scientific reports

OPEN

Systematic identification and expression profiles of the BAHD superfamily acyltransferases in barley (*Hordeum vulgare*)

Zhen Yuan[1], Hongliang Yang[1], Leiwen Pan[1], Wenhui Zhao[1], Lunping Liang[1], Anicet Gatera[1], Matthew R. Tucker[2] & Dawei Xu[1]✉

BAHD superfamily acyltransferases play an important role in catalyzing and regulating secondary metabolism in plants. Despite this, there is relatively little information regarding the BAHD superfamily in barley. In this study, we identified 116 HvBAHD acyltransferases from the barley genome. Based on phylogenetic analysis and classification in model monocotyledonous and dicotyledonous plants, we divided the genes into eight groups, I-a, I-b, II, III-a, III-b, IV, V-a and V-b. The Clade IV genes, including *Agmatine Coumarol Transferase (ACT)* that is associated with resistance of plants to Gibberella fungi, were absent in Arabidopsis. Cis-regulatory element analysis of the *HvBAHDs* showed that the genes respond positively to GA3 treatment. In-silico expression and qPCR analysis showed the *HvBAHD* genes are expressed in a range of tissues and developmental stages, and highly enriched in the seedling stage, consistent with diverse roles. Single nucleotide polymorphism (SNP) scanning analysis revealed that the natural variation in the coding regions of the *HvBAHDs* is low and the sequences have been conserved during barley domestication. Our results reveal the complexity of the *HvBAHDs* and will help facilitate their analysis in further studies.

Acyltransferases have multiple functions, fulfilling an important role in plant gene expression, metabolism, and signal transduction. Plant secondary metabolism results in the production of many important metabolites, such as phenols (flavonoids), isoprene compounds (such as terpenoids) and nitrogen-containing compounds (such as alkaloids) (Fig. 1a). The BAHD acyltransferases catalyze the acylation of many plant secondary metabolites. They represent a superfamily of enzymes found primarily in plants, algae and bacteria that are responsible for transfer of acylated moieties (RC(O)R') from an acyl-activated CoA thioester donor to an acceptor molecule, whose products include small volatile esters, modified anthocyanins as well as constitutive defense compounds and phytoalexins[1]. BAHD acyltransferases were named after the first letter of each of the first four enzymes identified biochemically in this superfamily[2]. These four enzymes are BEAT benzylalcohol O-acetyltransferase (BEAT), cloned from *Clarkia breweri* and involved in the production of volatile ester compounds, anthocyanin O-hydroxycinnamoyltransferase (AHCT), cloned from *Gentiana triflora* and involved in the acylation of anthocyanin, anthranilate N-hydroxycinnamoyl/benzoyltransferase (HCBT), cloned from *Dianthus caryophyllus* and involved in the production of anthramide plant antitoxins, and deacetylvindoline 4-Oacetyltransferase (DAT) cloned from *Catharanthus roseus*, involved in the final step of the synthesis of the alkaloid vindoline, deacetylated vindoline 4-O-acetyltransferase. Among these enzymes there are two kinds of acyltransferases, one that mainly acts on the release of plant volatile ester compounds, and another that is closely related to the synthesis of plant anthocyanins/flavonoids. Members of the HCBT and DAT families typically share two important conserved amino acid motifs including HXXXD (Histidine, XXX, Aspartic acid) that is typically found in the central region of the enzyme, and DFGWG (Aspartic acid, Phenylalanine, Glycine, Tryptophan, Glycine), a conserved motif that is typically found near the carboxyl terminus of the protein and is thought to participate in CoA binding (Fig. 1b)[1].

[1]School of Agronomy, Anhui Agricultural University, Hefei 230036, China. [2]School of Agriculture, Food and Wine, Waite Research Institute, University of Adelaide, Adelaide, SA 5064, Australia. ✉email: xudawei@sjtu.edu.cn
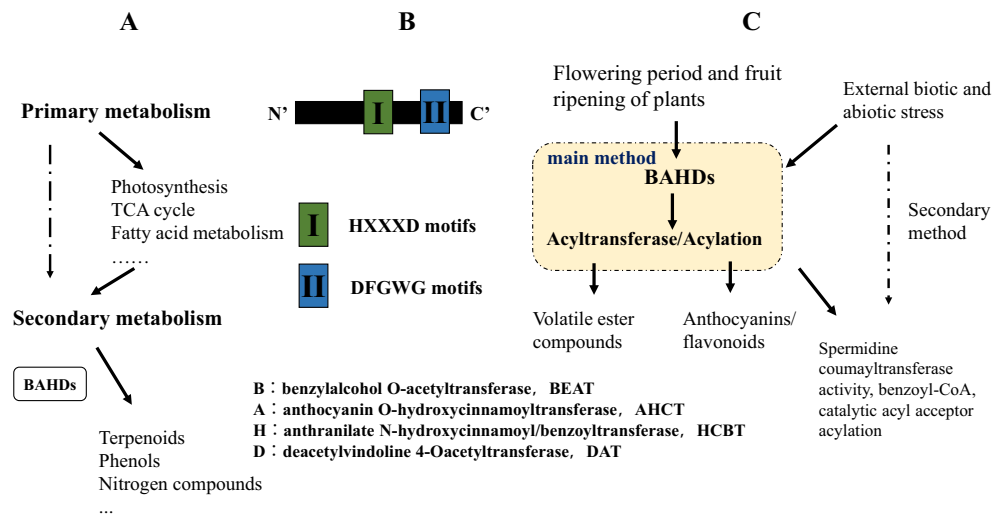
1

Figure 1. Functional roles of *BAHD* genes in plants. (**a**) *BAHDs* play a prominent role in secondary metabolism. (**b**) Typical protein structure of BAHDs. (**c**) Activation of BAHDs.

In recent years, based on the completion of rice and *Arabidopsis* whole genome sequencing, at least 119 rice genes have been identified as part of the *BAHD* superfamily, and 64 (including pseudogenes) have been identified in Arabidopsis[3]. This has expedited research around *BAHD* gene function, catalytic mechanisms, evolutionary relationships, and in particular, their relevance to medicine. Polyphenolic compounds are an important dietary component, and therapeutic benefits of hydroxycinnamate ester conjugates such as Monolignol hydroxycinnamate have been discussed[4]. Similarly, a BAHD enzyme Feruloyl-CoA:monolignol Transferase (AsFMT) from Chinese angelica (*Angelica sinensis*) is involved in the production of coniferyl ferulate, which has antibacterial, antioxidant and anti-Alzheimer's properties[5]. It has also shown potential reversal effect in antitumor adjuvant therapy[6]. The formation of chicory acid in echinacea also involves BAHD acyltransferase; this metabolite has high medicinal value in antiviral, anti-inflammatory, and glycolipid balance[7].

The most common way for eukaryotic proteins to be modified after translation is via acylation or deacylation, and this is controlled by the abundance and activity of acyltransferase enzymes. These enzymes can have diverse roles in growth and development. Several characterized BAHD members in *Arabidopsis* use hydroxycinnamoyl-CoAs as substrates, including Hyroxycinnamoyl-Coenzyme A: Shikimate/Quinate Hydroxycinnamoyltransferase (HCT), which is involved in lignin synthesis[8], the Spermidine Hydroxycinnamoyl Transferase (SHT), which is involved in the synthesis of hydroxycinnamoyl spermidines in the tapetum of *Arabidopsis* anthers[9], and poly-amine acyltransferase, responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed[10]. Acetyl CoA:(Z)-3-hexen-1-ol Acetyltransferase (CHAT) is responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in *Arabidopsis thaliana*[11], while *Defective in Cuticular Ridges* (*DCR*) is required for incorporation of the most abundant monomer into the polymeric structure of the *Arabidopsis* flower cutin[12]. There have also been reports of suberin and cutin feruloyl transferases and a wax fatty alcohol caffeoyl transferase, which transfer hydroxycinnamoyl-CoAs to ω-hydroxy fatty acid acceptors that accumulate in the extracellular matrix[13–15]. Recently a novel BAHD family acyltransferase *Brassinosteroid Inactivator2* (*BIA2*) was identified that is involved in BR homeostasis and may inactivate bioactive BRs by esterification, particularly in roots and hypocotyls under dark conditions[16]. In rice, BAHD acyltransferase OsAT10 functions as a p-coumaroyl coenzyme A transferase involved in glucuronoarabinoxylan modification[17], while the *Defective Pollen Wall 2* (*DPW2*) BAHD acyltransferase is important for the formation of rice pollen wall; after knockout of DPW2, the pollen wall is destroyed, leading to pollen abortion (Fig. 1c)[18]. Acyltransferases can also catalyze the esterification of cholesterol (acyl-CoA: cholesterol acyltransferase, ACAT)[19], catalyzing free cholesterol to form cholesterol ester.

Barley (*Hordeum vulgare*) is the fourth major cereal in terms of production just after maize, rice and wheat, and has high value in terms of animal feed, human food, and brewing industries. Barley also has the advantages of strong resistance to stress, strong adaptability, and wide-ranging cultivation. The International Barley genome Sequencing Consortium (IBSC) recently released the first version of the barley reference genome sequence[20], and with the sequencing of the barley pan-genome[21], considerable genetic information has been unearthed. At present, the genome-wide analysis of the barley *BAHD* gene family has not been reported. Such an analysis would provide scope to investigate interspecific similarities and putative unique functions in the Triticeae cereals. Here, we mined genomic sequence resources and constructed a Hidden Markov Model (HMM) of the barley *BAHD* gene family. This was used this to define the *BAHD* supergene family, assess gene expression and predict putative functions of barley *BAHD*-related genes[22,23].

Results

Identification of *HvBAHD* genes in the barley genome. The genome sequence of Barley cultivar Morex was used to identify putative *HvBAHD* genes. To achieve this, local protein BLASTs were carried out using HvBAHD protein sequences from *Oryza sativa* and *Arabidopsis thaliana*. In parallel, an HMM search of the Pfam website was performed using the conserved HvBAHD superfamily feature domain (Pfam:PF02458). A total of 116 genes were identified from *Hordeum vulgare* after deleting redundant sequences and designated as *HvBAHD001* to *HvBAHD116*. Almost all selected genes encode the HXXXD domain, but the DFGWG domain showed variability. The basic physical and chemical properties of the putative HvBAHD proteins such as amino acid length, molecular weight (Da), isoelectric point (pI), hydrophobicity prediction, and GC (%) content are summarized (Supplementary Table 1). The length of 75% of the HvBAHD proteins is around 412–499 amino acids, which is the same as *Arabidopsis* and rice, 63% (73/116) are hydrophilic, 72% exhibit an isoelectric point less than 7, and the molecular weight ranged from 40 to 59 kDa. The Gene ID for each sequence derived from Golden Promise is provided in Supplementary Table 1[24].

Phylogenetic analysis of HvBAHD proteins. To investigate the cross-species relationship of BAHD family proteins, phylogenetic evolutionary trees of 295 BAHD protein sequences (116 barley, 119 rice and 60 Arabidopsis) are constructed using MEGA X (https://www.megasoftware.net/), and the phylogenetic trees was generated by the maximum likelihood method. The protein sequences of the HvBAHD family proteins were classified according to the classification of rice and Arabidopsis following previous methods[1,3,25]. According to the clustering with known Arabidopsis and rice orthologs, barley BAHDs proteins were divided into 5 clades (Fig. 2, Supplementary Fig. S1).

Clade I consists of two subclades: Clade Ia (*15 HvBAHDs*) and Clade Ib (18 *HvBAHDs*) (Fig. 2) while Clade II (2 *HvBAHDs*) contains the fewest members (Fig. 2). Clade III can be divided into two subclades: Clade IIIa (0 *HvBAHDs*) and Clade IIIb (5 *HvBAHDs*) (Fig. 2). There are two genes showing homology to Eceriferum2 (CER2) in barley, *HvBAHD031* and *HvBAHD083* (Fig. 2), which appear highly conserved compared to rice and Arabidopsis. Clade IV (22 *HvBAHDs*) appears to be specific to the monocot species analyzed here (Fig. 2). One of the BAHD proteins in this clade, the agmatine coumarol transferase (ACT), is reported to catalyze the synthesis of hydroxycinnamate agmatine from barley agmatine. Hydroxycinnamoyl-CoA thioester, its precursor, can act as an antifungal compound hordatine[26,27]. Hydroxycarnitine agmatine derivatives have also been found at low levels in wheat[28]. Clade IIIa may be specific to dicots, as rice and barley sequences are absent. AT4G15400 in this clade is involved in the synthesis and modification of the triterpene analogue brassinosteroid (BR), and thus participates in the regulation of BR homeostasis and plant light signals[29]. Although most members in Clade III have different acylated alcohol acceptors, their donors are predominantly acetyl-CoA. Some acetyl-CoA BAHDs are involved in the synthesis of volatile esters in fruit (e.g., *Rosa chinensis* RhAAT1), and others are involved in the modification of alkaloids. Clade V incorporates the largest number of *HvBAHDs* and can be divided into two subclades, Clade Va (39 *HvBAHDs*) and Clade Vb (15 *HvBAHDs*). In Clade Va, *HvBAHD029* is the likely barley orthologue of *AT5G41040* (ASFT), *AT5G63560* (FACT) and *AT3G48720* (DCF), which are involved in suberin and cutin biosynthesis[13,30–32]. Also in this clade, *HvBAHD115* is the orthologous gene of *AT2G23510* (spermidine diprosyl acyltransferase; SDT), which is required for the production of disinapoyl spermidine and its glucoside in seeds, and *AT2G25150* (spermidine dicoumarolyl acyltransferase SCT), which is mainly expressed in roots and has spermidine coumaroyl CoA acyltransferase activity[33]. The characteristic protein of acyl CoA in Clade Va is involved in the formation of volatile esters and is related to hydroxycinnamoyl transferase (HCT) which is responsible for the synthesis of chlorogenic acid and monoxylitol. AtHCT (*AT5G48930*) is in Clade V, which includes *HvBAHD011*, *HvBAHD010* and *HvBAHD082*[17]. Bartley et al., and others indicate that HCT is involved in the synthesis of lignin precursors in most vascular plants, since several characterized members (e.g., hydroxycinnamoyl-coenzyme A shikimate/quinate HCT) use hydroxycinnamoyl-CoAs as substrates. HQT is closely related to HCT, both which have HHLVD and DFGWG motifs.

In general, close barley homologues of functional BAHD genes are present in every sub-clade, except for Clade IIIa. This indicates that their function as acyltransferases in secondary metabolism may have been retained through evolution.

Chromosomal location of *BAHD* genes. All 116 *HvBAHDs* genes were mapped onto barley chromosomes, designated as *HvBAHD001* to *HvBAHD113* according to the physical location, except for *HvBAHD114* to *HvBAHD116* whose locations remained unknown (Fig. 3). Further, chromosome position analysis showed that these genes were unevenly distributed on most chromosomes (chr) of barley. Chromosomes 2, 7, 4 and 1 contained the majority of *BAHD* genes (Fig. 3), which were distributed mainly at the ends of the chromosomes.

Analysis of conserved motif and cis-acting elements in the promoters of *HvBAHD* genes. Protein motif distributions may contribute to functional diversity among the members of a gene family. Therefore, we assayed the motif distribution in predicted full length protein sequences of HvBAHD proteins using MEME (Fig. 4; Supplementary Table S2). A total of 10 motifs were predicted. Most HvBAHDs contained both motif1 and motif4. Based on gene structural analysis we determined that motif1 and motif4, with sequences FTCG-GFVIGLRTNHAVADGTGAAQFLNAV and FDVYGNDFGWGRPV, correspond to domains HXXXD and DFGWG, respectively (Supplementary Table S2). The type and distribution of conserved motifs in the same subfamily of HvBAHDs are similar, which is consistent with the clade classification results.

To assess conservation of putative cis-regulatory elements in the *HvBAHD* genes we assessed all 116 genes to identify 2000 bp of upstream sequence. However, the barley intergenic sequence is currently interspersed with small regions of missing data. To ensure the reliability of prediction, we ignored *HvBAHDs* with missing sequence
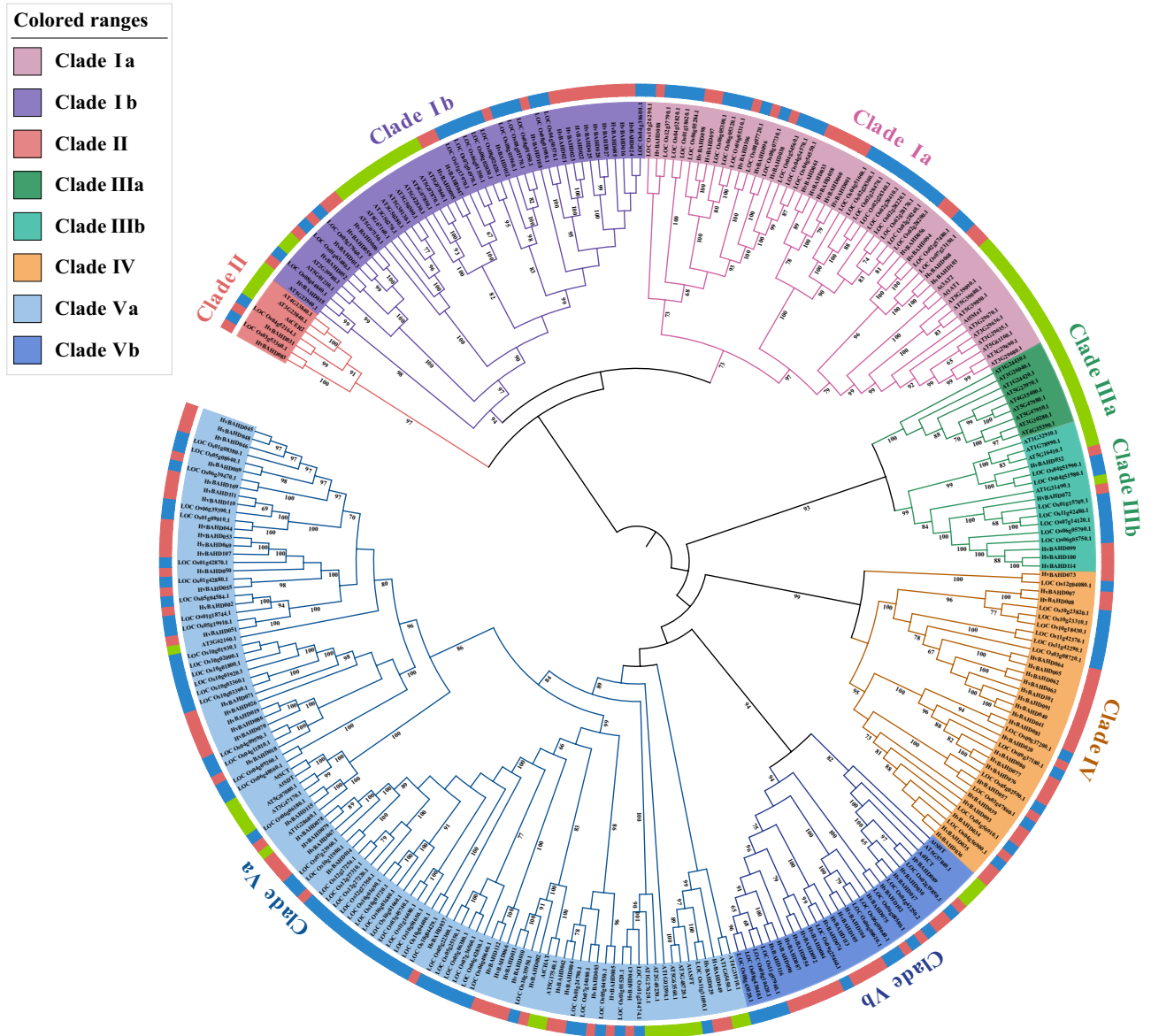
Figure 2. Phylogenetic analysis of BAHD proteins. The BAHD sequences of *Arabidopsis thaliana*, rice and barley were used to construct a phylogenetic tree using the maximum likelihood method. The sequences were divided into five clades. Different clades in the inner ring are distinguished by different colors, and different colors in the outer ring represent different species (green is Arabidopsis, red is barley, blue is rice). This phylogenetic tree was generated by MEGA X (https://www.megasoftware.net/) and iTOL online site (https://itol.embl.de/).

data and focused on 79 *HvBAHDs* containing complete upstream 2000-bp sequences. Sequences from each clade of *HvBAHDs* were represented in the 79 upstream regions, so the results were general to some extent, but not completely representative of the *HvBAHD* family. The results show that some cis-regulatory elements such as photo-responsive elements, plant hormone responsive elements, plant defense, and compressive stress were all related to putative *HvBAHD* family function. Details are shown in Supplementary Table S3. The two fundamental gene elements of the CAAT-BOX (Common Cis-Acting Element in Promotor and Enhancer Regions) and TATA-BOX (Core Promotor Element Around-30 of Prescription Start) were removed, and the remaining elements were plotted (Supplementary Fig. S2). Certain elements were clearly prevalent in a majority of *HvBAHDs*, include light-responsive (G-box), biotic and abiotic stress-responsive elements (As-1, ARE, STRE, MYB and MYC) which were identified at a high frequency. These elements are also found in *Arabidopsis thaliana* BAHD sequences but without special annotation information. It is possible they reflect important conserved elements of the BAHD family. The cis-element analysis identified several gibberellin-related reaction elements in the *HvBAHD* genes that may contribute to activation of *HvBAHD* genes. Moreover, previous studies have shown that the ACT-related genes in barley antagonize gibberellin function.
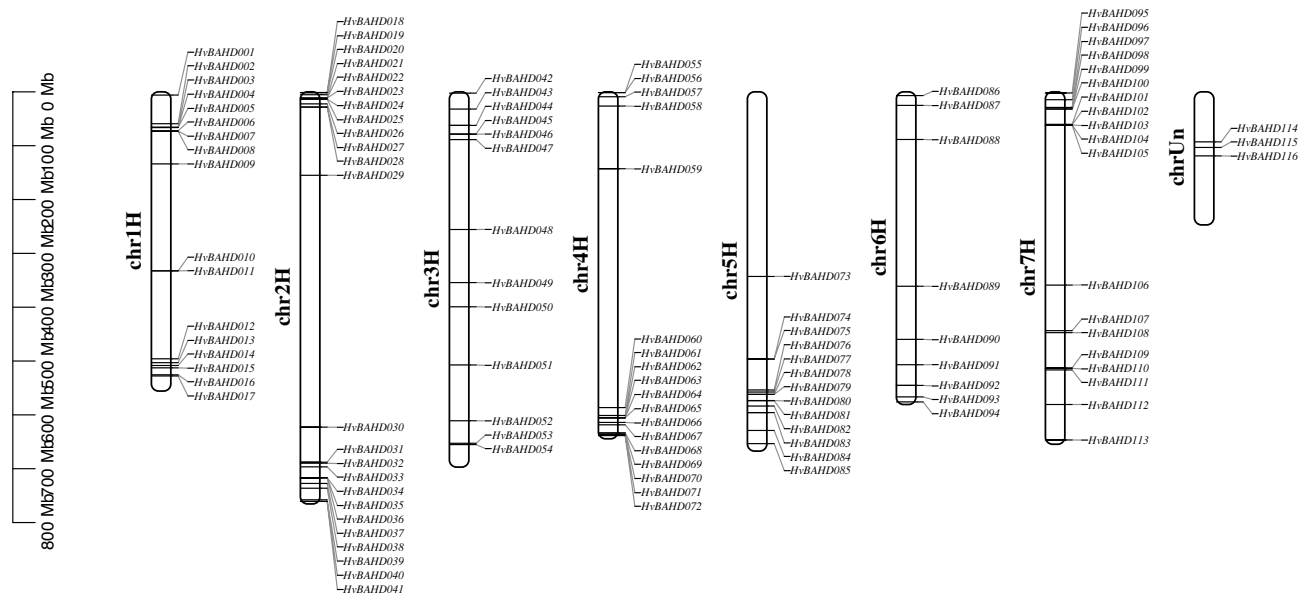
Figure 3. Distribution of 113 putative *HvBAHD* genes (3 genes on unknown chromosome; chrUn) across the seven chromosome (chr) groups of barley. All barley chromosomes are drawn to scale based on predicted lengths. This picture was constructed by TBtools v1.082 (https://doi.org/10.1016/j.molp.2020.06.009).

Evolutionary analysis of *HvBAHD* genes. Phylogeny can often reveal information regarding gene duplication events. The Ka/Ks value is an important parameter that is used to evaluate the evolution of coding sequences and determine the type of selection pressure after duplication. In our study, we searched for evidence of gene duplication through local blast comparisons of barley *BAHD* genes and screening of data to identify sequences with > 75% identity[34]. According to the phylogenetic analysis, there are 14 pairs of homologous genes in the *HvBAHDs* (Fig. 5); the putative tandem repeat genes were analyzed by KaKs, and the Ks, Ka, and the Ka/Ks ratio for the 14 gene pairs are shown in Supplementary Table S4.

Non-synonymous changes are typically subject to natural selection, while synonymous changes are not. In evolutionary analysis, it is of great significance to understand the rate of occurrence of synonymous and non-synonymous changes. If Ka/Ks > 1, it is considered that there is a positive selection effect, if Ka/Ks = 1, it is considered that there is a neutral choice, if Ka/Ks < 1, it is considered to have purification selection effect[35]. Among the 14 selected paralogue gene pairs, the ratio of Ka/Ks of 10 genes pairs was less than 1, which is consistent with synonymous substitution rather than nonsynonymous substitution to prevent the change of amino acid residues after duplication.

Gene expression analysis of *HvBAHDs* by qRT-PCR and in response to GA3 treatment. To analyze the expression patterns of the barley *BAHD* gene family, expression data from different stages, organs and tissues were extracted from the IPK website and examined using a hierarchical clustering model. The heatmap shows that 111 HvBAHDs were expressed in 15 different developmental stages based on web RNA-seq data (Fig. 6). The *HvBAHDs* were expressed to different degrees at all stages of barley growth. Generally, it can be seen that the expression level of *HvBAHDs* in the older roots is higher than young roots, and the expression level of several genes which belong to Clade V (like *HvBAHD030*, *HvBAHD044*, *HvBAHD055*) was high at all stages (Fig. 6).

To better understand the potential roles of *HvBAHDs* in different tissues, we performed qRT-PCR on 27 genes from 5 clades. The results showed that *HvBAHD012*, *HvBAHD013*, *HvBAHD015* and *HvBAHD095* from Clade I were highly expressed in stems (Fig. 7). *HvBAHD034* and *HvBAHD036* from Clade IV showed high expression at the seedling stage, (Fig. 7). *HvBAHD032*, *HvBAHD099* and *HvBAHD100* distributed in Clade IIIb, and many genes in this clade are involved in the formation of volatile substances and alkaloids; qRT-PCR showed that these genes were highly expressed not only in young leaves but also in roots (Fig. 7). *HvBAHD052* shares homology with genes involved in the synthesis of hydroxycinnamyl transferase (HCT) and is expressed in young leaves (Fig. 7). The qRT-PCR analysis showed that many of the *BAHD* genes are expressed in developing seedlings. However, there are clear differences in gene expression from different clades, which may relate to different functional roles in specific tissues within the BAHD family.

The cis-element analysis identified several GA-responsive elements in the promoters of selected *BAHD* genes, which may relate to the role in response the GA pathway. For example, GARE-motif and p-box (Gibberellin-responsive element) are related to gibberellin response in barley and were particularly prevalent in Clade IV genes. ACT in Clade IV was related to Gibberella zeae, while there was a subtle relationship between Gibberella zeae and gibberellin. To test this further, transcriptional responses of selected *HvBAHD* genes were examined in plants containing 5 leaves after treatment with 50 mg/L GA3 for 3 days, 5 days, and 9 days. The results of
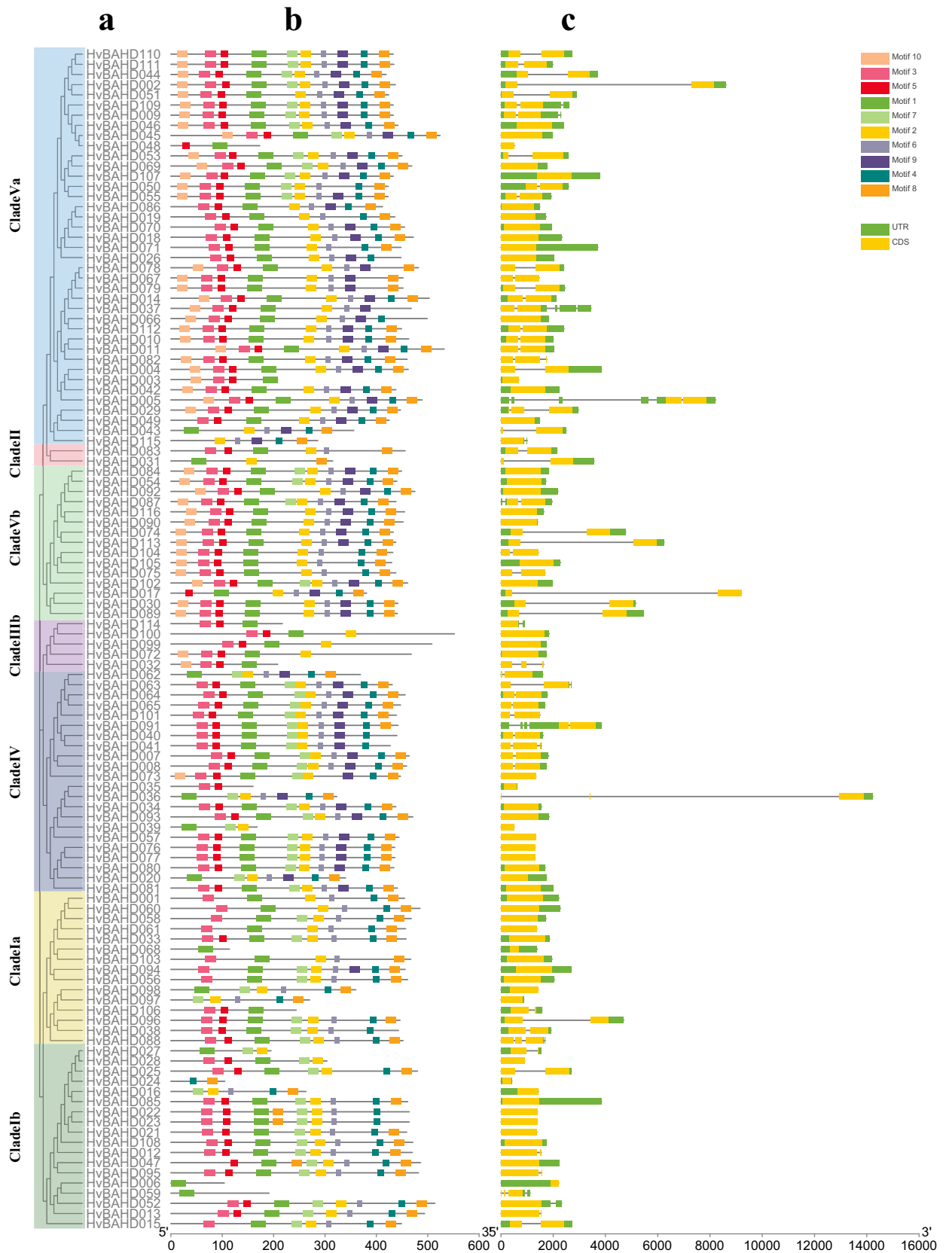
◀**Figure.4.** Phylogenetic relationships, gene structure and architecture of conserved protein motifs in *BAHD* genes from barley. (**a**) A maximum likelihood tree was constructed using the predicted amino acid sequences of HVBAHD proteins. This phylogenetic tree was generated by MEGA X (https://www.megasoftware.net/) and iTOL online site (https://itol.embl.de/). (**b**) The motif composition of barley BAHD proteins. The motifs are displayed in different colored boxes, motifs in the HvBAHD proteins were elucidated by MEME 5.0.1 (https://meme-suite.org/meme/tools/meme). Motif1 is HXXXD domain and Motif4 is DFGWG domain. The sequence information for each motif is provided in Supplementary Table S2. (**c**) Exon–intron structure of barley *BAHD* genes. The green and yellow squares represent the untranslated region (UTR) and coding sequences (CDS) respectively. The black lines represent the intronic. Genomic structure was constructed by TBtools v1.082 (https://doi.org/10.1016/j.molp.2020.06.009).
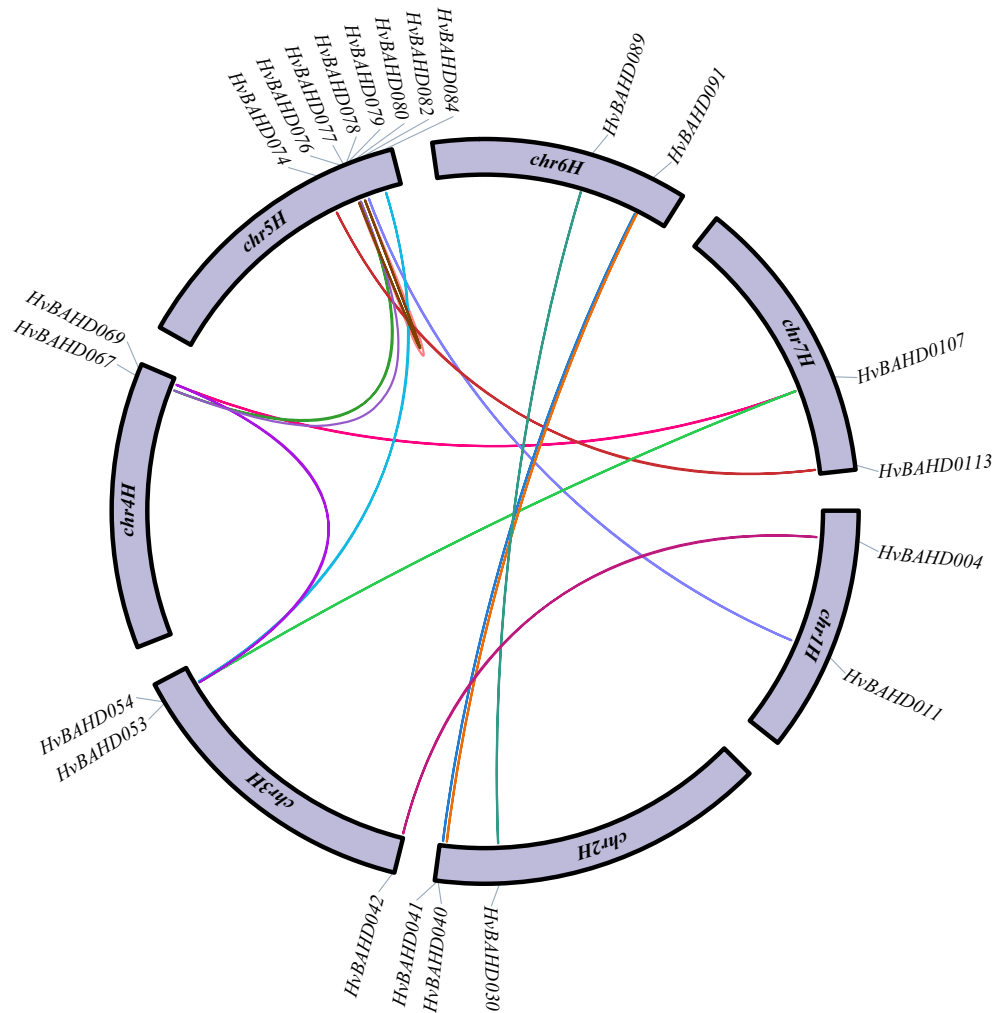

Figure 5. The homologous gene pairs and chromosomal locations of the barley *BAHD* members. Lines of different colors represents a certain replication relationship between the two genes. The outer circle is the gene name of the associated gene and the distribution position on the chromosome. This picture was constructed by TBtools v1.082 (https://doi.org/10.1016/j.molp.2020.06.009).

qRT-PCR are shown in Fig. 8. After treatment of leaves with GA3, we found that the expression of *HvBAHD* genes did not respond significantly on the 3rd day after treatment, but the gene expression was increased significantly after the 5th day of treatment (Fig. 8).

The high level of expression level of *HvBAHD* genes at seedling stage was also of interest, so we conducted another GA3 spray treatment on the seedlings. Leaves were collected from seedlings sprayed with GA3 or water after 5d. qRT-PCR showed that most *HvBAHD* genes assayed showed a significant increase in expression, indicating that GA3 had already affected the plants and *HvBAHD* genes were responding (Fig. 9).

Figure 6. Expression profiles of a subset of *BAHD* genes in different barley tissues and organs. The transcript levels are indicated by a graded color scale from white to green. Expression data were obtained from IPK (https://www.ipk-gatersleben.de/). Normalized expression levels were transformed by logarithmic function and hierarchical clustering of the subset of *HvBAHD* genes are shown. The heatmap is drawn by the TBtools v1.082 (https://doi.org/10.1016/j.molp.2020.06.009). ROO1: Roots from seedlings (10 cm shoot stage); LEA: Shoots from seedlings (10 cm shoot stage); INF1: Young developing inflorescences (5 mm); INF2: Developing inflorescences (1–1.5 cm); NOD: Developing tillers, 3rd internode (42 DAP); CAR5: Developing grain (5 DAP); CAR15: Developing grain (15 DAP); ETI: Etiolated seedling, dark cond. (10 DAP); LEM: Inflorescences, lemma (42 DAP); LOD: Inflorescences, lodicule (42 DAP); EPI: Epidermal strips (28 DAP); RAC: Inflorescences, rachis (35 DAP); ROO2: Roots (28 DAP); SEN: Senescing leaves (56 DAP).

Subcellular localization of selected HvBAHDs in barley. The subcellular localization of selected HvBAHD proteins was analyzed to assess possible differences between BAHD proteins in barley. HvBAHD041, HvBAHD044, HvBAHD063 and HvBAHD096 are in the cytoplasm of tobacco (Fig. 10). HvBAHD041 and HvBAHD063 belong to Clade IV, which appears to be a monocot-specific clade potentially linked to anti-fungal function. The biosynthetic pathway related to the Clade IV *ACT* gene in barley is therefore likely to occur within the cells. *HvBAHD096* is a representative gene from Clade Ia, which also contains *Arabidopsis At5MaT* that functions as an anthocyanin 5-glucoside malonyl CoA acyl transferase both in vitro and in vivo[36]. The localization results indicate that the transferases from At5MaT classes are most likely synthesized in the cytoplasm in barley. By contrast, HvBAHD074 is predicted to be predominantly involved in the synthesis of acetyl-CoA (Clade V branch), and localization studies indicate it is in the nucleus.

SNP data analysis of barley. To investigate the natural variation within the *HvBAHD* family during domestication, we utilised the IPK Gatersleben barley pan-genome database to search for SNPs, and collected information for 93 BAHD genes. Of these, only 52% (48/93) contained any SNPs in the varieties sampled in the SNP-browser, and 47% (44/93) exhibited amino acid changes (Supplementary Fig. S3). However, we also found that the proportion of genes which have SNPs in Clade IV is very high, reaching 71% (12/17), and the two barley genes retrieved in the Clade II sub-clade showed SNP mutation. This indicates that the members of these subunits may have undergone important mutations, and are "young genes" in evolution with relatively high change frequency. Only 32% (30/93) of the *HvBAHD* genes with MAF > 0.05 in SNPs were due to its own functional limitation, which also indicated that most genes were conservative and not affected by SNP site mutations in evolution.

Discussion

Through the retrieval, analysis and organization of the *BAHD* gene family in barley, a total of 116 *HvBAHDs* were identified, all of which have HXXXD characteristic domains. The total number of genes is similar to that of rice and poplar, which indicates that the *BAHD* family plays an important role in plant evolution. A phylogenetic tree was constructed using the maximum likelihood method, which included 116 barley, 60 *Arabidopsis thaliana*, and 119 rice BAHD protein sequences. During phylogenetic analysis, the 116 *HvBAHDs* were divided into five clades, which were similar to those in rice and Arabidopsis except that Clade IV was specific to Gramineae (Fig. 2). The sequence and molecular weight of the HvBAHDs varied considerably, but the characteristic domains and motif composition were conserved (Supplementary Table 1). The exon/intron composition analysis showed that the number of introns in *HvBAHDs* varied from 0 to 5, but most of the *HvBAHDs* only possess one or no introns. The number of introns in different groups was remarkably conserved, except for *HvBAHD005* and *HvBAHD037* (Fig. 4), similar to other plant species[1,25,37]. We found that most of the *HvBAHDs* are distributed towards the gene-rich ends of the chromosomes (Fig. 3), similar to reports for other barley gene families[38,39]. Gene duplication is one of the major mechanisms contributing to genome complexity, as well as to the rapid expansion and evolution of gene families[40,41]. Gene chromosomal distribution demonstrates that at least 14 pairs of *HvBAHDs* have undergone gene duplication (Fig. 5; Supplementary Table S4), and most of these gene duplication events correspond to segment duplication. Only two pairs of *HvBAHDs* (*HvBAHD76* and *HvBAHD80*) and (*HvBAHD77* and *HvBAHD80*) are the products of tandem replication (Fig. 5). The segment duplication events of *HvBAHDs* in barley is similar to the previous studies on *BAHDs* of other species[37].

Most members of the BAHD family are involved in different biochemical reactions, although there are subtle differences from lower plants to higher plants such as BAHD acyltransferase function in various ester synthetic pathways. The *AT4G24510.1* (CER2) located in Clade II contributes to wax formation in the epidermal stratum corneum and physically resists the invasion of external pathogens and controls water loss[42,43]. CER2 have been characterized in *Arabidopsis*, maize, and rice, having similar effects on the extension of waxy VLCFA precursors. CER2 homologues have not been found in *bryophytes*, which suggests CER2 may be a vascular plant innovation[44]. Including these sequences as reference members in our phylogenic analysis provided hints that the HvBAHDs include a large number of potentially novel enzymes, especially for subclades that don't contain any enzymes of known function (Fig. 2). The phylogenetic tree shows clade IV is a specific clade of Gramineae; there are many related response elements in this branch potentially to improve the function of plants against *Gibberella zeae*. In evolution, a special branch of Gramineae has slowly evolved, and there may be more interesting relationships (Fig. 2). The Agmatine coumaroyltransferase (ACT) in this clade was the first amine N-hydroxycinnamoyltransferase characterized in plants and responsible for catalyze the biosynthesis of antifungal hydroxycinnamoylagmatin derivatives in barley and represents a new class of N-hydroxycinnamoyltransferases[45]. The significant number of HvBAHD proteins in Clade IV suggests a possible role in preventing the invasion of fungi, which may
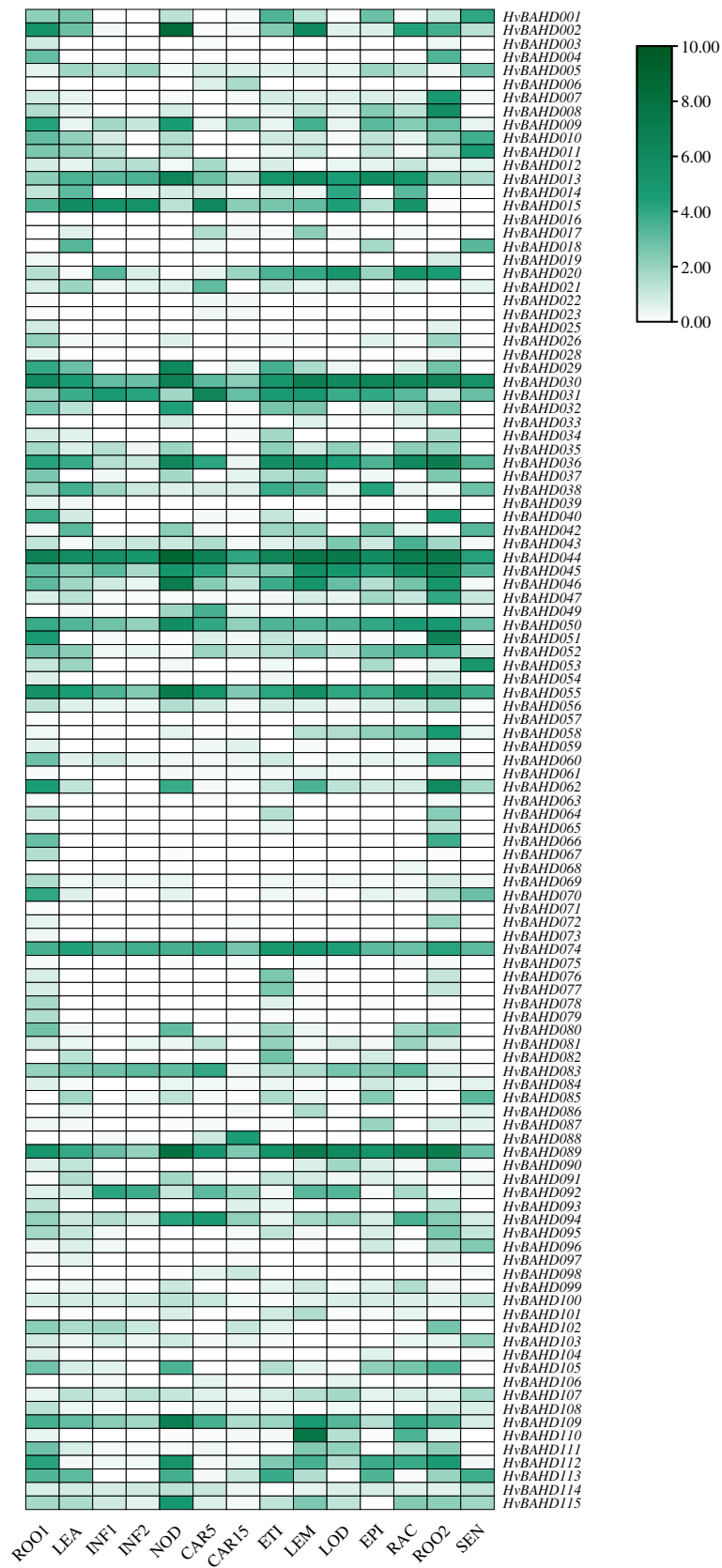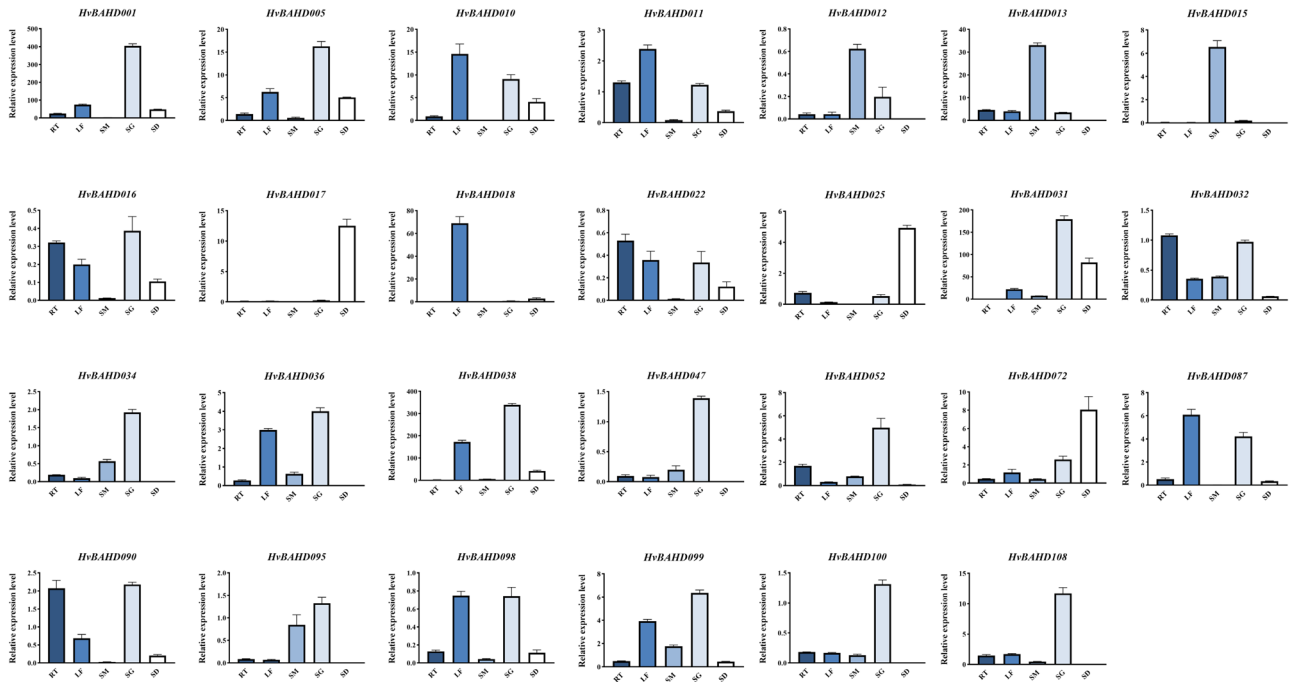
Figure 7. The relative expression levels of the *HvBAHD* genes in five different barley tissues as determined by qRT-PCR. RT: root; SM: stem; LF: leaf; SG: 2-week-old seedling; SD: ripe seed. This picture was generated by GraphPad 8 (https://www.graphpad-prism.cn/) .
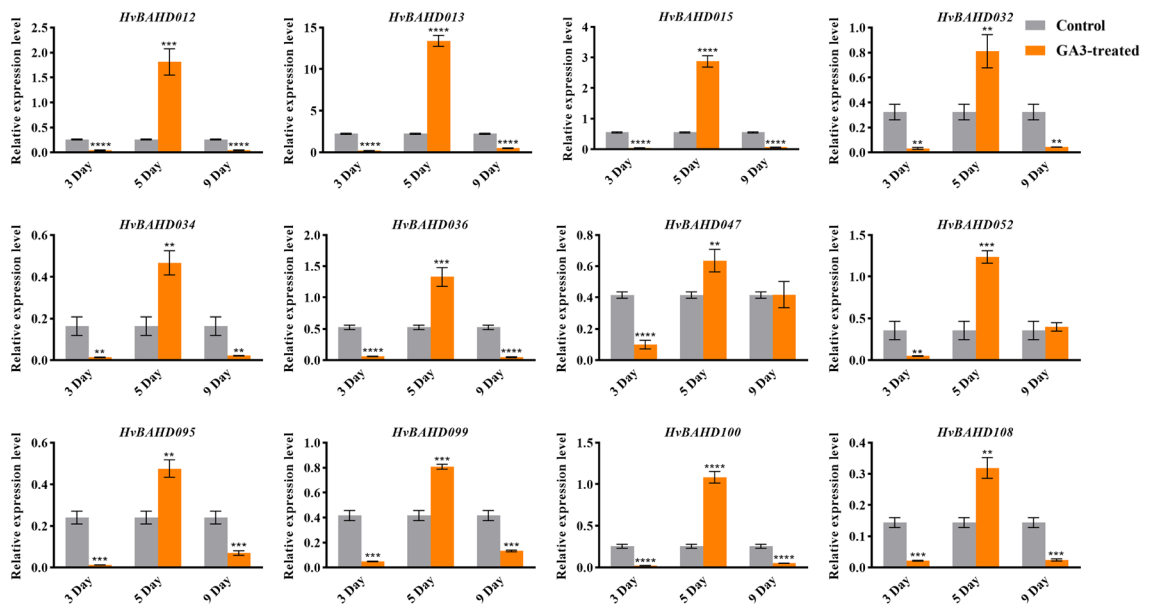

Figure 8. Relative expression level of *HvBAHD* genes on the 3rd, 5th and 9th day after GA3 treatment in barley leaves. Gray shows the control group and orange indicates the GA3 treatment group. (P < 0.0001, ****, P < 0.001, ***, P < 0.01, **, P < 0.1, *). This picture was generated by GraphPad 8 (https://www.graphpad-prism.cn/) .

be related to the evolution of barley. Further experiments are needed for the research in Clade IV, potentially in relation to the regulation of gibberellin which has a certain relationship with *Gibberella zeae*.

The cis-regulatory element analysis can provide an important basis for further functional analysis of *HvBAHDs*. In this study, we found that biotic and abiotic stress-responsive elements accounted for the greatest number of cis-regulatory elements predicted in the promoter regions of *HvBAHDs*. This finding indicates the *HvBAHD* genes are likely to participate in responses against various adverse environmental stresses. In addition,
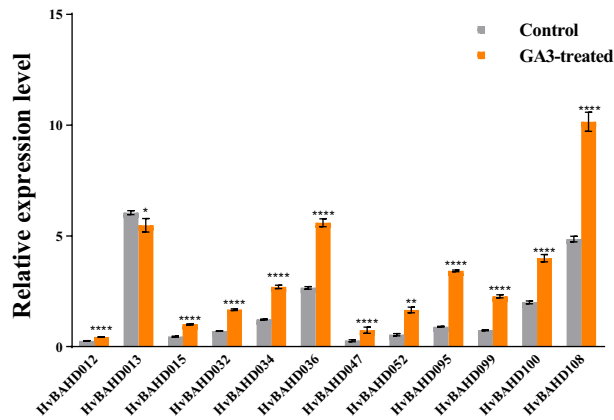
Figure 9. Relative expression of *HvBAHD* genes after 5 days of GA3 treatment in barley seedlings. Gray was used as the control group and orange was used as the GA3 treatment group. (P < 0.0001, ****, P < 0.001, ***, P < 0.01, **, P < 0.1, *). This picture was constructed by GraphPad 8 (https://www.graphpad-prism.cn/) .

GARE-motif and p-box (Gibberellin-responsive element) related to gibberellin response were particularly prevalent in this family. These results are consistent with our GA experiment, which showed that HvBAHDs respond positively to GA3 within a few days of treatment. We also identified several TGACG and CGTCA-motifs, which mainly function as response elements of methyl jasmonate. This is consistent with BAHD function in secondary metabolism and the ability of jasmonate compounds to increase the activity levels of defense proteins such as peroxidase, chitosanase, and lipoxygenase[46,47]. This can result in the accumulation of alkaloids and phenolic acid secondary substances, which increase and change the release of volatile signaling compounds and even form defense structures, such as trichomes and resin conduits. The qRT-PCR results showed that there were expression differences in genes of different branches of *BAHD* family; further experiments are needed to ascertain which members may are associated to JA responses and whether this directly coincides with the presence of cis-elements.

The large number of SNP sites distributed in clade II and clade IV indicated that these two clades may have evolved more specialized functions compared to other branches during plant evolution (Supplementary Fig. S3). Some recent studies have found that the introduction of cell wall BAHD acyltransferase-p-CA monoenol transferase (PMTs) into Arabidopsis and poplar[48], can promote the quasi-coumarinization of lignin and improve saccharification. This illustrates the importance of BAHD acyltransferases in modifying the cell walls of grasses and dicots, and also illustrates their potential to improve saccharification in feed and bioenergy engineering[49].

Biochemical functions of specific *BAHD* gene family have now been elucidated, and it is evident that they play an important role in research of various fields. However, the function of many *BAHD* genes remains unclear. This study provides a baseline for further research into the *BAHD* gene family barley, a key representative of the Triticeae cereals, and one for which functional genomics and genetics resources are becoming readily available.

Conclusion

In this study, the barley BAHD gene family was examined to identify 116 *HvBAHD* genes within the reference genome of cv. Morex. The *HvBAHD* genes are present in similar numbers and categories compared to rice and *Arabidopsis,* based on their phylogenetic analysis. The gene structure, MEME, Ka/Ks, and SNP scanning analyses disclose that most of the *HvBAHDs* were highly conserved. However, the Clade IV genes appear to be specific to the monocots, and include the ACT gene that is associated with resistance of plants to Gibberella fungi. Moreover, we investigated promoter sequences of the *HvBAHD* sequences and found the genes respond positively to GA3 treatment. Gene expression analysis identified most genes were expressed in seedlings. To date, no members of BAHD gene family in barley have been characterized in detail. Hence, the results lay a foundation for further research towards addressing the functional characterization, transcriptional regulation, and role of barley BAHD family members.

Materials and methods

Source of plants and data.   The 2-row spring barley Golden Promise seed was sourced from Shanghai Jiao Tong University. Rice OsBAHD and Arabidopsis AtBAHD protein sequences were downloaded from the Rice Genome Annotation Project (RGAP) (http://rice.plantbiology.msu.edu/)[50] and The Arabidopsis Information Resource (TAIR) (https://www.arabidopsis.org/), respectively. We obtained the barley nucleotide and protein sequences from the plant genomics database Phytozome (https://phytozome-next.jgi.doe.gov/)[51]. HMMER (https://www.ebi.ac.uk/Tools/hmmer/) was used to retrieve candidate barley protein sequences (E-value < 10⁻¹⁰) containing *HvBAHD* conserved domains.

Sequence identification and collection.   TBtools was used to extract the required Rice and Arabidopsis BAHDs protein sequences[52] and MEGA-X (https://www.megasoftware.net/) was used for alignments[53]. Bio-
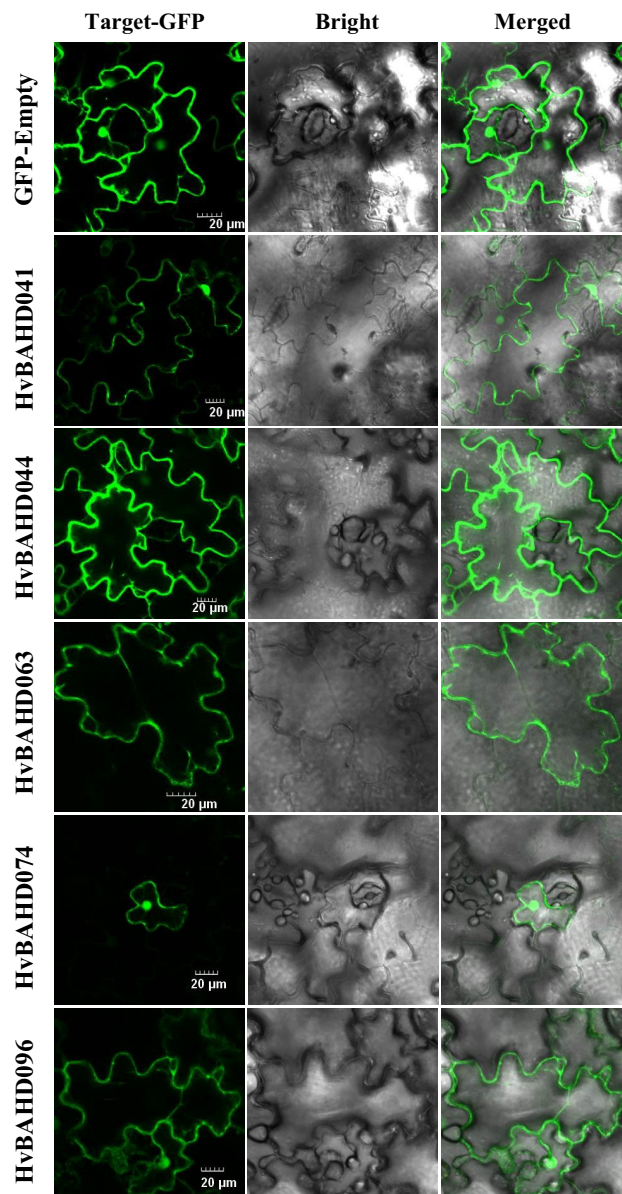
Figure 10. Subcellular localization of HvBAHD proteins in barley. Selected HvBAHD-GFP fusion protein as well as GFP-Empty as the control were independently transiently expressed in tobacco leaves and imaged under a confocal microscope. Bars, 20 μm.

linux software was used to build HMM with the results[54,55]. The sequences with E-values less than e^{-10} were selected as candidate genes and finally 116 *HvBAHDs* were retained. In order to ensure the reliability of the data, we construct the HMM with the BAHD feature domain (PF02458) provided by Pfam (http://pfam.xfam.org/)[56,57].

Phylogenic analysis. The HvBAHDs protein sequences were aligned using ClusterW with default parameters[58]. The phylogenetic tree was constructed with the BAHD proteins from barley, rice and *Arabidopsis* by MEGA-X using the optimal replacement model ML (WAG + G + F) method with 1000 bootstrap replicates[59]. 119 BAHD protein sequences from *Oryza sativa*, 60 BAHD protein sequences from *Arabidopsis thaliana*, and 116 BAHD protein sequences were selected from *Hordeum vulgare*. The constructed phylogenetic tree was rendered using iTOL (https://itol.embl.de/)[60].

Gene structure and motif identification. To identify the exon/intron structure of the *HvBAHD* analysis, barley gene CDS and protein sequences were downloaded from the Barley Genome Database Annotation on Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The MEME tool was used to search for conserved motifs within HvBAHD proteins (https://meme-suite.org/meme/tools/meme)[61]. The maximum value of the

motif was set to 50, and the motif length was set between 6 and 200, the number of motifs is 10, and other parameters by using default settings. Tbtools[52] was used to analyze the gene structure and conserved motifs. The ExPasy (Compute pI/Mw tool) bioinformatics online tool (https://web.expasy.org/translate/)[62] was used to predict the isoelectric point (pI) and relative molecular mass (kDa) of each HvBAHD. The 2 Kb upstream region of each *HvBAHD* sequence was chosen for cis-element analysis using the PlantCARE tool[63] (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

Chromosome position information and collinearity analysis. The predicted location of each gene was collected from the BARLEY IPK website (https://apex.ipk-gatersleben.de/apex/f?p=284:10), and then the chromosome location information was drawn using TBtools. The Ka/Ks were predicted using the KaKs_Calculator2.0[64] program of Bio-linux[65].

Expression analysis of barley *HvBAHD* genes. To explore the specific expression patterns of *HvBAHDs* in different barley tissues and developmental stages, RNA-seq data of different developmental stages were downloaded from IPK (https://galaxy-web.ipk-gatersleben.de/). The gene expression values are represented by fragments per kilobase of exon per million fragments mapped (FPKM). The heatmap is drawn by the online TBtools.

Plant growth, tissue processing and GA (gibberellin) treatment. The spring barley Golden Promise was cultivated at Anhui Agricultural University (31.85°N, 117.26°E) under normal outdoor temperature (15–25 °C). Seed germination was carried out in early March and seedlings were subsequently transferred to pots. Tissue samples collected from the mature root, stem, leaf, seedling, and spike of barley were wrapped in aluminum foil and dipped into liquid nitrogen. All harvested samples were stored at − 80 °C for future experiments including RNA isolation. For GA3 treatment, 50 ml/L GA3 was applied onto the leaves of five-leaf stage plants. Water spraying was used as the control. In short, two pots of barley plants with representative growth habit were selected. One pot was sprayed with exogenous GA3 (50 mg/L) on the leaves, and the control group was sprayed with water to the same degree. Plants were subsequently sprayed once every other day. Plants were watered and fertilized uniformly during the plant growth phase. RNA was extracted from leaves of the two groups on the 3rd, 5th and 9th days and assessed by qRT-PCR to detect *HvBAHD* gene expression. Changes in the growth of plants were documented. Barley seedling was treated with 10^{-4} M gibberellin for spraying, and the control group was treated with water. All descibed methods including the collection of the barley seeds were performed according to the relevant guidelines and regulations of China.

Quantitative real-time PCR (qRT-PCR) analysis. Total RNA was isolated from various barley tissues using total RNA Kit trnzol reagent (Tiangen, Beijing, China) following the manufacturer's protocol. The quality and quantity of all the RNA samples were verified with a Nanodrop 1000 UV spectrophotometer (ThermoFisher Scientific, UT) and checked by 1% agarose gel electrophoresis. For qRT-PCR, first-strand cDNA was reverse transcribed from total RNA with the PrimeScript™ RT reagent Kit with gDNA Eraser (Perfect Real Time; TaKaRa). Real-time RT-PCR was performed with Hieff® qPCR SYBR Green Master Mix (Yeasen, Shanghai, China), using the real-time PCR system (Roche Light Cycler 96). The reaction procedure was a two-step method. The results of relative gene expression were calculated by $2^{\Delta\Delta Ct}$ and the significance was analyzed by t-test[66]. The *HvACTIN* (HORVU1Hr1G074350.1) was used as the internal reference gene[67]. Primers used to quantify the expression of *HvBAHDs* in this study are listed (Supplementary Table S5).

Subcellular localization of the *HvBAHD* genes in barley. For transient expression in tobacco (*Nicotiana benthamiana*) leaves, full-length *HvBAHD* CDS were PCR amplified using *HvBAHDs*-CDS-F and *HvBAHDs*-CDS-R (sequences in Supplementary Table S5), and introduced into the pCAMBIA1301-GFP vector digested with Bgl II and Spe I to generate 1301-35Spro:*HvBAHD*CDS-GFP. Constructed vectors were transformed into Agrobacterium strain GV3101 and infiltrated into leaves of 4-week-old tobacco plants. After 72 h dark culture, the leaf epidermis of tobacco with 1301-35Spro:*HvBAHD*CDS-GFP was mounted in water and observed under a confocal microscope (Leica TCS SP5), using empty 1301-GFP as a control.

Cis-regulatory elements prediction of promoter. The 2000 bp upstream sequence of each barley BAHD-related gene was extracted from the barley genome GFF3 file for cis-regulatory element analysis. The data were screened and uploaded to the PlantCARE[63,68] website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for cis-regulatory element prediction and subsequently visualized in the R studio software.

SNP (single nucleotide polymorphism) data of *HvBAHD* genes. In the IPK-SNP Browser (https://bridge.ipk-gatersleben.de/#snpbrowser)[69], the subgroup of barley "core50" was selected as the research object, and the SNP numbers appearing in the *HvBAHD* gene family were counted.

References

1. D'Auria, J. C. Acyltransferases in plants: A good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340. https://doi.org/10.1016/j.pbi.2006.03.016 (2006).
2. St-Pierre, B. & De Luca, V. *Evolution of Metabolic Pathways Recent Advances in Phytochemistry* 285–315 (2000).
3. Yu, X. *et al.* BAHD superfamily of acyl-CoA dependent acyltransferases in Populus and *Arabidopsis*: Bioinformatics and gene expression. *Plant Mol. Biol.* **70**, 421–442. https://doi.org/10.1007/s11103-009-9482-1 (2009).
4. Eudes, A. *et al.* Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. *Microb. Cell Fact.* **15**, 198. https://doi.org/10.1186/s12934-016-0593-5 (2016).
5. Wilkerson, C. G. *et al.* Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science* **344**, 90–93 (2014).
6. Chen, C. *et al.* Coniferyl ferulate, a strong inhibitor of glutathione s-transferase isolated from radix angelicae sinensis, reverses multidrug resistance and downregulates P-glycoprotein. *Evid. Based Complement. Altern. Med* **2013**, 639083. https://doi.org/10.1155/2013/639083 (2013).
7. Fu, R. *et al.* Versatility in acyltransferase activity completes chicoric acid biosynthesis in purple coneflower. *Nat. Commun.* **12**, 1563. https://doi.org/10.1038/s41467-021-21853-6 (2021).
8. Hoffmann, L. *et al.* Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenyl-propanoid biosynthesis. *Plant Cell* **16**, 1446–1465. https://doi.org/10.1105/tpc.020297 (2004).
9. Grienenberger, E. *et al.* A BAHD acyltransferase is expressed in the tapetum of *Arabidopsis* anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* **58**, 246–259. https://doi.org/10.1111/j.1365-313X.2008.03773.x (2009).
10. Luo, J. *et al.* A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed. *Plant Cell* **21**, 318–333. https://doi.org/10.1105/tpc.108.063511 (2009).
11. D'Auria, J. *et al.* Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in *Arabidopsis thaliana*. *Plant J.* **49**, 194–207. https://doi.org/10.1111/j.1365-313X.2006.02946.x (2007).
12. Panikashvili, D. *et al.* The *Arabidopsis* DCR encoding a soluble BAHD acyltransferase is required for cutin polyester formation and seed hydration properties. *Plant Physiol.* **151**, 1773–1789. https://doi.org/10.1104/pp.109.143388 (2009).
13. Rautengarten, C. *et al.* *Arabidopsis* Deficient in Cutin Ferulate encodes a transferase required for feruloylation of omega-hydroxy fatty acids in cutin polyester. *Plant Physiol.* **158**, 654–665. https://doi.org/10.1104/pp.111.187187 (2012).
14. Kosma, D. *et al.* Identification of an *Arabidopsis* fatty alcohol:caffeoyl-Coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. *Plant Physiol.* **160**, 237–248. https://doi.org/10.1104/pp.112.201822 (2012).
15. Gou, J. *et al.* A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **106**, 18855–18860. https://doi.org/10.1073/pnas.0905555106 (2009).
16. Zhang, Z. & Xu, L. *Arabidopsis* BRASSINOSTEROID INACTIVATOR2 is a typical BAHD acyltransferase involved in brassinosteroid homeostasis. *J. Exp. Bot.* **69**, 1925–1941. https://doi.org/10.1093/jxb/ery057 (2018).
17. Bartley, L. E. *et al.* Overexpression of a BAHD acyltransferase, *OsAt10*, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiol.* **161**, 1615–1633. https://doi.org/10.1104/pp.112.208694 (2013).
18. Xu, D. *et al.* Defective Pollen Wall 2 (DPW2) encodes an acyl transferase required for rice pollen development. *Plant Physiol.* **173**, 240–255 (2017).
19. Rogers, M. A. *et al.* Acyl-CoA:cholesterol acyltransferases (ACATs/SOATs): Enzymes with multiple sterols as substrates and as activators. *J. Steroid Biochem. Mol. Biol.* **151**, 102–107. https://doi.org/10.1016/j.jsbmb.2014.09.008 (2015).
20. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433. https://doi.org/10.1038/nature22043 (2017).
21. Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**, 284–289. https://doi.org/10.1038/s41586-020-2947-8 (2020).
22. Krogh, A. *et al.* Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580. https://doi.org/10.1006/jmbi.2000.4315 (2001).
23. Henderson, J. *et al.* Finding genes in DNA with a Hidden Markov Model. *J. Comput. Biol.* **4**, 127–141. https://doi.org/10.1089/cmb.1997.4.127 (1997).
24. Schreiber, M. *et al.* A genome assembly of the barley "transformation reference" cultivar golden promise. *G3* **10**, 1823–1827. https://doi.org/10.1534/g3.119.401010 (2020).
25. Tuominen, L. *et al.* Differential phylogenetic expansions in BAHD acyltransferases across five angiosperm taxa and evidence of divergent expression among Populus paralogues. *BMC Genomics* https://doi.org/10.1186/1471-2164-12-236 (2011).
26. Burhenne, K. *et al.* A new class of N-hydroxycinnamoyltransferases. Purification, cloning, and expression of a barley agmatine coumaroyltransferase (EC 2.3.1.64). *J. Biol. Chem.* **278**, 13919–13927. https://doi.org/10.1074/jbc.M213041200 (2003).
27. Burhenne, K. *et al.* A new class of N-hydroxycinnamoyltransferases. Purification, cloning, and expression of a barley agmatine coumaroyltransferase (EC 2.3.1.64). *J. Biol. Chem.* **278**, 13919–13927 (2003).
28. Jin, S. *et al.* Accumulation of hydroxycinnamic acid amides in winter wheat under snow. *Biosci. Biotechnol. Biochem.* **67**, 1245–1249. https://doi.org/10.1271/bbb.67.1245 (2003).
29. Wang, Z. *et al.* Brassinosteroid signaling network and regulation of photomorphogenesis. *Annu. Rev. Genet.* **46**, 701–724. https://doi.org/10.1146/annurev-genet-102209-163450 (2012).
30. Gou, J. *et al.* A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA*. **106**, 18855–18860. https://doi.org/10.1073/pnas.0905555106 (2009).
31. Molina, I. *et al.* Identification of an *Arabidopsis* feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiol.* **151**, 1317–1328. https://doi.org/10.1104/pp.109.144907 (2009).
32. Kosma, D. *et al.* Identification of an *Arabidopsis* fatty alcohol: Caffeoyl-coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. *Plant Physiol.* **160**, 237–248. https://doi.org/10.1104/pp.112.201822 (2012).
33. Luo, J. *et al.* A novel polyamine acyltransferase responsible for the accumulation of spermidine conjugates in *Arabidopsis* seed. *Plant Cell Online* **21**, 318–333 (2009).
34. Gu, Z. *et al.* Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol. Biol. Evol.* **19**, 256–262. https://doi.org/10.1093/oxfordjournals.molbev.a004079 (2002).
35. Akhunov, E. D. *et al.* Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiol.* **161**, 252–265. https://doi.org/10.1104/pp.112.205161 (2013).
36. Luo, J. *et al.* Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from *Arabidopsis thaliana*. *Plant J.* **50**, 678–695. https://doi.org/10.1111/j.1365-313X.2007.03079.x (2007).
37. Liu, C. *et al.* Genome-wide comparative analysis of the BAHD superfamily in seven Rosaceae species and expression analysis in pear (*Pyrus bretschneideri*). *BMC Plant Biol.* **20**, 019–2230 (2020).
38. Pourabed, E. *et al.* Basic leucine zipper family in barley: Genome-wide characterization of members and expression analysis. *Mol. Biotechnol.* **57**, 12–26. https://doi.org/10.1007/s12033-014-9797-2 (2015).
39. Guo, B. *et al.* Genome-wide analysis of APETALA2/ethylene-responsive factor (AP2/ERF) gene family in barley (*Hordeum vulgare* L.). *PLoS ONE* **11**, e0161322 (2016).

40. Pegueroles, C. *et al.* Accelerated evolution after gene duplication: A time-dependent process affecting just one copy. *Mol. Biol. Evol.* **30**, 1830–1842. https://doi.org/10.1093/molbev/mst083 (2013).

41. Zhang, P. *et al.* Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.* **4**, R56. https://doi.org/10.1186/gb-2003-4-9-r56 (2003).

42. Negruk, V. *et al.* Molecular cloning and characterization of the CER2 gene of *Arabidopsis thaliana*. *Plant J.* **9**, 137–145. https://doi.org/10.1046/j.1365-313x.1996.09020137.x (1996).

43. Xia, Y. *et al.* Cloning and characterization of CER2, an *Arabidopsis* gene that affects cuticular wax accumulation. *Plant Cell* **8**, 1291–1304 (1996).

44. Haslam, T. *et al.* The unique role of the ECERIFERUM2-LIKE clade of the BAHD acyltransferase superfamily in cuticular wax metabolism. *Plants* **6**, 23. https://doi.org/10.3390/plants6020023 (2017).

45. Burhenne, K. A new class of N-hydroxycinnamoyltransferases. *J. Biol. Chem.* **278**, 13919–13927 (2003).

46. Raza, A. *et al.* Jasmonic acid: A key frontier in conferring abiotic stress tolerance in plants. *Plant Cell Rep.* **40**, 1513–1541. https://doi.org/10.1007/s00299-020-02614-z (2021).

47. Martin, M. J. Enzymes involved in jasmonic acid biosynthesis. *Physiol. Plantarum* **100**, 653–663 (1997).

48. Hasegawa, K. *et al.* Rice putative methyltransferase gene OsPMT16 Is required for pistil development involving pectin modification. *Front. Plant Sci.* **11**, 475. https://doi.org/10.3389/fpls.2020.00475 (2020).

49. Li, G. *et al.* Overexpression of a rice BAHD acyltransferase gene in switchgrass (*Panicum virgatum* L.) enhances saccharification. *BMC Biotechnol.* **18**, 54. https://doi.org/10.1186/s12896-018-0464-8 (2018).

50. Yuan, Q. *et al.* The institute for genomic research Osa1 rice genome annotation database. *Plant Physiol.* **138**, 18–26. https://doi.org/10.1104/pp.104.059063 (2005).

51. International Barley Genome Sequencing, C *et al.* A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**, 711–716. https://doi.org/10.1038/nature11543 (2012).

52. Chen, C. *et al.* TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).

53. Kumar, S. *et al.* MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. https://doi.org/10.1093/molbev/msy096 (2018).

54. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195. https://doi.org/10.1371/journal.pcbi.1002195 (2011).

55. Finn, R. *et al.* HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29-37. https://doi.org/10.1093/nar/gkr367 (2011).

56. Finn, R. D. *et al.* The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-285. https://doi.org/10.1093/nar/gkv1344 (2016).

57. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222-230. https://doi.org/10.1093/nar/gkt1223 (2014).

58. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404 (2007).

59. Douady, C. *et al.* Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* **20**, 248–254. https://doi.org/10.1093/molbev/msg042 (2003).

60. Nguyen, L. *et al.* IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. https://doi.org/10.1093/molbev/msu300 (2015).

61. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).

62. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597-603. https://doi.org/10.1093/nar/gks400 (2012).

63. Rombauts, S. *et al.* PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* **27**, 295–296. https://doi.org/10.1093/nar/27.1.295 (1999).

64. Wang, D. *et al.* KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinform.* **8**, 77–80. https://doi.org/10.1016/S1672-0229(10)60008-3 (2010).

65. Krampis, K. *et al.* Cloud BioLinux: Pre-configured and on-demand bioinformatics computing for the genomics community. *BMC Bioinform.* **13**, 42. https://doi.org/10.1186/1471-2105-13-42 (2012).

66. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408. https://doi.org/10.1006/meth.2001.1262 (2001).

67. Schmittgen, T. D. & Zakrajsek, B. A. *Biochem. Biophys. Methods* **46**, 69–81 (2000).

68. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327. https://doi.org/10.1093/nar/30.1.325 (2002).

69. Konig, P. *et al.* BRIDGE—A visual analytics web tool for barley genebank genomics. *Front. Plant Sci.* **11**, 701. https://doi.org/10.3389/fpls.2020.00701 (2020).

Author contributions

D.X. and Z.Y. conceived the study and design the experimental study. Z.Y., H.Y., W.Z. and A.G. performed the experiments. Z.Y., L.P. and L.L provided useful suggestions regarding the study. Z.Y. and D.X. analyzed the data and wrote the original draft of this paper. M.T. and D.X. revised the paper. All authors have read and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-08983-7.

Correspondence and requests for materials should be addressed to D.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.