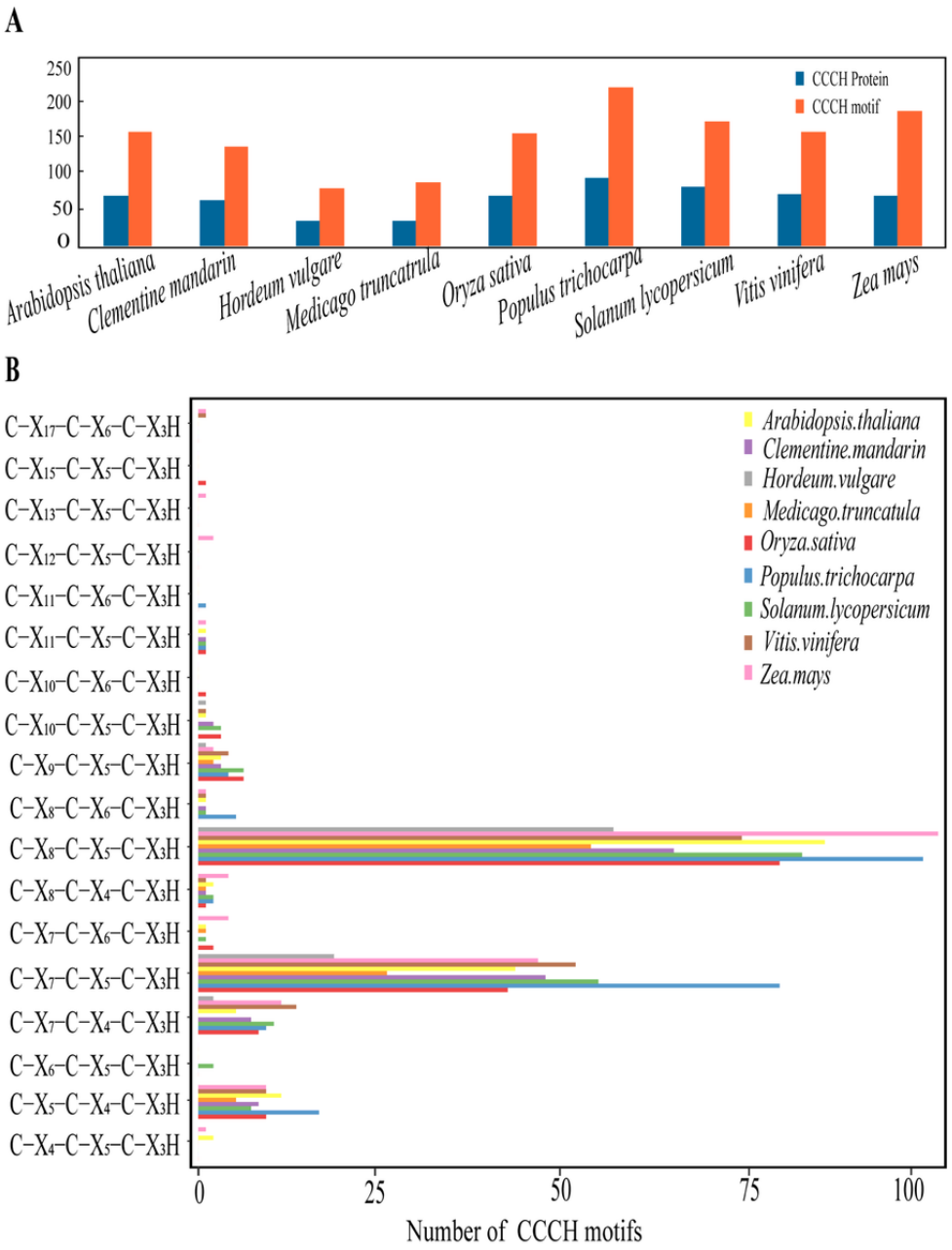

Figure 1

Characterization of CCCH-type zinc finger proteins from nine representative plant species. (A) The number of CCCH proteins and CCCH motifs identified in Arabidopsis thaliana, Clementine mandarin, Hordeum vulgare, Medicago truncatula, Oryza sativa, Populus trichocarpa, Solanum lycopersicum, Vitis vinifera and Zea mays. (B) The number of each type of CCCH motifs in

Arabidopsis thaliana, Clementine mandarin, Hordeum vulgare, Medicago truncatula, Oryza sativa, Populus trichocarpa, Solanum lycopersicum, Vitis vinifera and Zea mays nine plant species.

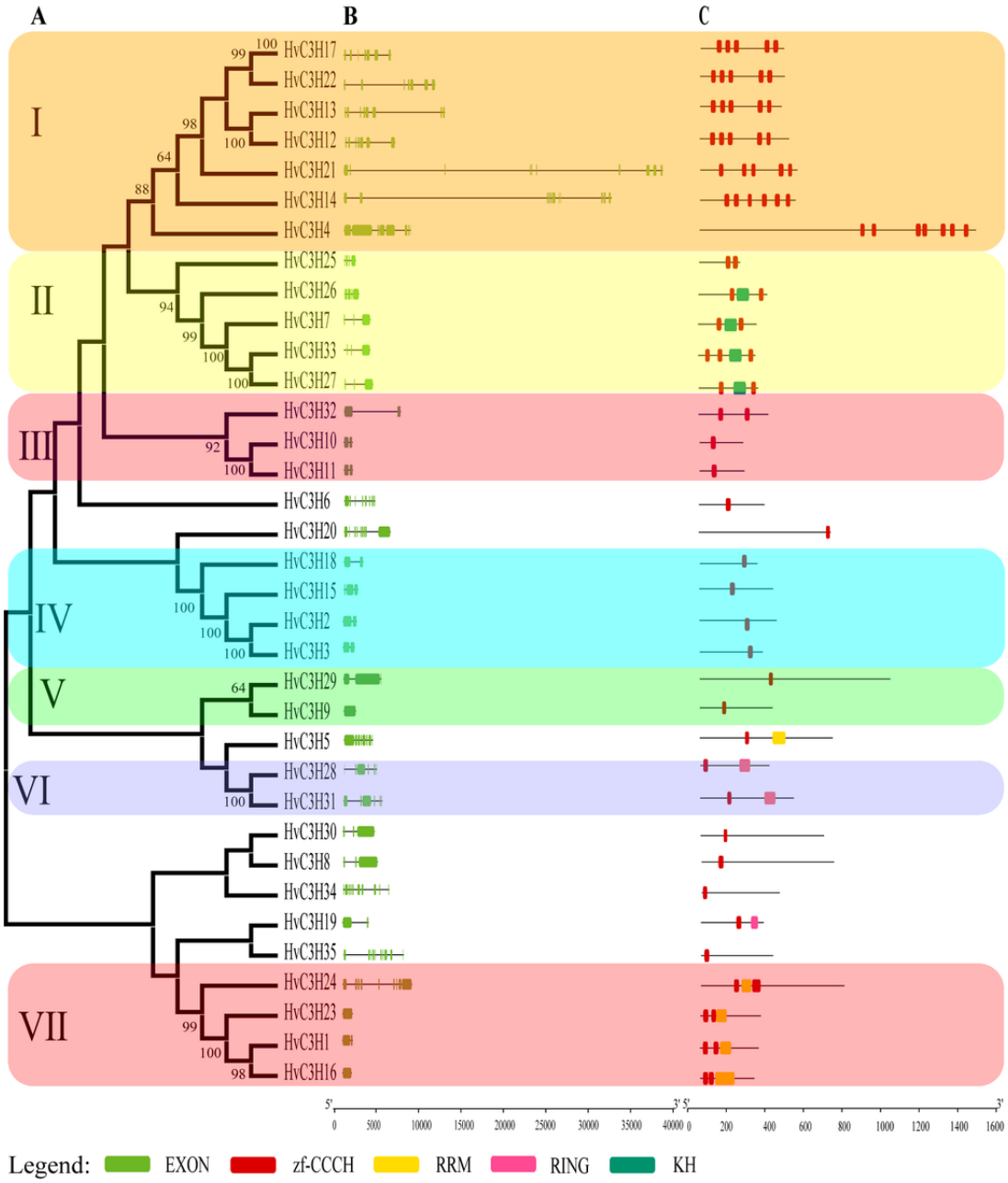

Figure 2

Phylogenetic relationships, gene structure, and motif compositions of HvC3Hs. The following parts are shown from left to right. (A) The amino acid sequences of the 35 barley CCCH proteins were aligned with ClustalX v2.1 and the phylogenetic tree was constructed using the neighbor-joining method in MEGA X software with 1000 bootstrap replications. The percentage bootstrap scores higher than 60% are indicated on the nodes. The tree shows the seven major phylogenetic subfamilies (left column, numbered I-VII and marked with different alternating tones of different colors to clarify subfamily identification easier). (B) Intron-exon organizations of barley CCCH genes. The introns and exons are represented by the black lines and green boxes, respectively. (C) Schematic structure of the CCCH protein motifs identified in barley. Different motifs are indicated by different color boxes at the bottom.

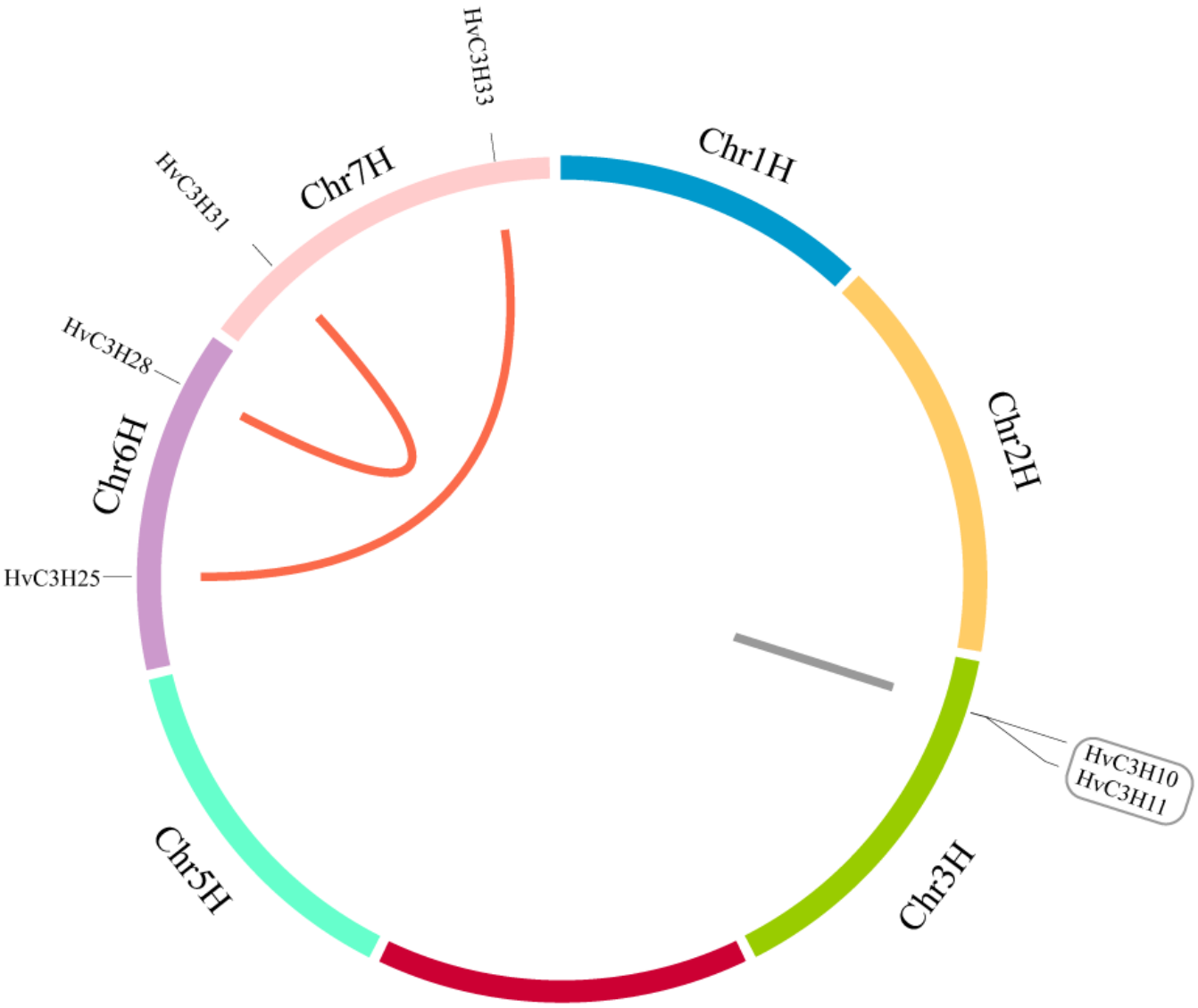

Figure 3

Chromosomal location and gene duplication of HvC3Hs. The segmental duplicated gene pairs are connected by curved lines, and the tandem duplicated genes are highlighted with black boxes.

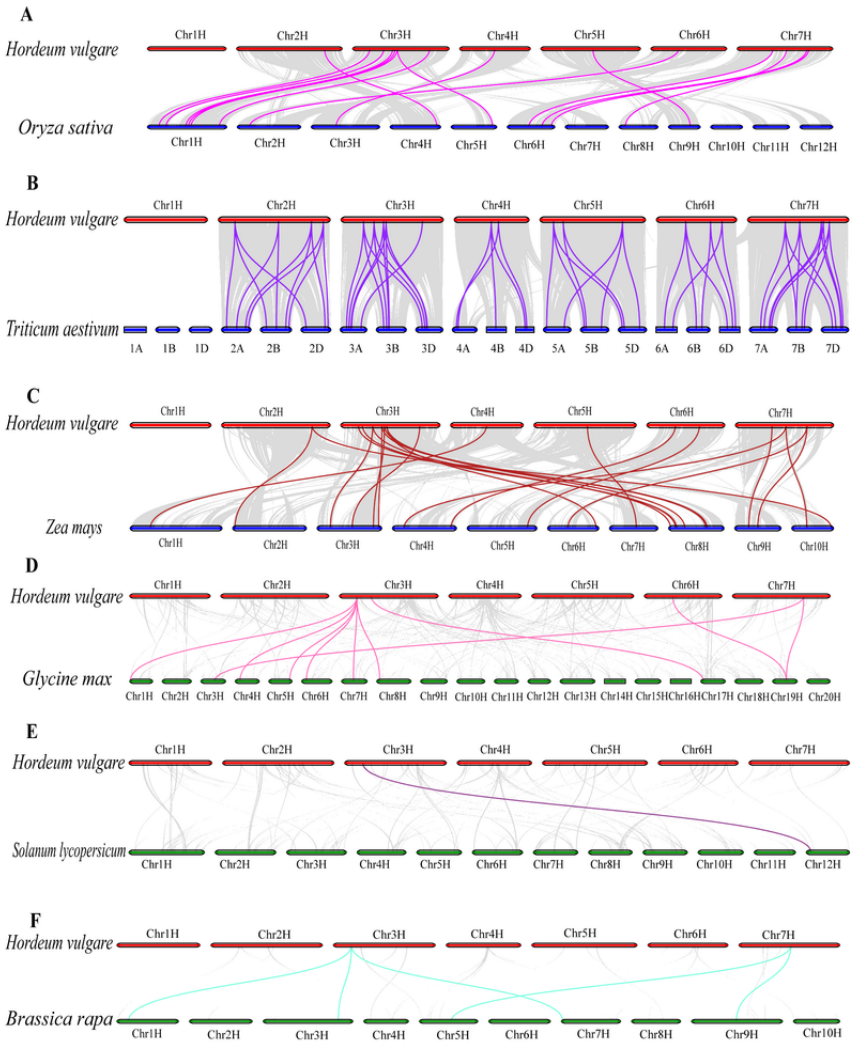

Figure 4

Synteny relationships analysis of HvC3Hs between barley and three Monocotyledons, three Dicotyledons. (A) Oryza sativa. (B) Triticum aestivum. (C) Zea mays. (D) Glycine max. (E) Solanum lycopersicum. (F) Brassica rapa.

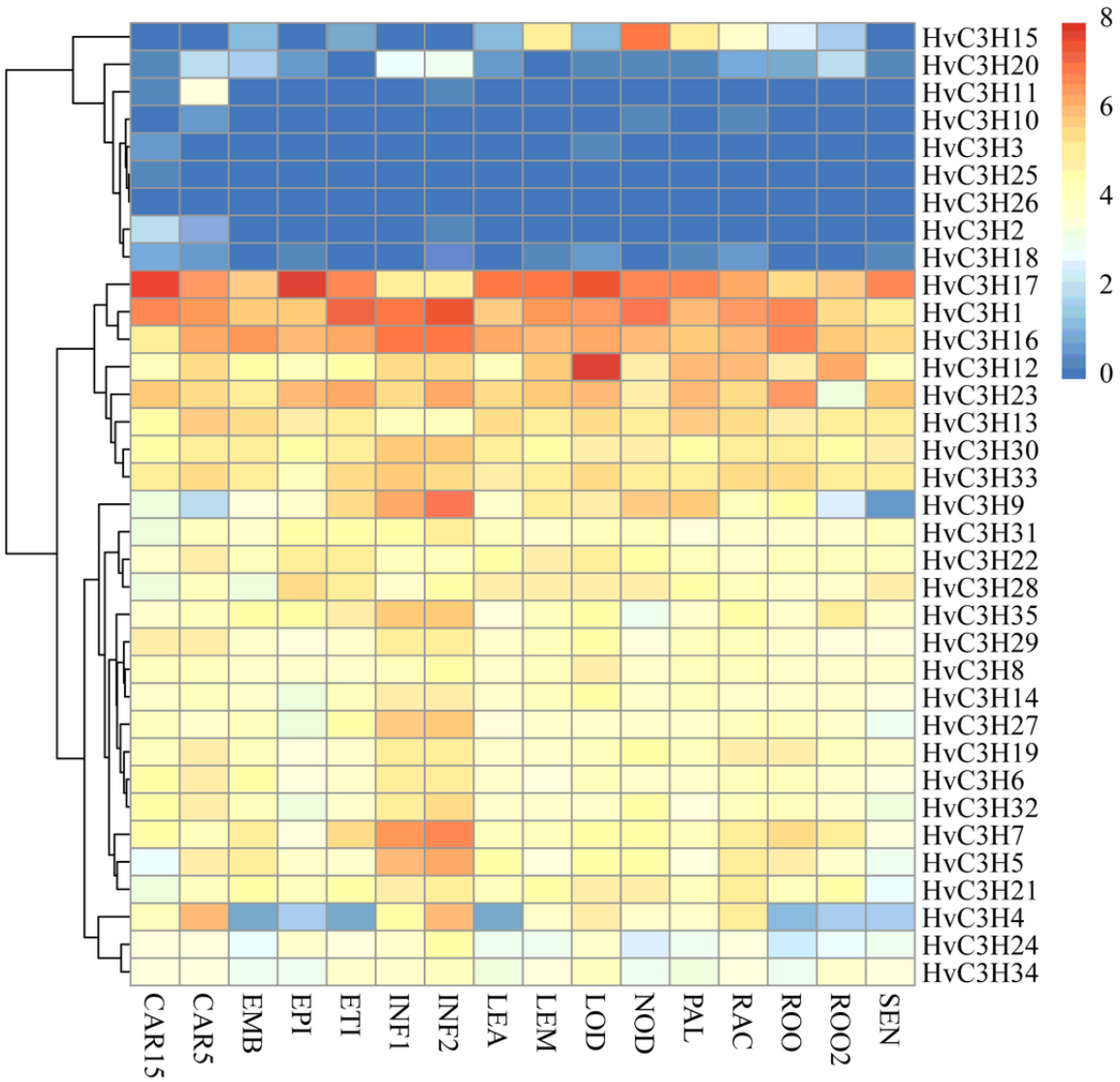

Figure 5

The spatiotemporal expression profile of HvC3H genes at different tissues or stage of barley. FPKM values were normalized by log2(FPKM+1) transformation to display the heatmap color scores. CAR15: bracts removed grains at 15DPA; CAR5: bracts removed grains at 5DPA; EMB: embryos dissected from 4 d-old germinating grains; EPI: epidermis with 4 weeks old; ETI: etiolated from 10-day old seedling; INF1: young inflorescences with 5 mm; INF2: young inflorescences with 1–1.5 cm; LEA: shoot with the size of 10 cm from the seedlings; LEM: lemma with 6 weeks after anthesis; LOD: lodicule with 6 weeks after anthesis; NOD: developing tillers at six-leaf stage; PAL: 6-week old palea; RAC: rachis with 5 weeks after anthesis; ROO2: root from 4-week old seedlings; ROO: Roots from the seedlings at 10 cm shoot stage; SEN: senescing leaf.

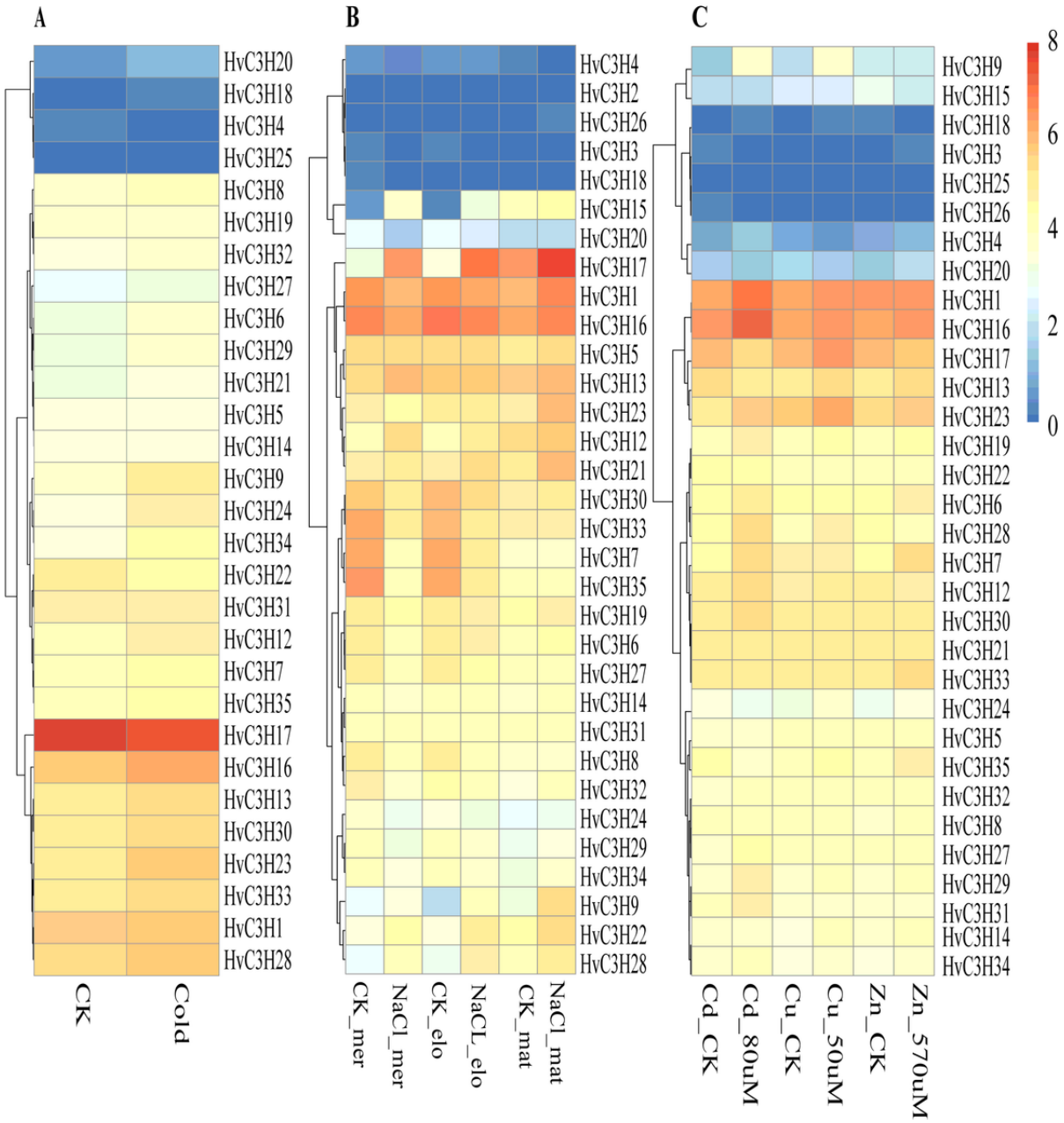

Figure 6

Expression profiles of HvC3H genes under five stress conditions. (A) cold stress treatment. From left to right: control check (CK) and cold stress. (B) salt stress. From left to right: CK in the meristematic zone, salt stress in the meristematic zone, CK in the elongation zone, salt stress in the elongation zone, CK in the maturation zone and salt stress in the maturation zone. (C) cadmium, copper and zinc stress. (C) cadmium, copper and zinc stress. CK and cadmium stress, CK and copper stress, and CK and zinc stress.

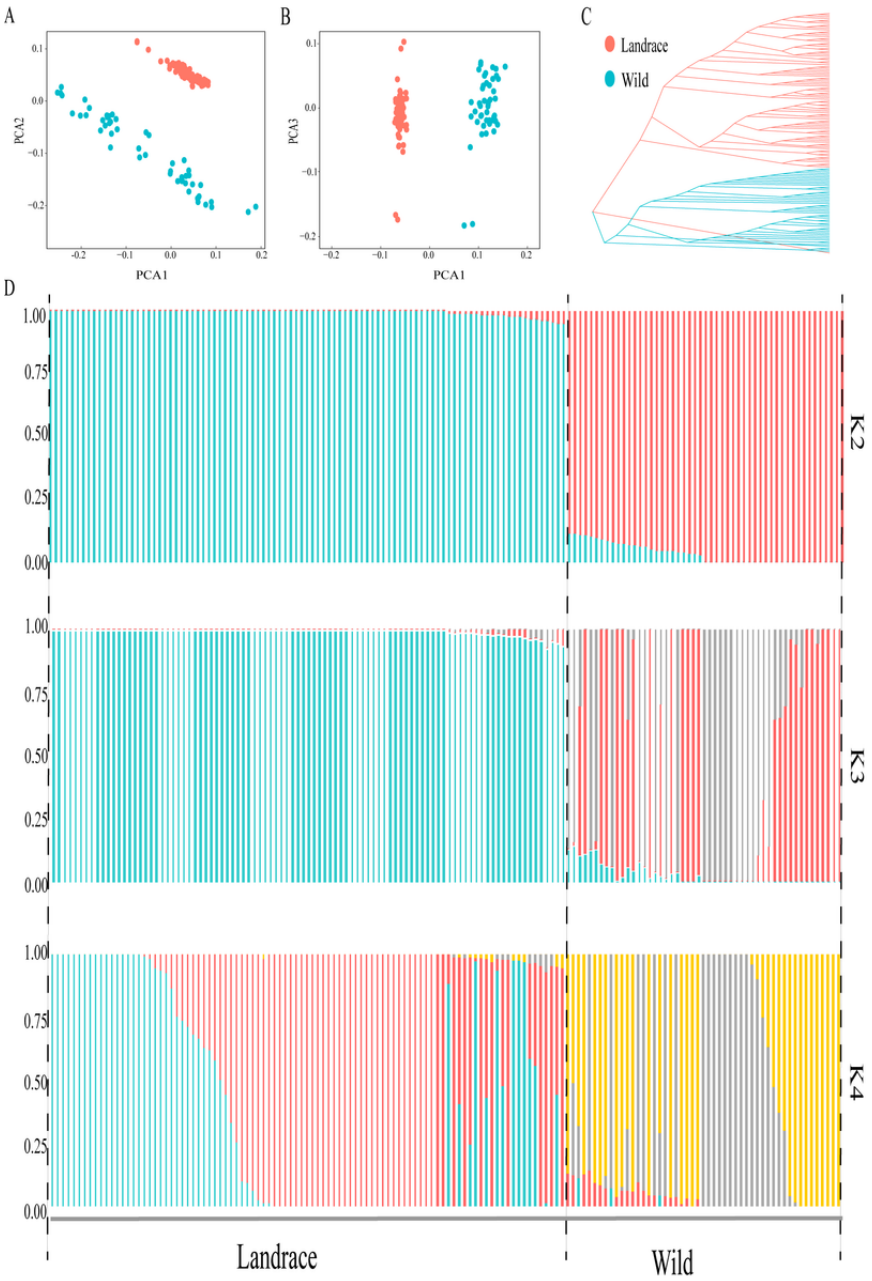

Figure 7

Population structure of 95 landraces and 51 wild barley accessions based on the HvC3H-related SNPs. (A) PCA plots of the first component (PC1) and second component (PC2), The color of dots separately indicates the population and location. (B) PCA plots of the first component (PC1) and third component (PC3), The color of dots separately reflects the population and location. (C) The NJ phylogenetic tree. Branch colors indicate different populations. (D) Population structure with K ranging from 2 to 4.

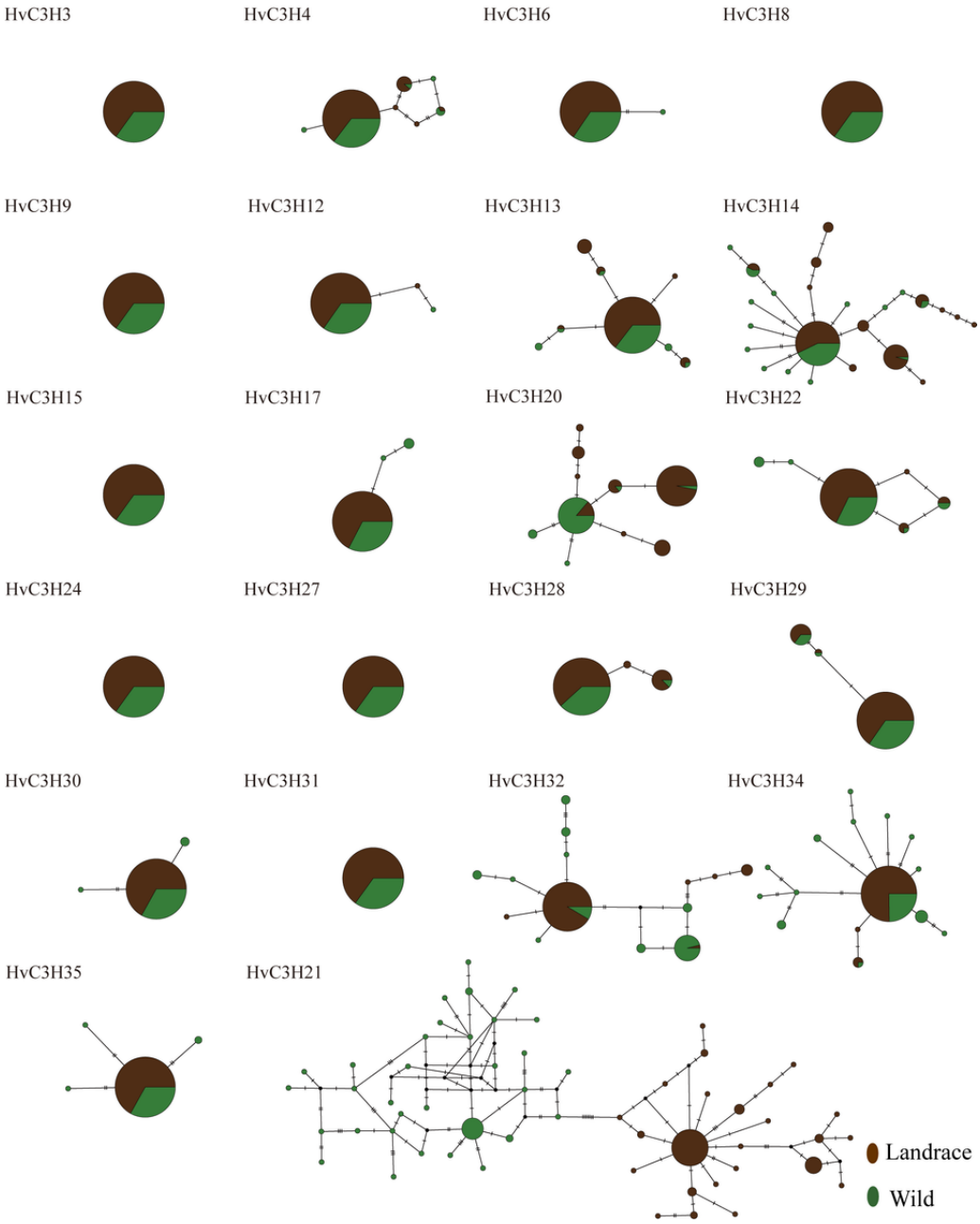

Figure 8

Median-Joining haplotype network analysis of 22 HvC3Hs in wild and domesticated barley populations. The circle size represents the number of accessions holding a particular haplotype. Circle colors refers to different populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.jpg
- FigureS2.jpg
- FigureS3.jpg
- FigureS4.jpg
- FigureS5.jpg
- FigureS6.jpg

- FigureS7.jpg
- FigureS8.jpg
- FigureS9.jpg
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx
- TableS10.xlsx
- TableS11.xlsx
- TableS12.xlsx