Identification of the histone acetyltransferase gene family in the Platycodon grandiflorus genome

Yang Xiang
China Three Gorges University
Jingzhi Wan
China Three Gorges University
Yong Liao
liaoyong_1972@163.com

China Three Gorges University

Article

Keywords: Histone acetyltransferases, Platycodon grandiflorus, Gene family, Epigenetics, Medical plant

Posted Date: January 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8358743/v1

Additional Declarations: No competing interests reported.

Abstract

Triterpenoid saponins, the bioactive principles of the traditional medicinal plant *Platycodon grandiflorus*, possess significant pharmaceutical potential. Their production is highly sensitive to the plant's growth conditions. Histone acetyltransferases (HATs) are characterized by the presence of acetyl groups, which influence mRNA transcription and consequently participate in plant adaptations to their surroundings. While extensive studies on *HAT*s have been conducted across various plant species, there remains a gap in the systematic identification of these enzymes within medicinal plants. In this research, we discovered seven *PgHAT*s and classified these genes into four distinct categories determined by their preserved protein structures. These findings implicated the *HAT* genes from *A. thaliana*, *O. sativa*, and *P. grandiflorus* in specific functional roles. The findings suggest that PgHAT has a well-preserved evolutionary lineage and contains highly variable regions, making it an excellent candidate for further investigation in the context of medicinal plant research. Moreover, motifs from the genome may correlate with the functionality of PgHATs. The cis-regulatory impact on these acetyltransferases modulates their role in specific gene functions, with downstream effects on phytohormone responsiveness, stress adaptation, and developmental growth. We conducted expression analyses to explore the potential functions of PgHATs under three different environmental stress conditions. Our findings highlighted a group of PgHATs that may significantly contribute to how plants respond to changing environmental factors.

Introduction

Histones are primarily modified post-transcriptionally within the N-terminal tails. These changes extend from the nucleosome and encompass processes such as acetylation, methylation, and phosphorylation [1]. The influence of these histone modifications is vital in modulating diverse mRNA expression [2–5]. Acting as a critical epigenetic switch, histone (de)acetylation enables the precise regulation of diverse biological processes and is fundamentally important for orchestrating plant stress responses to environmental challenges [6]. Acetyl moieties are subject to dynamic regulation by histone acetyltransferases (HATs) and are removed by histone deacetylases (HDACs) [7]. The acetylation of specific lysine residues on H3/H4 tails by HATs neutralizes their positive charge, relaxing chromatin structure by reducing histone-DNA affinity, thereby promoting transcription [8].

Typically, acetylation in histone facilitated with HATs is linked to transcriptional evenets. Nonetheless, this mechanism remains inadequately understood in the context of species [9], particularly concerning herb varieties. HATs are categorized into four groups: HACs (responsive to element-binding proteins), HAFs (part of the TATA-binding protein-associated factor family), HAGs (belonging to the N-terminal acetyltransferase family), and HAMs/MYST family-all containing an acetyltransferase domain. [10]. Furthermore, HATs have been identified in various model organisms, including *Arabidopsis thaliana* [11], *Oryza sativa* [12], litchi [13], barley [14], *Vitis vinifera* [15], tomato [16]. The regulatory mechanisms governing HATs are influenced by dynamic histone acetylation, a process potentially connected to physiology and development [6].

AtGCN5 protein are established regulators of cell differentiation and contribute to a broad spectrum of biological functions [17]. Particularly, AtGCN5 plays a crucial role in sustaining the root stem cell reservoir by activating the transcription factors associated with root stem cells [18]. Similarly, with the recruitment of the rice ADA2/GCN5 is recruited by the WUSCHEL homeobox gene to regulate the root growth [19]. Research indicates that light-responsive mRNA transcription is diminished in plants with mutations in *AtGCN5* and *AtHAF2* [20]. Furthermore, flowering phenotypes are suppressed in plants where *AtHAC* has been knocked down [21]. Plants silenced for AtHAM1 and AtHAM2 show significant defects in the development of reproductive structures [22]. Additionally, *AtHAC1* is implicated in various developmental phases of flowering, as well as in root elongation and the regeneration of shoots from scratch [21]. When it comes to regulating flowering time, the acetylation of H4K5, facilitated by HAM1/HAM2 in arabidopsis, influences the transcription of MADS-box genes [23]. Moreover, HATs also participate in ethylene signal transduction [24].

Platycodon grandiflorus is extensively utilized in Asia both as a medicinal herb and as a food source, primarily due to its anti-inflammatory and liver-protective characteristics. The clinical significance of *P. grandiflorus* stems from the presence of active triterpenoid saponins found in its roots [25] [26]. Triterpenoid saponins (TSs) are a class of natural plant secondary metabolites that exhibit amphipathic properties and display a wide range of structural and functional diversity. These compounds consist of triterpenoid or steroidal aglycones that are connected to oligosaccharide units [27]. TSs serve crucial ecological roles, aiding in plant resistance against pests and pathogens, enhancing crop quality, and finding applications across various industries, including pharmaceuticals, pesticides, cosmetics, and food [27, 28]. Prior research has shown that TSs possess a variety of therapeutic benefits, such as wide-ranging therapeutic potential, notably immunomodulatory, cytoprotective, antipathogenic, and antioxidant properties [27]. Consequently, enhancing the artemisinin levels and improving stress tolerance in challenging environmental conditions is crucial. HATs play a role in modulating histone acetylation, which could affect the plant growth performance and adaptation [6]. The whole-genome sequence for *P. grandiflorus* was published [29]. The development opened avenues for investigating functional genes, particularly epigenetic regulators, as there has been nearly no research on the genes within herbs.

This research investigated the complete genome of *P. grandiflorus*, identifying and characterizing the homologs of the *PgHAT* gene family. Our study characterized these proteins through multiple analyses: phylogeny, gene structure, expression patterns, promoter cis-elements, subcellular localization, and interaction networks. Collectively, our findings present a collection of *PgHAT* genes that are specifically responsive to viral infections, paving the way for future research in medicinal plants.

Results

HAT gene identification in the P. grandiflorus genome

Previous research indicated that *A. thaliana* contains 12 *HAT*s, while rice has 8 *HAT*s [11, 12]. We discovered and analyzed 7 HATs in *P. grandiflorus* by performing BLASTP searches using HAT protein sequences from *Arabidopsis* and rice as references (Table S1). The identified HATs were categorized into 4 groups: HAC, HAG, HAF, and HAM [11]. In accordance with their evolutionarily conserved functional modules and their systematic categorization system used for HATs in *Arabidopsis*, the 7 PgHATs were organized into 5 groups, specifically, HAC, HAG1, HAG3, HAF, and HAM. Each of these groups possesses unique conserved domains that underscore the significance of this classification (Fig. 1). The The characteristics and classification of the *PgHAT* gene family, encompassing IDs of gene, IDs of protein, the locations, coding sequence (CDS) and protein sizes, molecular weights (MWs), isoelectric points (pIs), as well as exon counts and group numbers, can be found in Table S1. Approximately half of the PgHACs exhibited sizes exceeding 1000 amino acids (aa), whereas PGRA_09720 (436 aa), PGRA_01018 (482 aa), PGRA_16872 (546 aa), and PGRA_20711 (564 aa) were notable exceptions. The protein sequences of each group displayed a significant degree of similarity. The MWs for the PgHATs ranged between 50.77 and 213.1 kDa. The pIs were observed to fall between 5.47 and 8.44. Among them, PGRA_08965 encoded the longest protein, which also had the highest MW of 213.1 kDa, while PGRA_09720 was responsible for the shortest proteins, characterized by the lowest MW of 50.77 kDa (Table S1). The characteristics of the PgHATs were processed in a manner that closely resembled those of HATs in other plant species [11, 12], implying that the functions of these PgHAT proteins have been evolutionarily preserved.

Evolutionary relationship analysis of the PgHAT proteins

In order to grasp the evolutionary connections and evolutionary background of plant PgHATs, an unrooted neighbor-joining phylogenetic analysis was developed for protein sequences of HATs derived from *P. grandiflorus* and various other plant species. Consequently, chains of 12 *A. thaliana*, 8 *O. sativa*, and 7 *P. grandiflorus* HAT amino acids were employed to create a neighbor-joining (NJ) tree. This analysis did not make any assumptions regarding ancestry, instead illustrating the relative branching order of taxa in an unrooted phylogeny. Consequently, HAT proteins from *A. thaliana*, rice, and *P. grandiflorus* were categorized into 6 clades as anticipated (Fig. 2). A high degree of amino acid sequence identity was observed between PgHAT and model plant HAT proteins. They grouped into the same clades alongside AtHATs and OsHATs, supported by high bootstrap values. One AtHAT, one OsHAT, and one PgHAT were grouped within the HAG1 category; meanwhile, each of the three HAG groups contained one HAT from each species (Fig. 2). The data showed a different biological evolution of the HAT present in the *P. grandiflorus* genome in comparison to that in other model plants. Despite the species, the HAC group emerged as the most extensive, comprising 5 AtHATs, 3 OsHATs, and 2 PgHATs (Fig. 2). Additionally, the HAM and HAF groups contained two AtHATs and one OsHAT, respectively (Fig. 2). Together, the findings imply that the HAT within the *P. grandiflorus* genome demonstrates significant evolutionary conservatism while also being abundant in highly variable regions. This characteristic makes it an ideal candidate for studies in medical plant identification and systematics.

Furthermore, these findings align with earlier research on HATs from *A. thaliana* and rice, exhibiting similar phylogenetic patterns among these genes [11, 12].

Structural prediction and conserved motif identification in the PgHAT family

Recently, it was popular with homology modelling, which has become a vital technique in the field of structural biology. To illustrate the different protein structures associated with histone acetylation transferase, HATs from each group across three different species were randomly selected. Modelling was performed using the SWISS-MODEL software (Fig. 3). The HAM proteins exhibited similar structural characteristics among a range of plant species (Fig. 3). Conversely, this investigation revealed that the structures of HAC, HAF, and HAG proteins diverged from one another (Fig. 3 and Table S2). However, upon closer examination, it became evident that the conserved protein structures were intact, with only minor variations in certain folding directions (Fig. 3). Additionally, since *PgHACs* possess a unique gene (*PGRA_08998*) that has evolutionarily lost specific regional introns, we also created a protein model for it using SWISS-MODEL (Fig. S1). Consequently, it was shown that the gene occupies comparable 3D protein structures alongside other PgHAC family groups (Fig. 1 and Fig. S1). It was discovered that within the same HAT groups, proteins from various plant species exhibit similar structures. The converse of this statement is also valid. Since the protein structures and characteristics determine the organism's properties and functions, our findings implied that HATs from three different plants could carry out analogous biological functions.

Many additional significant methods exist through which introns enhance gene expression, either via the general effects of splicing or through particular characteristics of specific introns that operate through established mechanisms. To gain a clearer insight into the evolution of gene families, we conducted a comparison of the structures of *PgHAT* genes (Fig. 3A). DNA sequence from genome examinations displayed that the exons ranged from 9–23 (Fig. 3A and Table S1). The PgHATs, characterized by highly conserved structures, clustered into three primary clades on the neighbor-joining tree (Fig. 3A). Most *PgHACs* displayed comparable quantities of introns and exons, except for one *PgHAF* gene (*PGRA_08965*). This particular *PgHAF* gene contained the highest intron count (22) across all groups. What's more, to identify potential motifs within the HAT family in *P. grandiflorus*, the protein chains of the seven PgHATs were predicted utilizing MEME. As a result, there are twenty predicted signature motifs (Fig. 3B; Table S3). Members within the same category showed comparable motifs, implying that these proteins might fulfill analogous roles. The group of HAC genes, which are evolutionarily conserved across various plants, displayed the highest quantity of motifs. Each PgHAT protein contained at least one motif, and these motifs were consistently organized within each group, with the exception of PgHAG1 (PGRA_20711), which lacked any predicted motifs. Overall, motifs frequently existing in histone acetyltransferases within the *P. grandiflorus* genome are likely linked to the conserved functions of *PgHATs*, whereas those unique to certain *PgHATs* appear to contribute to specific gene functions.

Subcellular localization pattern of PgHAT proteins

Owing to the value of subcellular localization as an indicator of protein function, prediction of the localizations for PgHAT proteins was performed using the classical software Wolf PSORT and UniProt (Table S4). Utilizing the Wolf PSORT approach, it was demonstrated that all PgHAT proteins, with the exception of HAG3 (PGRA_20711), are predicted to localize to the nucleus with high confidence (RI > 6). HAG3 likely targets both the chloroplast and mitochondria (Table S4). Specifically, HAC (PGRA_12289) was exclusively located in the nucleus (RI = 14), while other PgHATs proteins were anticipated in more than one subcellular organelle (Table S4). UniProt recognized the potential nuclear signal for PgHATs, with the exception of HAG3 (PGRA_20711) and HAG2 (PGRA_01018), which could not be predicted. The findings indicated that the recently identified PgHAT proteins in *P. grandiflorus* displayed diverse subcellular distributions, potentially linked to functional diversification during growth and development.

Spatiotemporal expression patterns of PgHATs during P. grandiflora development

HATs and HDACs influence histone acetylation levels [30]. To facilitate a more profound comprehension of the manner in which *PgHAT* genes are expressed across diverse tissues, a heatmap was created using RNA-seq data sourced from the *P. grandiflorus* database (Table S5). The findings illustrated intricate and overlapping transcriptional events of *PgHATs* from different tissues and organs. The plant samples were divided into six tissues: stem, pistil, stamen, petal, seed, leaf, root, and sepal. The analysis indicated that seven *PgHAT* genes exhibited unique expression across different developmental stages of these tissues. The expression levels of individual genes varied among the various tissues and organs. For instance, *PGRA_12289* (*HAC*) showed minimal expression throughout the entire plant, whereas the *PGRA_08965* (*HAF*) gene displayed significantly higher transcriptional signals in the overall plantlet (Fig. 4). Furthermore, the transcriptional activities of different genes varied considerably within corresponding tissues/organs. As an illustration, in the eight *P. grandiflorus* tissues, the expressions of *PgHAC*s were notably low (Fig. 4). Certain genes showed highly tissue-restricted expression patterns. For instance, *PGRA_08965* (*HAF*) was highly expressed in the root. These findings illustrate that the transcriptional patterns of *PgHATs* differ among the tissues of *P. grandiflorus* and contribute to the growth and development processes in this plant.

Mining promoter regions of PgHAT genes for cis-acting elements

A variety of surrounding stresses impact plant development. Consequently, examining the cis-acting elements linked to these stresses is crucial [31]. To delve deeper into the potential roles of *PgHAT* genes, a search was performed using a promoter database in plant, focusing on the promoters located 2 kb upstream of the transcription initiation site for these genes. As a result, we discovered 126 *cis*-acting elements, which are associated with hormone responsiveness, stress response, light sensitivity, developmental processes, promoter and enhancer elements, site-binding related, and additional

elements (Fig. 5). Hormone- and stress-responsive elements were the most prevalent among all identified categories (Fig. 5). Furthermore, the classification of plant hormones was pertinent to the category, as *cis* elements were detected in *PgHAF* gene promoter sequences (Fig. 5). Consequently, all seven *PgHAT* genes exhibited the presence of abscisic acid responsiveness (ABRE), while elements responsive to gibberellin (P-box elements) and auxin (TGA-element) were detected in most promoters of *P. grandiflorus* HAT (Fig. 5). Notably, cis-element composition differed significantly between the paralogous gene pair, with a marked contrast in gibberellin and MeJA response elements within the *PgHAG* promoter (Fig. 5). Importantly, no elements responsive to cytokinin were found in these promoter regions (Fig. 5), aligning with earlier findings [32]. These findings suggest that *PgHATs* might influence phytohormone responsiveness, stress adaptability, and developmental growth.

Construction and Analysis of the HAT-Mediated Regulatory Network in P. grandiflorus

Functional association networks of gene family members can greatly benefit the study of gene function. A regulatory network involving *P. grandiflorus* HATs was developed using STRING (Table S6). The PCC values associated with the *P. grandiflorus* transcriptome were determined from various dynamic organs within the network. Within the figure, the red rectangles symbolize PgHATs, while the red edges indicate the PCC values and interaction scores of PgHATs that exceed 0.5 (Fig. S2). Among the seven *PgHATs*, four transcripts exhibited PCCs exceeding 0.5 for PGRA_08998 (PgHAC), which showed high expression levels in both early-stage flowers and leaves. Additionally, four genes displayed PCC values above 0.5 for PGRA_12289 (PgHAC), expressed across all tissues with low levels (Fig. S2). For PGRA_20711 (PgHAG), four members recorded PCC values above 0.5, indicating a constitutive expression pattern in all tissue types. Furthermore, PGRA_16872 (PgHAG) had PCCs above 0.5 for two positively co-expressed genes, which also demonstrated low transcription levels in young leaves (Fig. S2). Four members showed PCC values more than 0.5 for PGRA_08965 (PaHAF), with these genes being highly expressed in root tissues (Fig. S2). Three genes had PCCs exceeding 0.5 for PGRA_09720 (PgHAM), which was relatively low in expression in older leaves, stems, and roots (Fig. S2). Consequently, we hypothesized that PgHAG1 might cooperate with binding factors associated with the PgHATs, suggesting *PgHAT*s could possess self-regulating mechanisms.

PgHAT gene Functions in response to stress

Given the established role of HATs in plant stress responses, this study investigated the transcriptional profiles of *PgHAT* genes under cold, NaCl, and ABA stresses. Our results indicated that the expression of *PgHAT* genes generally rose during most cold treatment time points, with the exception of *PgHAC2*. Furthermore, *PgHAG3* showed a rapid increase in response to NaCl treatment but declined after one hour of exposure compared to the control group (Fig. 6). Notably, following NaCl treatment, there was an observed rise in the expression of the rice orthologs of these genes, suggesting that *HAT* genes are evolutionarily conserved across plants. Additionally, High positive correlations (all PCCs > 0.6) were observed in the transcriptional patterns of *PgHAC*s under different tissues and stress treatments, consistent with the possibility of functional redundancy among them. (Fig. S2). The findings supported

the hypothesis that HAC proteins found in monocot plants possess common ancestral genes. These genes might activate the biosynthesis of artemisinin-related genes through ectopic acetylation when exposed to cold and NaCl conditions. Nevertheless, it was unclear whether *PgHATs* were triggered by ABA treatment, in particular proteins in the downstream pathway (Fig. 6). Overall, the data suggested that different *PgHATs* play unique roles in response to various environmental stimuli, potentially activating genes linked to artemisinin biosynthesis to enhance the output derived from plants.

Discussion

Triterpenoid saponins are amphipathic compounds found in *Platycodon grandiflorus* root, to which they confer medicinal value. These compounds function in plant defense mechanisms and are also extensively utilized in the pharmaceutical, agrochemical, cosmetic, and food industries [25, 27, 28]. Primarily, the complete genome sequence of *P. grandiflorus* was published [29]. The HAT family typically forms complexes that exert significant regulatory influence over various cellular functions, for example, regulating the cell cycle, DNA replication and repair, transcriptional activation, and mRNA silencing [33, 34]. Therefore, functions of HATs have been linked to plant growth and responses to stress [35]. The discovery and analysis of *HAT*s have opened avenues for studying functional epi-regulators that respond to environment changes in medicinal plants. To date, the *HAT* gene family has been identified and examined in certain higher land plants, for instance *Arabidopsis thaliana* [11], *Oryza sativa* [36], litchi [13], barley [14], *Vitis vinifera* [15], tomato [16].

In this study, we characterized seven *HATs* within the genome of *P. grandiflorus* and examined the phylogeny, gene structures, transcriptions, *cis*-acting elements, predictions of sub-localization, and interaction analyses. Based on previous findings, *HAT*s can be categorized into HAC, HAF, HAM, and HAG families [11]. Additionally, the *PgHAGs* were further classified into *HAG1* and *HAG3* due to their conserved domain variations, while the *HAG2* group was evolutionarily lost (Fig. 1). Consequently, we discovered that the structures of *PgHAG1* and *PgHAG3* exhibited significant diversity, implying that they might possess different conserved functions. Each group of *PgHATs* contains unique conserved domains that support the rationale for this classification (Fig. 1). The properties of the PgHAT proteins exhibited a high degree of similarity to HATs from other plants [11, 12], suggesting an evolutionarily conserved functions of these PgHAT proteins.

Histone modifications are recognized for their significant influence on developmental growth and responses to stress, yet the functional investigation of *PgHATs* has not progressed as rapidly [10]. Earlier research indicated that analyzing orthologs could serve as a promising approach for predicting previously uncharacterized functional homologous genes across various species. As orthologs derive from a single ancestral gene, they typically retain conserved functions across descendant plant lineages. [37]. To enhance our understanding of the phylogenetic connections between *PgHATs* and other plant modules, NJ trees were used to illustrate the protein chains of HATs from *P. grandiflorus*, *Arabidopsis*, and rice. As anticipated, the NJ trees were categorized with five distinct clades (Fig. 2). Furthermore, considering that protein structure influences gene function, our analysis suggested that HATs from the 3

plant species could share homologous biological roles (Fig. 3). In *Arabidopsis*, AtHAC1, AtHAC5, and AtHAC12 act redundantly to regulate flowering through the repression of *FLC* transcription (*FLOWERING LOCUS C*) [21]. Consequently, the findings suggest that PgHAC, being homologous with AtHACs, might participate in the flowering of *P. grandiflorus* (Fig. 2). Based on conserved phylogeny and protein architecture, PgHAFs are predicted to be functional orthologs of AtHAF1 and AtHAF2 (Fig. 2). The knock-down variants of AtHAF1 demonstrated resistance to Agrobacterium infection [38]. Additionally, AtHAF2 was a key regulator modulating the expression of cold-responsive genes [39]. The PgHAMs exhibited significant homology with HAM proteins found in module plant. They grouped into the same groups of AtHATs and OsHATs, supported by high bootstrap values. Regarding the HAGs group, each species contains a single HAT (Fig. 2). Considering the involvement of AtHAM1 and AtHAM2 in developing both male and female gametophytes [40], we proposed that the closest orthologs of AaHAM might fulfill a similar role. Interestingly, only one PgHAT exists within the HAG2 group (Fig. 2), suggesting a unique evolutionary path for the HAT in the *P. grandiflorus* genome compared to those in module plants. Aside from HAG2, PgHAG1 and HAG3 show similarities to AtHAG1 and AtHAG3 (Fig. 2). It has been reported that *AtHAG1* is crucial for the processes of cell differentiation as well as the formation of leaf and flower organs (Servet C et al., 2010). The transcriptions of this gene differed across various tissues and organs; for instance, *PGRA_12289* (*HAC*) exhibited minimal expression in the entire plant, whereas the transcriptional activity of the *PGRA_08965* (*HAF*) gene was found to be significantly higher in the overall plantlet (Fig. 4). Furthermore, in silico localization predictions indicated that almost all PgHATs are nuclear-localized. (Table S4), suggesting that these PgHATs might influence epigenetic modifications in the cell nucleus, leading to transcriptional events. In the same way, AtHAG3 is also capable of modulating plant ABA responsiveness [41], which aligns with the abscisic acid responsiveness (TCA-element) identified in promoters of *PgHAG3* (Fig. 6). Collectively, the findings implied that the HATs in the *P. grandiflorus* genome exhibit a high degree of evolutionary conservation, making them an ideal subject for studies in medical plant identification and systematics.

It raises inquiries regarding the functional redundancy of multicopy *PgHAT* genes. In accord to evolutionary narratives, duplicated genes can enter into various selective conditions: neo-functionalization, in which one copy possesses an uncharacterized role; hypo-functionalization, where one copy experiences a reduction in transcription or functionality; non-functionalization, in which one copy becomes nonfunctional; or sub-functionalization, in which both copies evolve to perform distinct functions [42]. The variation in expression patterns or protein structures may suggest evolutionary relationships. Increasing evidence points to the divergence among multiple copy genes, which can be linked to the transcriptional patterns observed in the *PgHAT* gene sets. Our findings revealed a complex interplay of specific and overlapping *PgHAT* expressions across different tissues and organs (Fig. 5). The plants were categorized into early-leaves, old leaves, flowers, seeds, stems, and roots. The data indicated that seven *PgHAT* genes exhibited distinct expression levels at various developmental stages of these tissues. Furthermore, the shifts in expression patterns of gene pairs hinted at potential functions. For instance, *PGRA_12289* (*HAC*) showed minimal expression throughout the plants; in contrast, the *PGRA_08965* (*HAF*) gene displayed strong transcriptional signals in the overall plantlet

(Fig. 4), aligning with earlier findings that many identified HATs influence seed development [35]. Furthermore, HAG1 and GCN5 is a well-researched and functionally defined histone acetyltransferase that regulates pathways related to leaf and floral meristem patterning [30]. *PWA77834.1* (*HAG1*) had a comparable expression profile, showing high levels in both flowers and leaves (Fig. 5). The ADA2-GCN5 complex is known to recruit WOX11, which is responsible for controlling root [19]. Similarly, *PGRA_16872* (*HAG*) demonstrated significant expression in roots and is likely to be crucial for root development (Fig. 4).

It has been reported that *cis*-acting elements are linked to environmental stress [31]. We discovered 126 *cis*-acting elements in the *PgHAT* gene promoters. These are related with hormone responsiveness, environment response, light sensitivity, developmental processes, promoter and enhancer functions, site-binding characteristics, and additions (Fig. 5). Consequently, all 11 *PgHAT* genes exhibited ABRE, while P-box and TGA-element were found in nearly all promoters of *P. grandiflorus* HAT (Fig. 5). It is important to note that no elements responsive to cytokinins were found within these promoter regions (Fig. 5), which aligns with earlier findings [32]. The data indicated that *PgHATs* might influence the responsiveness to phytohormones, stress adaptability, and growth during development. Furthermore, a particularly intriguing finding was that the expression patterns of four AaHAC proteins (Fig. S2) exhibited a significant positive correlation with all PCCs exceeding 0.5 in both tissues we examined, implying a functional redundancy among the four HAC proteins in *P. grandiflorus*. This outcome supports our hypothesis that the AaHAC proteins in monocots originated from a common ancestral genes before the divergence of the plants.

Conclusions

The production of triterpenoid saponins, the bioactive compounds in the traditional medicinal plant *Platycodon grandiflorus*, is highly dependent on its cultivation environment. HAT activity influences chromatin modifications and mRNA transcription, playing a crucial role in how plants adapt to changing environments. Therefore, a thorough investigation of *PgHATs* throughout the development of *P. grandiflorus* offers valuable insights in the connection between HAT genes and environment adaptation in herbs. In this study, the entire genome of *P. grandiflorus was* analyzed, identifying and characterizing the *PgHAT* members. We also conducted a extensive analysis of their phylogenetic relationships with model plants, examined gene structures, transcriptions, *cis*-acting elements, predicted sub-localization, and performed interaction analyses. Our findings present a collection of *PgHAT* genes that play specific roles in responding to environmental changes, paving the way for future research in medicinal plants.

Methods

Characterization of PgHAT gene sequences

We collected data that included gene ID, protein/genomic sequences, and the conserved domain database pertaining to *P. grandiflorus* genome [29]. The *AtHAT* and *OsHAT* gene sequences were

downloaded from the Phytozome to facilitate the identification of all *HAT* genes present in *P. grandiflorus*, utilizing the AtHAT and OsHAT sequences as search queries. Through BLASTP searches ($E < 10^{-10}$) with the NCBI, we characterized seven putative PgHATs. Subsequently, we downloaded the Hidden Markov Model (HMM) profiles for HAG (PF00583), HAC (PF08214), HAF (PF09247), and HAM (PF01853) from Pfam. We employed a similar approach to examine the conserved domains within the seven putative PgHATs, confirming their classification within the HATs, with TBtools utilized for visualization [43]. In conclusion, we retrieved detailed information regarding the PgHATs, including CDS, pI, and MW from the NCBI.

Comparative sequence analysis and evolutionary tree reconstruction

The protein sequences of AtHAT, OsHAT, and PgHAT were subjected to a multiple sequence alignment and then entered into MEGA. This data was utilized to create an unrooted phylogeny employing NJ approach, along with 1000 bootstrap replicates [44].

Analysis of gene, protein structures and motif

The genome annotation file pertaining to *P. grandiflorus* was obtained from the NCBI database. The gene structures were examined utilizing Tbtools alongside GTF for *P. grandiflorus* and the NewickTree String associated with PgHATs. To predict the structures of PgHAT proteins, we employed SWISS-MODEL. Additionally, the tool MEME was utilized to investigate motifs, allowing for a selection of the maximum number of motifs.

Cis -acting elements in the PgHAT promoters

Sequences located 2000 bp upstream of *PgHAT*s were retrieved from NCBI to pinpoint *cis*-elements within the potential promoter areas, utilizing PlantCARE software. The *cis*-acting elements were categorized based on the various function and visualized with the TBtools.

Gene expression profiles analysis

Details regarding the expressions of *PgHAT*s across six different tissue types of *P. grandiflorus* were obtained from an existing database. The tissues analyzed comprised leaf, root, stem, pistil, stamen, petal, seed, and sepal, as outlined in Table S5. The preparation of all RNA sequencing libraries was carried out following the methodology described in a prior study [4]. To visualize the gene expression data, the HemI tools was utilized [45].

Declarations

Authors' contributions

Y.L. was responsible for initiating and designing the experiments. The experiments were carried out by Y.X. and J.W., who also gathered the data. Y.L. conducted data-set analysis and write the manuscript,

incorporating feedback from Y.X. and J.W. Each author has reviewed the manuscript and consented to its submission for publication.

Funding

This work was supported by the grants from Guidance project of scientific research plan of Hubei Provincial Department of Education (No. B2024022) to Jingzhi Wan.

Availability of data and materials

The research data are openly provided in this article and the supplementary information. Genomic sequences for *Arabidopsis thaliana*, *Oryza sativa*, and *Platycodon grandiflorus* can be found in both the Ensemble Plants database and the National Genomics Data Center.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D. & Zhao, Y. *SnapShot: histone modifications* (Cell, 2014).
2. Gong, F. & Miller, K. M. Mammalian DNA repair: HATs and HDACs make their mark through histone acetylation. *Mutat. Research/Fundamental Mol. Mech. Mutagen.* **750**, 23–30 (2013).
3. Zhou, C. et al. Accessible chromatin regions and their functional interrelations with gene transcription and epigenetic modifications in sorghum genome. *Plant. Commun.* **2**, 100140–100140 (2020).
4. Zhou, C. et al. Genome-Wide Identification and Characterization of Main Histone Modifications in Sorghum Decipher Regulatory Mechanisms Involved by mRNA and Long Noncoding RNA Genes. *J. Agric. Food Chem.* **69**, 2337–2347 (2021).
5. Lu, L., Chen, X., Sanders, D., Qian, S. & Zhong, X. *High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in Arabidopsis and rice* (Epigenetics, 2015).
6. Shen, Y., Wei, W. & Zhou, D-X. *Histone Acetylation Enzymes Coordinate Metabolism and Gene Expression* (Trends Plant Sci, 2015).

7. Chen, X., Ding, A. B. & Zhong, X. *Functions and mechanisms of plant histone deacetylases* (Sci China Life Sci, 2020).
8. Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Rep. (2002).
9. Chen, Z. J. & Tian, L. *Roles of dynamic and reversible histone acetylation in plant development and polyploidy* (Biochim Biophys Acta, 2007).
10. Liu, X., Yang, S., Yu, C. W., Chen, C. Y. & Wu, K. in *Chapter Six - Histone Acetylation and Plant Development*. 173–199 (eds Lin, C. & Luan, S.) (Academic, 2016).
11. Pandey, R. et al. Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res.* **30**, 5036–5055 (2002).
12. Liu, X. et al. Histone acetyltransferases in rice (Oryza sativaL.): phylogenetic analysis, subcellular localization and expression. *BMC Plant Biol.* **12**, 145 (2012).
13. Peng, M. et al. *Genome-Wide Identification of Histone Modifiers and Their Expression Patterns during Fruit Abscission in Litchi* (Front Plant Sci, 2017).
14. Papaefthimiou, D., Likotrafiti, E., Kapazoglou, A., Bladenopoulos, K. & Tsaftaris, A. *Epigenetic chromatin modifiers in barley: III. Isolation and characterization of the barley GNAT-MYST family of histone acetyltransferases and responses to exogenous ABA* (Plant Physiol Biochem, 2010).
15. Aquea, F., Timmermann, T. & Arce-Johnson, P. *Analysis of histone acetyltransferase and deacetylase families of Vitis vinifera* (Plant Physiol Biochem, 2010).
16. Aiese Cigliano, R. et al. *Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles* (BMC Genomics, 2013).
17. Vlachonasios, K. E., Thomashow, M. F. & Triezenberg, S. J. *Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect Arabidopsis growth, development, and gene expression* (Plant Cell, 2003).
18. Kornet, N. & Scheres, B. *Members of the GCN5 histone acetyltransferase complex regulate PLETHORA-mediated root stem cell niche maintenance and transit amplifying cell proliferation in Arabidopsis* (Plant Cell, 2009).
19. Zhou, S. et al. *Rice Homeodomain Protein WOX11 Recruits a Histone Acetyltransferase Complex to Establish Programs of Cell Proliferation of Crown Root Meristem* (Plant Cell, 2017).
20. Bertrand, C., Bergounioux, C., Domenichini, S., Delarue, M. & Zhou, D-X. *Arabidopsis histone acetyltransferase AtGCN5 regulates the floral meristem activity through the WUSCHEL/AGAMOUS pathway* (J Biol Chem, 2003).
21. Deng, W. et al. *Involvement of the histone acetyltransferase AtHAC1 in the regulation of flowering time via repression of FLOWERING LOCUS C in Arabidopsis* (Plant Physiol, 2007).
22. Engstrom, E. M. et al. *Arabidopsis homologs of the petunia hairy meristem gene are required for maintenance of shoot and root indeterminacy* (Plant Physiol, 2011).

23. Xiao, J. et al. Requirement of histone acetyltransferases HAM1 and HAM2 for epigenetic modification of FLC in regulating flowering in Arabidopsis. *J. Plant Physiol.* **170**, 444–451 (2013).
24. Li, C., Xu, J., Li, J., Li, Q. & Yang, H. *Involvement of Arabidopsis histone acetyltransferase HAC family genes in the ethylene signaling pathway* (Plant Cell Physiol, 2014).
25. Zhang, L. et al. Platycodon grandiflorus – An Ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol.* **164**, 147–161 (2015).
26. <51CTO□□-. Sed□awk□□□(□□□)_IT168□□_P86.pdf&gt.
27. Moses, T., Thevelein, J. M., Goossens, A. & Pollier, J. Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. *Phytochemistry* **108**, 47–56 (2014).
28. Thimmappa, R. et al. Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **65**, 225–257 (2014).
29. Jia, Y. et al. A Chromosome-Level Reference Genome of Chinese Balloon Flower (Platycodon grandiflorus). *Front. Genet.* Volume 13–2022. (2022).
30. Servet, C., Conde e Silva, N. & Zhou, D-X. *Histone acetyltransferase AtGCN5/HAG1 is a versatile regulator of developmental and inducible gene expression in Arabidopsis* (Mol Plant, 2010).
31. Li, Z. et al. AcYABBY4Genome-Wide Analysis of the YABBY Transcription Factor Family in Pineapple and Functional Identification of Involvement in Salt Stress. *Int. J. Mol. Sci.* (2019).
32. Xing, G. et al. Genome-wide investigation of histone acetyltransferase gene family and its responses to biotic and abiotic stress in foxtail millet (Setaria italica [L.] P. Beauv). *BMC Plant Biol.* **22**, 292 (2022).
33. Mai, A. et al. Identification of 4-hydroxyquinolines inhibitors of p300/CBP histone acetyltransferases. *Bioorg. Med. Chem. Lett.* **19**, 1132–1135 (2009).
34. Kikuchi, H. & Nakayama, T. GCN5 and BCR signalling collaborate to induce pre-mature B cell apoptosis through depletion of ICAD and IAP2 and activation of caspase activities. *Gene* **419**, 48–55 (2008).
35. Wang, Z., Cao, H., Chen, F. & Liu, Y. The roles of histone acetylation in seed performance and plant development. *Plant Physiol. Biochem.* **84**, 125–133 (2014).
36. Song, X. et al. Rice RNA-dependent RNA polymerase 6 acts in small RNA biogenesis and spikelet development. *Plant. J.* **71**, 378–389 (2012).
37. Zhao, Y. et al. Genome-Wide Identification and Analysis of the AP2 Transcription Factor Gene Family in Wheat (Triticum aestivum L.). 10. (2019).
38. Crane, Y. M. & Gelvin, S. B. *RNAi-mediated gene silencing reveals involvement of Arabidopsis chromatin-related genes in Agrobacterium-mediated root transformation* (Proc Natl Acad Sci U S A, 2007).
39. Pavangadkar, K., Thomashow, M. F. & Triezenberg, S. J. *Histone dynamics and roles of histone acetyltransferases during cold-induced gene regulation in Arabidopsis* (Plant Mol Biol, 2010).
40. Nelissen, H. et al. *The elongata mutants identify a functional Elongator complex in plants with a role in cell proliferation during organ growth* (Proc Natl Acad Sci U S A, 2005).

41. Chen, Z. et al. *Mutations in ABO1/ELO2, a subunit of holo-Elongator, increase abscisic acid sensitivity and drought tolerance in Arabidopsis thaliana* (Mol Cell Biol, 2006).
42. Duarte, J. M. et al. *Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis* (Mol Biol Evol, 2006).
43. Chen, C. et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant.* **13**, 1194–1202 (2020).
44. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms* (Mol Biol Evol, 2018).
45. Deng, W., Wang, Y., Liu, Z., Cheng, H. & Xue, Y. *HemI: a toolkit for illustrating heatmaps* (PLoS One, 2014).

Figures

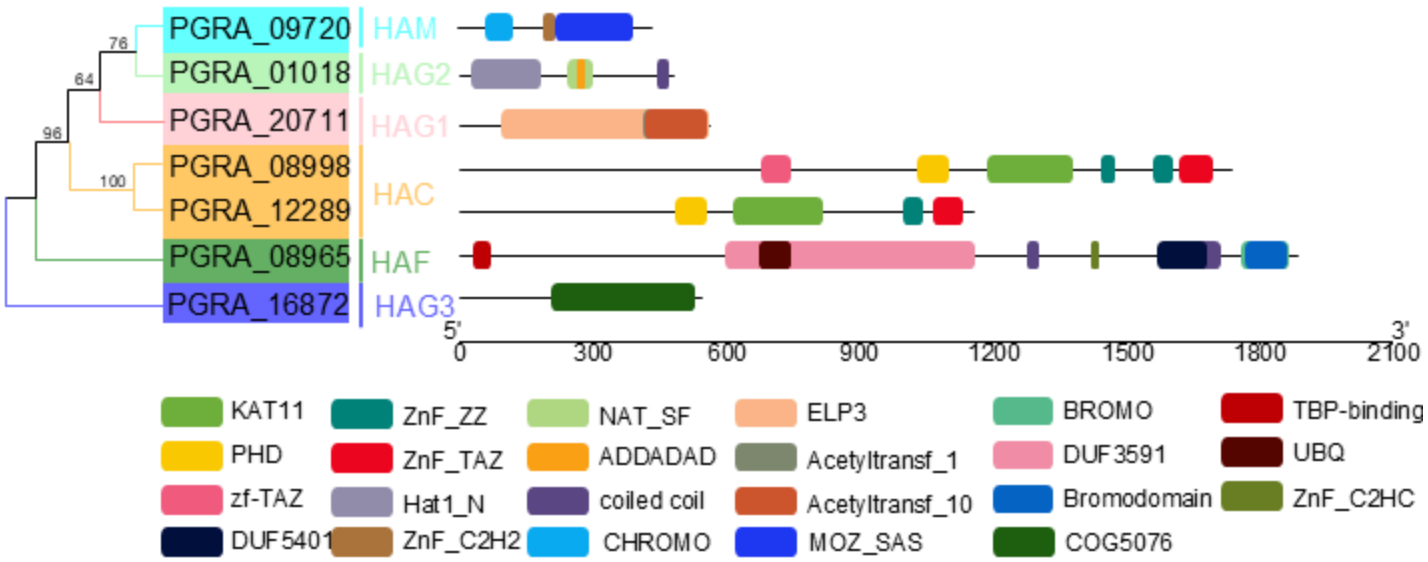

Fig. 1 Conserved domain analysis of the *PgHAT* gene family.
The seven *PgHAT* can be divided into six groups (HAC, HAM, HAF, HAG1, HAG2, and HAG3) based on conserved domain analysis. Individual conserved domains are indicated by different colored boxes.

Figure 1

See image above for figure legend.

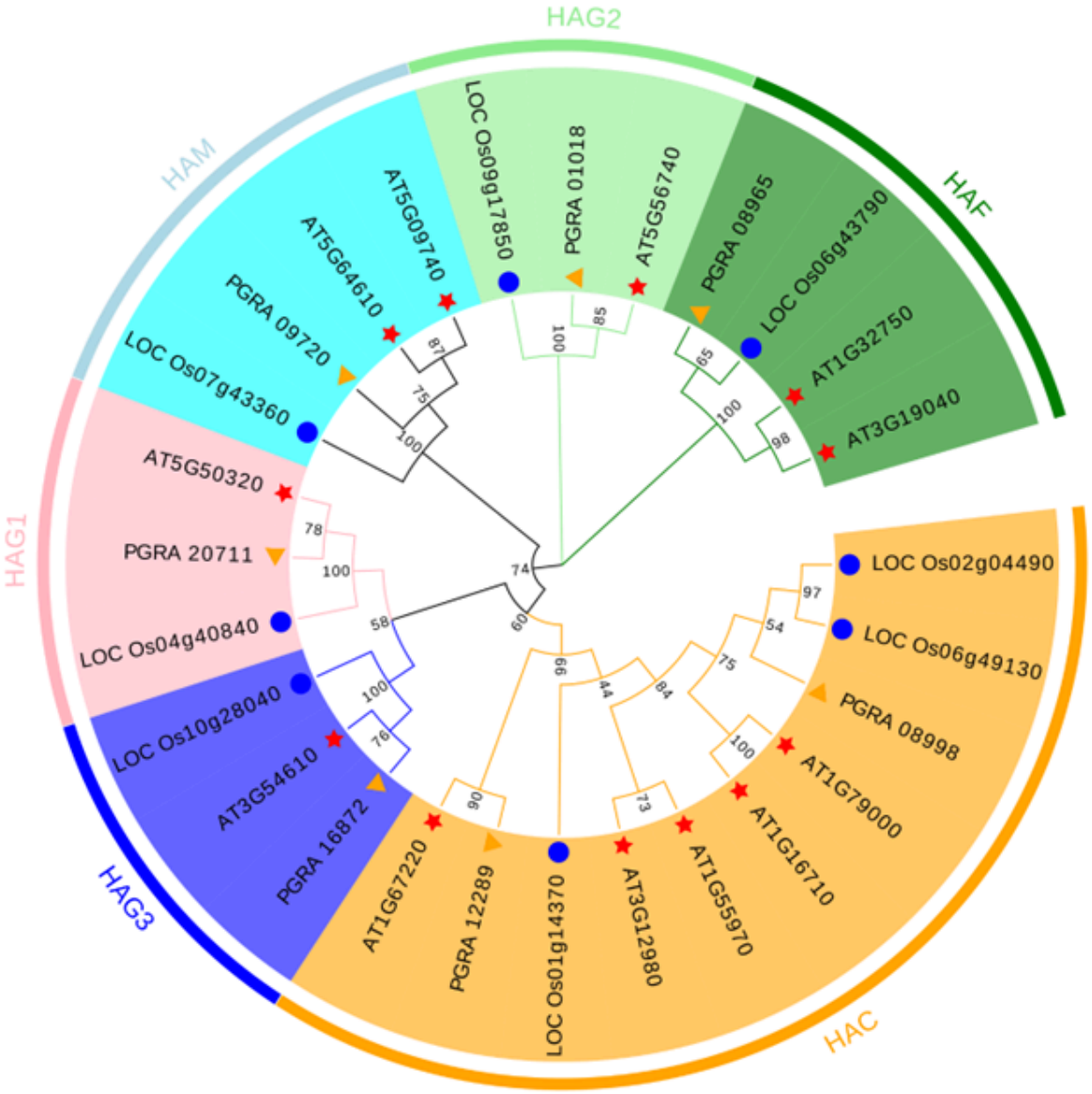

Fig. 2 Phylogenetic tree of HAT proteins from *Arabidopsis thaliana*, *Oryza sativa* and *Platycodon grandiflorus* constructed by the neighbor-joining method in MEGA-X.
The numbers at nodes represent bootstrap values after 1000 iterations. Each group is indicated by a different color. Stars represent *A. thaliana*, circles represents *O. sativa*, and triangles represent *P. grandiflorus*.

Figure 2

See image above for figure legend.

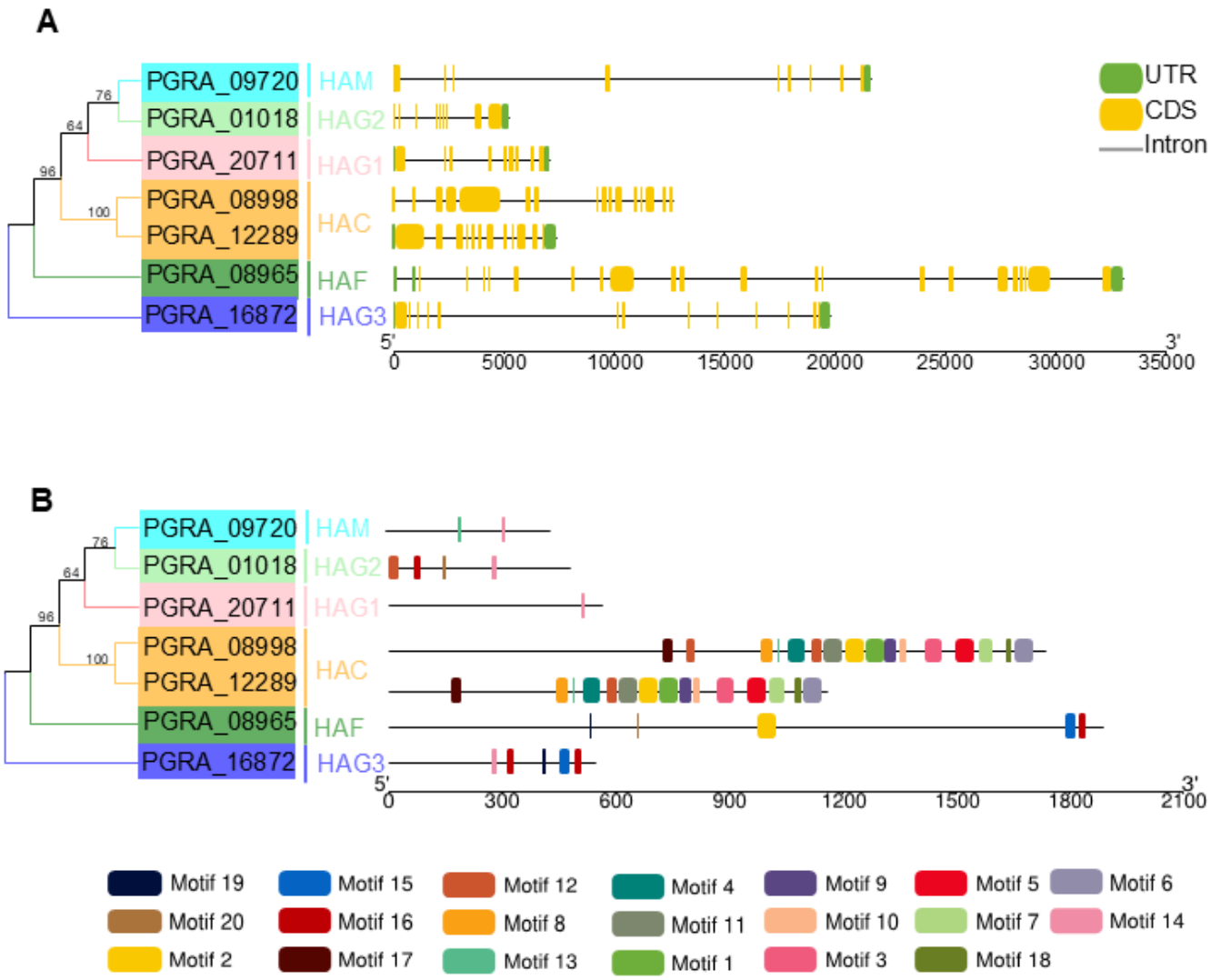

Fig. 3 Genetic and motif structures of 7 *PgHAT* genes.
(A) Exons, introns, and untranslated regions (UTR) are indicated by yellow frames, gray lines, and green frames on the right, respectively. The number on the gray line represents the number of introns. (B) Schematic representation of conserved motifs in the PgHATs. Different colored frames represents different protein motifs, and each motif has its own number.

Figure 3

See image above for figure legend.

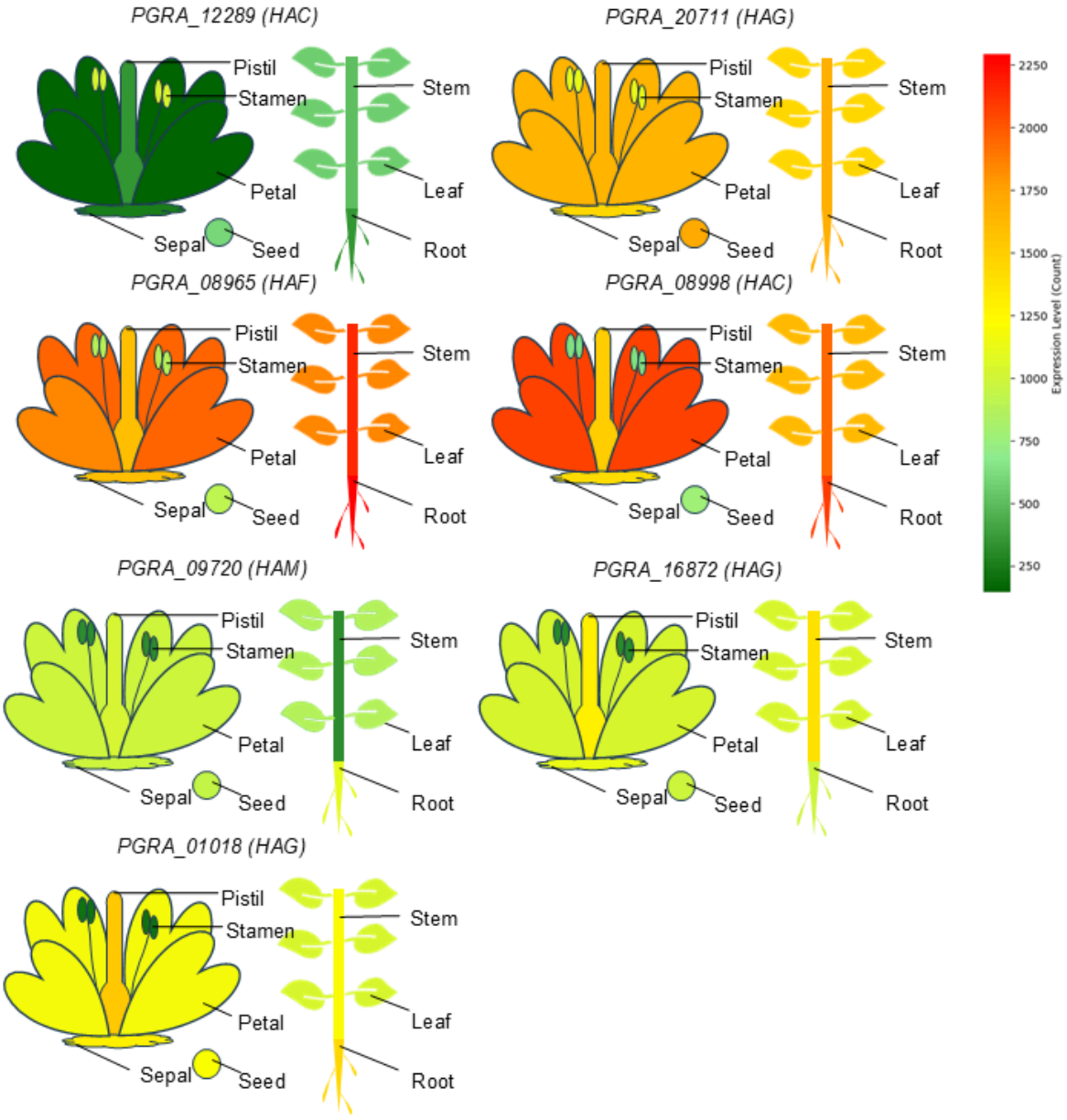

Fig. 4 Differential expression of representative *PgHATs* in different tissues.
The mean expression values were visualized by Tbtools; red represents a high expression level and green represents a low expression level. These RNA-seq data different tissues was download from public database. Red represents a high expression level and green represents a low expression level. These RNA-seq data of different tissues was download from public database.

Figure 4

See image above for figure legend.

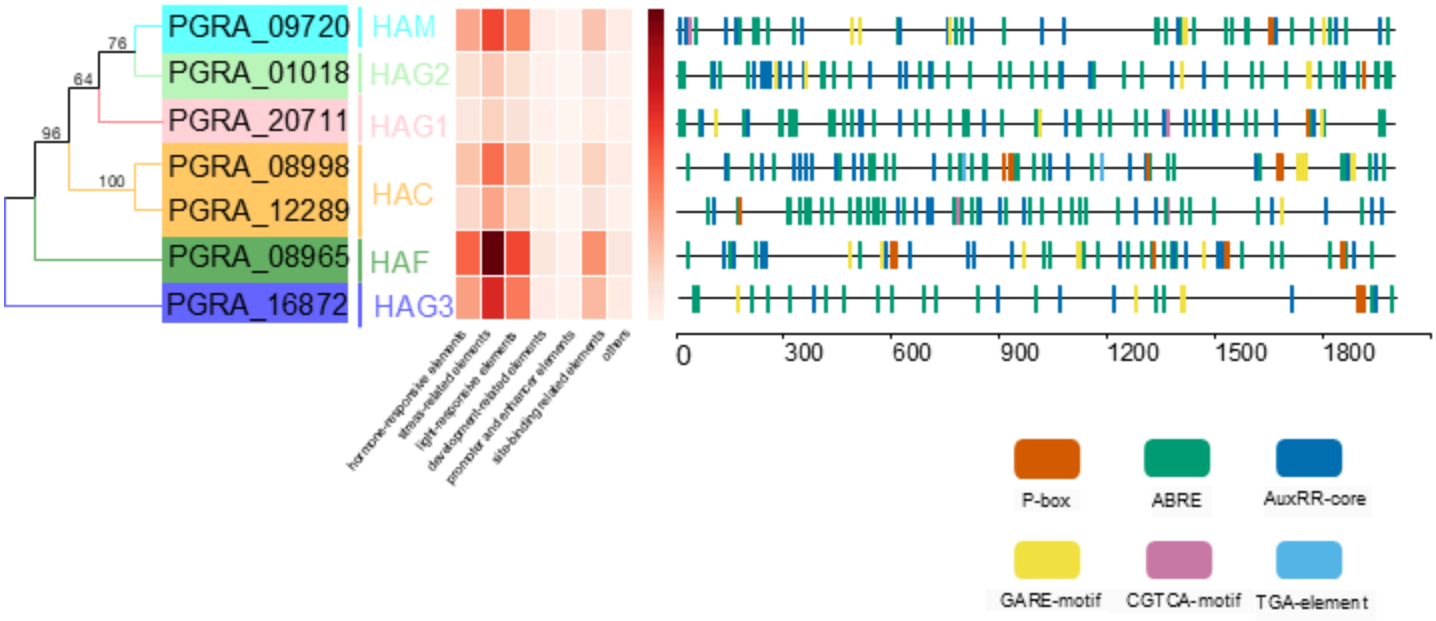

Fig. 5 Prediction of cis-acting elemets in the AaHAT promoters.
(A) Number of cis-acting elements detected in the promoter region of each AaHAT gene; elements were divided into seven types.
(B) Type, quantity, and position of hormone-responsive elements in AaHATs' promoters. MeJA responsiveness (CGTCA-motif), auxin-responsive (TGA-box, TGA-element), abscisic acid responsiveness (ABRE, TCA-element), gibberellin-responsive (GARE-motif, P-box, TATC-box).

Figure 5

See image above for figure legend.

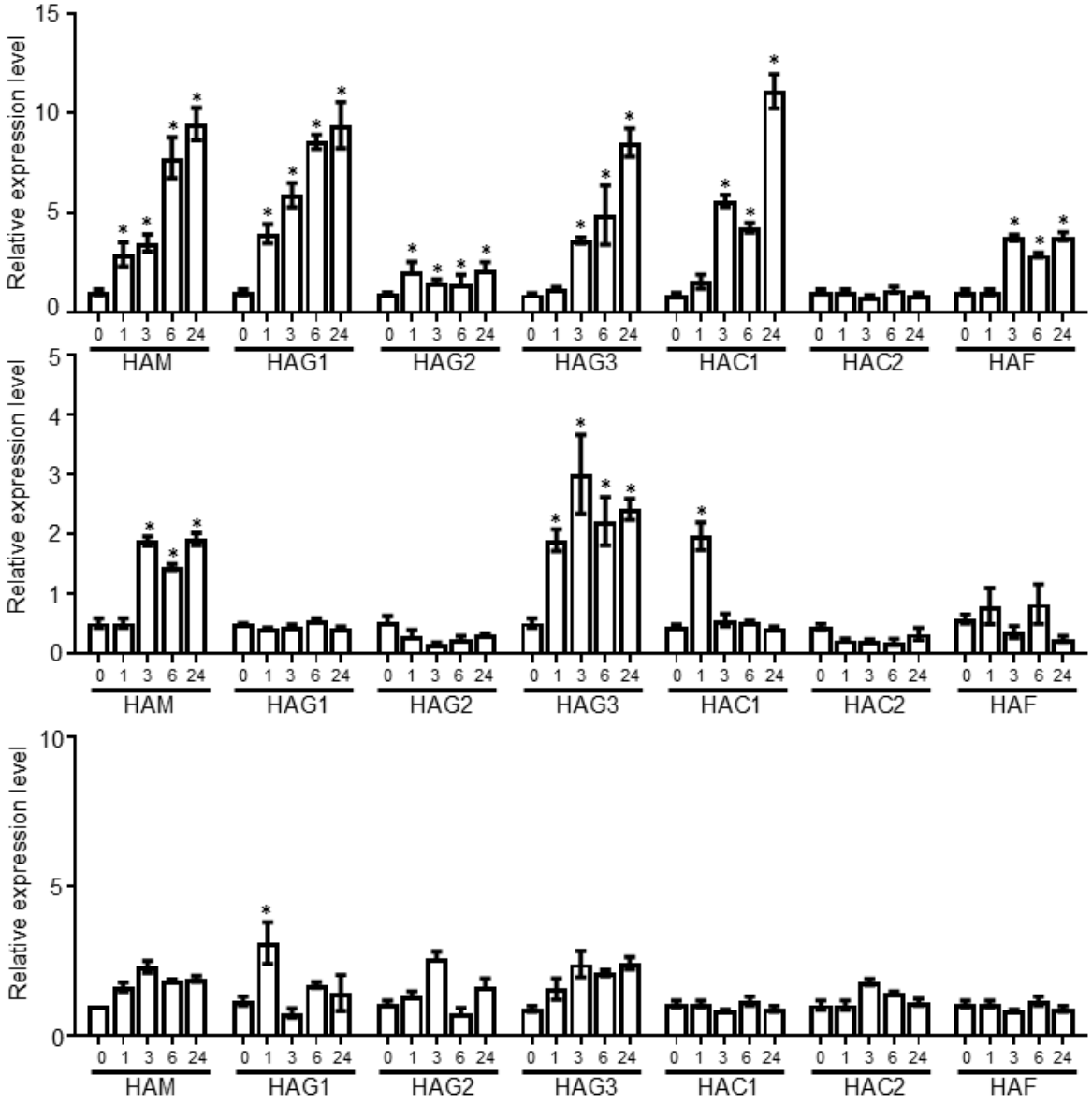

Fig. 6 qRT-PCR analyses of *PgHAT* genes under the treatments of three different abiotic stresses. The transcript levels of *PgHAT* genes under cold, NaCl and ABA treatments and mock were normalized. The average ± SD values from three biological repeats are shown. * $P < 0.05$, Student's *t*-test.

Figure 6

See image above for figure legend.

Supplementary Files

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

- SupplementaryFiles.zip