Response of phytohormone mediated plant homeodomain (PHD) family to abiotic stress in upland cotton (Gossypium hirsutum spp.)

Huanhuan Wu
Chinese Academy of Agricultural Sciences Cotton Research Institute
Lei Zheng
Chinese Academy of Agricultural Sciences Cotton Research Institute
Ghulam Qanmber
Chinese Academy of Agricultural Sciences Cotton Research Institute
Mengzhen Guo
Chinese Academy of Agricultural Sciences Cotton Research Institute
Zhi Wang
Chinese Academy of Agricultural Sciences Cotton Research Institute
Zuoren Yang (yangzuoren4012@163.com)
Chinese Academy of Agricultural Sciences Cotton Research Institute https://orcid.org/0000-0003-0337-9573

Research article

Keywords: Cotton, PHD, Transcription factor, Phytohormone, Stress tolerant, Co-expression network, Transcriptome analysis

Posted Date: February 4th, 2020

DOI: https://doi.org/10.21203/rs.2.22544/v1

Version of Record: A version of this preprint was published on January 6th, 2021. See the published version at https://doi.org/10.1186/s12870-020-02787-5.

Abstract

Background

The sequencing and annotations of cotton genomes provide strong theoretical support to reveal more physiological phenomena and functions. Plant homeodomain (PHD) protein family have been reported to be involved in regulating diverse biological processes in plants. However, their functions have not yet been carried out in cotton.

Results

In this study, we performed a genome-wide analysis of the PHD genes in cotton, including the chromosomal location, phylogenetic relationship, gene structure, and conserved domains. Using a phylogenetic analysis, we divided the 297 PHD genes into five subgroups. The GhPHDs were unevenly distributed across all 26 chromosomes in upland cotton, and whole genome duplication events analyses showed that purifying selection might contributed greatly to the maintenance of function in the GhPHD family. Expression pattern analysis based on RNA-seq data and qRT-PCR results showed that the most of GhPHD genes have significant tissue-specific spatial and temporal expression patterns, indicating GhPHD have multiple functions in growth and development. We further summarized the cis -acting elements in response to abiotic stresses and plant hormones, and treated cotton seedlings with abiotic stresses and plant hormones, respectively. Then, GhPHD gene expression level were detected by qRT-PCR, which indicated that GhPHD could response to stresses and plant hormones. Co-expression network analysis also indicated that GhPHDs were essential for plant growth and development, and phytohormone mediate GhPHD response to abiotic stress can improve plant tolerance to adverse environment.

Conclusion

This study provides useful information to facilitate the further study of the function of the GhPHD family.

Background

Plants are often subjected to a wide range of abiotic and biotic stresses. Abiotic stress conditions include heat, cold, drought, and salinity, whereas biotic stresses are mainly derived from bacteria, fungi, viruses, and insects. Both abiotic and biotic stresses significantly reduce crop productivity and quality worldwide [1, 2]. In order to adapt in such unfavorable environment, plants have evolved mature mechanisms to perceive the stress signals and achieve optimal growth response [3]. The main phytohormones produced by plant are auxin, gibberellin (GA), abscisic acid (ABA), ethylene (ET), salicylic acid (SA), jasmonate (JA), cytokinin (CK), brassinosteroid (BR), and strigolactone (SL). Among these, GA, IAA, SA, JA, and BL are known to play important roles in mediating the defense response of plants to biotic and abiotic stresses [4–7]. Upland cotton (G. hirsutum) is an important economic crop and is a source of natural fiber, protein,

and vegetable oil. Cotton is usually planted on saline-alkali and arid areas accompanied by high temperature, cold, and other adverse environmental conditions.

Zinc finger protein motifs are part of many protein families and widely distributed in eukaryotic organisms. The term “zinc finger” represents the sequence motifs in which cysteines and/ or histidines coordinate the zinc atom(s) to form local peptide structures that are required for their specific functions. The “finger” structural motifs have been divided into different types, such as TFIIIA-type zinc finger (EPF1, SUPERMAN) [8, 9], the WRKY family (WRKY1, 2, and 3), GATA1-type protein (NTL1) [10, 11], the Dof family (Dof1) [12, 13], RING-finger type (COP1) [14], PHD-finger family (AtHAT3.1 and Zmhox1a) [15, 16], the LIM family (SF3) [17, 18], and other uncategorized types. The plant homeodomain (PHD) zinc fingers are small reader domains found in several chromatin-binding proteins. In plants, PHD proteins are usually zinc finger proteins with one or more PHD domains, which have Cys4-His-Cys3 zinc-binding motif consisting of about 60 amino acids [19]. It is worth noting that the number of amino acids between cysteine and histidine or between cysteine residues in the PHD domain are usually conserved, while the second before the penultimate cysteine amino acid residues is usually an aromatic amino acids, such as tryptophan [20].

Since the first PHD protein HAT3.1 was found in Arabidopsis, more PHD proteins have been identified to involved in some physiological and biochemical processes in fungi, animals, and plants, regulating chromatin structure and transcription [21]. For example, NURF (nucleosome remodeling factor) is an ATP-dependent chromatin-remodeling complex containing ISWI and mediates direct preferential association with H3K4me3 tails to regulate chromatin remodeling and state [22]. In Arabidopsis, the PHD protein MMD1/DUET is necessary to regulate the progress of meiotic prophase I chromosome [23]. MMD1 is preferentially expressed during male meiosis and may be involved in the regulation of gene expression during meiosis. The mmd1 mutant plant exhibits meiotic defects and monocytes appear normal until the diakinesis stage show signs of apoptosis, which subsequently trigger the death of male meiotic cells [24]. VIM proteins are involved in the regulation of chromatin state with histone and regulate epigenetic gene silencing in the genome by the coordinated modulation of DNA methylation and histone modification status with MET1 [25]. Studies have shown that mammal PHD proteins UHRF1/2 are necessary for methylation maintenance and can interact with MPG protein [26]. UHRF1 can mediate the crosstalk between DNA methylation and H3K9 methylation at the level of DNA methylation maintenance, and UHRF1 targets DNMT1 (a maintenance methyltransferases) by binding either hemi-methylated CpG or H3K9me2/3 to ensure high fidelity DNA maintenance methylation during the global epigenetic reprogramming of oocytes and early embryos [27, 28]. PHD protein ING2 is a binding module and nuclear receptor of phosphoinositides, which play key roles in cytoplasmic signal transduction pathways, and ING2-phosphoinositides interactions could directly regulate the DNA damage signaling [29]. The cellular kinase MEKK1 has PHD-dependent E3 ubiquitin ligase activity, and the close structural relationship between PHD domains and RING fingers suggests that many other PHD proteins may also be involved in the regulation of ubiquitination [30].

Many PHD proteins may be involved in regulating plant response to abiotic stress. For example, six soybean PHD proteins have different expression levels for salt, cold, drought, and ABA treatment [31]. The PHD proteins could be applied to alter the various developmental processes in plants. The effect of VIN3 is to adjust in mediating a proper vernalization response in Arabidopsis, which is a long-term cold induced and is necessary for PRC2-mediated trimethylation of H3K27me3 at the FLC locus [32]. AL PHD-PRC1 complexes constructed around H3K4me3 causes seed development genes to transition from the H3K4me3-associated active to the H3K27me3-associated repressive transcription state [33]. GSR1 (germostatin resistance locus 1) is a member of the auxin-mediated genetic network for seed germination. GSR1 physically interacts with the transcriptional repressor ARF16 and attenuates the interaction intensity of IAA17/ARF16 by directly interacting with IAA17 to release ARF16. IAA17-GSR1-ARF16 may form auxin signaling pathway in which ARF10 and ARF16 are downstream components, which are necessary for auxin to regulate seed germination and dormancy [34].

At present, nine plant species have identified the PHD protein family, such as Arabidopsis (Arabidopsis thaliana), poplar (Populus trichocarpa), maize (Zea mays), moso bamboo (Phyllostachys edulis), and five Rosaceae species: apple (Malus × domestica), peach (Prunus persica), pear (Pyrus bretschneideri), mei (Prunus mume), and strawberry (Fragaria vesca). However, compared with these species, the research works on cotton PHD protein family have not yet been carried on. Recently, the availability of the completed genome sequence and annotation of G. hirsutum [35], G. arboreum [36], and G. raimondii [37] provide us with an excellent opportunity to identify and characterize PHD transcription factors in cotton species. In this study, we performed a genome-wide analysis, tissue expression pattern analysis, expression levels analysis under different stresses and hormone treatments, and co-expression network analysis of PHD genes in upland cotton, which revealed that phytohormone mediate PHD genes in response to abiotic stress.

Results

Genome-wide identification of PHD proteins in cotton

Whole genome sequences of three cotton species (G. hirsutum, G. arboreum, and G. raimondii) were used to identify PHD proteins. A total of 108 PHD proteins in G. hirsutum, 52 in G. arboreum, and 55 in G. raimondii were identified in three cotton species (Table S1). Besides this, 39 and 43 PHD proteins in Arabidopsis and rice were also identified respectively (Table S1). Among these 108 GhPHD proteins, 56 members belong to At sub-genome and 52 from the Dt sub-genome (Table S1). The biophysical characteristic of GhPHDs are provided in Table 1. The length of GhPHD proteins ranged from 159 aa (GhPHD28) to 2231 aa (GhPHD39), with an average length of 741 aa (Table 1). The molecular weight of GhPHD proteins ranged from 17.76 kD (GhPHD28) to 247.42 kD (GhPHD39) with an average value of 93.09 kD (Table 1). The isoelectric point (pI) of GhPHD proteins ranged from 4.58 (GhPHD38) to 10.41 (GhPHD103) with an average value of 6.89 (Table 1). We predicted the subcellular localization of 108 GhPHD proteins and found that 93 GhPHD proteins were located in the nucleus, while ten were in the cytoplasm and five were located extracellular (Table 1).

Table 1
Physicochemical parameters of 108 GhPHD genes in G. hirsutum genome

Name	Protein length (aa)	Molecular weight (kDa)	Charge	Isoelectric point	Grand average of hydropathy	Subcellular localization
GhPHD1	217	24.915	5	7.895	-0.694	Nuclear
GhPHD2	1,033	114.441	32.5	8.49	-0.274	Nuclear
GhPHD3	1,030	114.182	36	8.594	-0.306	Nuclear
GhPHD4	815	90.024	5	6.895	-0.323	Nuclear
GhPHD5	1,303	144.878	-9.5	6.002	-0.713	Nuclear
GhPHD6	700	79.363	-4.5	6.21	-0.308	Nuclear
GhPHD7	700	79.363	-4.5	6.21	-0.308	Nuclear
GhPHD8	345	39.474	12	8.648	-0.573	Nuclear
GhPHD9	251	28.253	-8	4.891	-0.621	Nuclear
GhPHD10	786	86.704	-35.5	4.631	-0.966	Nuclear
GhPHD11	216	24.846	8.5	8.262	-0.789	Nuclear
GhPHD12	375	42.862	5.5	7.542	-0.708	Nuclear
GhPHD13	237	26.757	-4	5.421	-0.596	Nuclear
GhPHD14	959	104.731	-3	6.227	-1.126	Nuclear
GhPHD15	733	82.869	40.5	9.936	-0.907	Nuclear
GhPHD16	252	28.482	-8.5	4.84	-0.717	Nuclear
GhPHD17	1,680	189.096	46.5	8.1	-0.669	Nuclear
GhPHD18	252	28.35	-7.5	4.894	-0.661	Nuclear
GhPHD19	238	27.277	4.5	7.669	-0.648	Cytoplasmic
GhPHD20	493	55.306	5	7.03	-0.485	Nuclear
GhPHD21	600	67.261	5.5	7.049	-0.611	Nuclear
GhPHD22	1,084	122.987	34	8.276	-0.574	Nuclear
GhPHD23	253	28.577	-5.5	5.132	-0.736	Nuclear
GhPHD24	259	29.239	-7.5	4.915	-0.708	Nuclear
GhPHD25	224	25.723	6.5	8.087	-0.682	Cytoplasmic

Name	Protein length (aa)	Molecular weight (kDa)	Charge	Isoelectric point	Grand average of hydropathy	Subcellular localization
GhPHD26	870	95.472	3	6.876	-0.472	Nuclear
GhPHD27	1,358	154.506	35	7.891	-0.677	Nuclear
GhPHD28	159	17.763	0	6.496	-0.666	Extracellular
GhPHD29	733	80.954	22	8.271	-0.664	Nuclear
GhPHD30	1,247	141.67	24	7.655	-0.43	Nuclear
GhPHD31	949	104.907	2.5	6.779	-0.416	Nuclear
GhPHD32	1,618	180.35	45	8.404	-0.446	Nuclear
GhPHD33	1,618	180.725	41.5	8.289	-0.442	Nuclear
GhPHD34	216	24.95	8.5	8.399	-0.783	Nuclear
GhPHD35	321	35.88	-4.5	5.599	-0.049	Extracellular
GhPHD36	822	88.768	2	6.651	-0.539	Nuclear
GhPHD37	1,305	143.316	-33.5	4.951	-0.624	Nuclear
GhPHD38	705	78.949	22	8.309	-0.315	Extracellular
GhPHD39	2,231	247.421	-36	5.321	-0.444	Nuclear
GhPHD40	226	25.942	6	8.086	-0.788	Nuclear
GhPHD41	1,685	187.671	-4.5	6.321	-0.389	Nuclear
GhPHD42	1,239	138.122	-0.5	6.487	-0.735	Nuclear
GhPHD43	253	28.585	-6	5.13	-0.76	Nuclear
GhPHD44	531	58.455	-4	5.77	-0.564	Nuclear
GhPHD45	389	44.625	26.5	9.906	-0.405	Nuclear
GhPHD46	803	90.452	-16.5	5.132	-0.836	Nuclear
GhPHD47	851	94.805	-26	4.895	-1.011	Nuclear
GhPHD48	212	23.802	26	10.41	-0.745	Nuclear
GhPHD49	1,019	116.396	18	7.609	-0.584	Nuclear
GhPHD50	655	74.308	11	7.433	-0.207	Cytoplasmic
GhPHD51	1,091	124.448	45	8.516	-0.59	Nuclear

Name	Protein length (aa)	Molecular weight (kDa)	Charge	Isoelectric point	Grand average of hydropathy	Subcellular localization
GhPHD52	237	26.997	-3.5	5.244	-0.666	Nuclear
GhPHD53	216	24.723	6.5	8.049	-0.785	Nuclear
GhPHD54	716	81.912	-38.5	4.581	-1.1	Nuclear
GhPHD55	1,367	152.049	3	6.689	-0.451	Nuclear
GhPHD56	254	28.492	-6.5	5.136	-0.576	Cytoplasmic
GhPHD57	217	24.929	5	7.895	-0.684	Nuclear
GhPHD58	1,031	114.246	33.5	8.459	-0.299	Nuclear
GhPHD59	1,031	113.878	36.5	8.592	-0.306	Nuclear
GhPHD60	1,299	144.45	-11.5	5.886	-0.711	Nuclear
GhPHD61	290	32.821	-8.5	4.832	-0.419	Nuclear
GhPHD62	216	24.835	8.5	8.262	-0.789	Nuclear
GhPHD63	699	79.319	-4.5	6.21	-0.283	Nuclear
GhPHD64	345	39.311	13	8.745	-0.544	Nuclear
GhPHD65	1,084	123.101	34	8.275	-0.579	Nuclear
GhPHD66	237	26.615	-2	5.973	-0.561	Nuclear
GhPHD67	945	103.676	-2	6.329	-1.136	Nuclear
GhPHD68	684	77.195	23	9.073	-0.859	Nuclear
GhPHD69	733	83.057	34.5	9.713	-0.91	Nuclear
GhPHD70	252	28.414	-8.5	4.84	-0.697	Nuclear
GhPHD71	1,731	194.29	53	8.282	-0.675	Nuclear
GhPHD72	252	28.35	-7.5	4.894	-0.661	Nuclear
GhPHD73	224	25.664	7.5	8.248	-0.774	Cytoplasmic
GhPHD74	367	41.019	5	7.341	-0.49	Nuclear
GhPHD75	601	67.379	2	6.696	-0.611	Nuclear
GhPHD76	241	27.506	-0.5	6.269	-0.502	Extracellular
GhPHD77	253	28.677	-5.5	5.139	-0.752	Nuclear

Name	Protein length (aa)	Molecular weight (kDa)	Charge	Isoelectric point	Grand average of hydropathy	Subcellular localization
GhPHD78	252	28.407	-7.5	4.889	-0.682	Nuclear
GhPHD79	186	21.734	1.5	6.851	-0.752	Cytoplasmic
GhPHD80	237	27.077	-4	5.221	-0.689	Extracellular
GhPHD81	1,356	154.398	31	7.737	-0.674	Nuclear
GhPHD82	236	26.789	-12	4.605	-0.598	Cytoplasmic
GhPHD83	676	74.819	22.5	8.253	-0.665	Nuclear
GhPHD84	1,382	156.796	34	7.925	-0.467	Nuclear
GhPHD85	949	104.967	2.5	6.779	-0.426	Nuclear
GhPHD86	1,653	183.656	37.5	8.15	-0.449	Nuclear
GhPHD87	1,618	180.589	39	8.234	-0.447	Nuclear
GhPHD88	216	24.836	5.5	7.902	-0.775	Cytoplasmic
GhPHD89	822	88.688	0	6.506	-0.531	Nuclear
GhPHD90	1,301	142.837	-35.5	4.873	-0.619	Nuclear
GhPHD91	705	78.888	23.5	8.373	-0.305	Nuclear
GhPHD92	2,182	241.654	-34	5.362	-0.441	Nuclear
GhPHD93	226	25.984	6	8.086	-0.767	Nuclear
GhPHD94	1,685	187.566	-4	6.345	-0.397	Nuclear
GhPHD95	1,237	137.855	-0.5	6.486	-0.718	Nuclear
GhPHD96	253	28.613	-6	5.13	-0.75	Nuclear
GhPHD97	696	78.217	18.5	8.023	-0.17	Nuclear
GhPHD98	503	55.441	-5.5	5.455	-0.597	Nuclear
GhPHD99	385	44.394	31	10.216	-0.428	Nuclear
GhPHD100	812	91.657	-29	4.818	-0.838	Nuclear
GhPHD101	859	95.86	-32.5	4.789	-1.023	Nuclear
GhPHD102	1,019	116.3	19	7.67	-0.593	Nuclear
GhPHD103	655	74.268	11.5	7.443	-0.219	Cytoplasmic

Name	Protein length (aa)	Molecular weight (kDa)	Charge	Isoelectric point	Grand average of hydropathy	Subcellular localization
GhPHD104	1,091	124.625	43	8.46	-0.581	Nuclear
GhPHD105	889	101.493	-32	4.811	-0.882	Nuclear
GhPHD106	1,305	145.225	9.5	7.121	-0.424	Nuclear
GhPHD107	801	90.038	-12.5	5.251	-0.805	Nuclear
GhPHD108	252	28.259	-6.5	5.136	-0.619	Cytoplasmic

Phylogenetic analysis, chromosomal location, and gene duplication

In order to better understand the phylogenetic relationships of the PHD proteins among rice, A. thaliana, and cotton, we constructed the NJ phylogenetic tree, which categorizes PHD proteins into five groups (A-E). Among them, most orthologous PHD proteins between diploid and allotetraploid cotton are classified under the same clade (Fig. 1), which have similar organization in phylogenetic relationship. The group A contains maximum numbers (97 members) of PHD proteins followed by group D comprising second largest group with 79 members. Groups B, C, and E contain relatively small numbers of PHD proteins, including 39, 42, and 40 members, respectively. Similarly, each group contains different numbers of PHD proteins members from all observed five species.

To determine the chromosomal location of GhPHD genes in G. hirsutum, the physical distribution of GhPHD genes along with their corresponding chromosomes were determined. A total of 108 GhPHD genes were mapped on 26 chromosomes, including 13 chromosomes from the At subgenome and 13 from the Dt subgenome (Fig. 2). Table S2 presents the chromosomal location and gene annotation information of GhPHD genes in G. hirsutum genome. Among these, each chromosome of At_10, At_11, Dt_03, and Dt_11 contains only two GhPHD genes, and each of At_05, At_07, and Dt_05 has eight members. Additionally, PHD genes from G. arboreum and G. raimondii are distributed on all 13 chromosomes. The number of PHD members in the At subgenome (56) and Dt subgenome (52) in G. hirsutum are different from that of PHD members in G. arboreum (52) and G. raimondii (55). Based on these results, we speculated that this difference may be due to gene loss during a long period of evolution or incomplete genome assembly.

We further investigated the putative GhPHD genes created by whole-genome duplication (WGD). Among all GhPHD genes, these were 73 segmental duplication and four tandem duplication events (Table 2), suggesting that WGD events were essential in the expansion of GhPHD genes. Duplicated gene pairs might have undergone three alternative fates during their evolution, i.e., non-functionalization, neo-functionalization, and sub-functionalization [38]. In order to investigate the evolutionary history of GhPHD genes, the Ka/Ks calculator 2.0 program was used to calculate synonymous and non-synonymous

substitution rates. In this study, the Ka/Ks ratio of 76 pairs of duplicated GhPHDs was lower than 1, indicating that they experienced purification selection pressure and limited functional divergence after GhPHD duplication (Table 2). However, only one pair of duplicated GhPHDs with a Ka/Ks ratio greater than 1 indicted that this pair of GhPHD genes has undergone positive selection (Table 2). These observations indicate that the function and purification selection of duplication GhPHD genes may greatly contribute to the functional maintenance in G. hirsutum.

Table 2
Ka/Ks analysis for the duplicated PHD gene pairs from G. hirsutum

Duplicated gene 1	Duplicated gene 2	Ka	Ks	Ka/Ks	Purifying selection	Duplicate type
GhPHD1	GhPHD11	0.064	0.533	0.119	Yes	Segmental
GhPHD1	GhPHD62	0.064	0.533	0.119	Yes	Segmental
GhPHD2	GhPHD58	0.011	0.043	0.248	Yes	Segmental
GhPHD5	GhPHD60	0.016	0.039	0.411	Yes	Segmental
GhPHD5	GhPHD95	0.126	0.389	0.324	Yes	Segmental
GhPHD6	GhPHD63	0.004	0.048	0.089	Yes	Segmental
GhPHD9	GhPHD61	0.007	0.032	0.212	Yes	Segmental
GhPHD10	GhPHD14	0.232	0.441	0.526	Yes	Segmental
GhPHD10	GhPHD47	0.245	0.488	0.502	Yes	Segmental
GhPHD10	GhPHD67	0.241	0.471	0.512	Yes	Segmental
GhPHD10	GhPHD101	0.235	0.446	0.527	Yes	Segmental
GhPHD11	GhPHD53	0.039	0.606	0.064	Yes	Segmental
GhPHD11	GhPHD62	0.004	0.014	0.282	Yes	Segmental
GhPHD13	GhPHD66	0.009	0.053	0.169	Yes	Segmental
GhPHD14	GhPHD47	0.376	0.666	0.564	Yes	Segmental
GhPHD14	GhPHD67	0.018	0.046	0.387	Yes	Segmental
GhPHD15	GhPHD68	0.030	0.063	0.468	Yes	Segmental
GhPHD16	GhPHD28	0.089	0.311	0.285	Yes	Segmental
GhPHD16	GhPHD77	0.045	0.352	0.129	Yes	Segmental
GhPHD16	GhPHD82	0.053	0.405	0.130	Yes	Segmental
GhPHD17	GhPHD71	0.008	0.034	0.235	Yes	Segmental
GhPHD17	GhPHD81	0.075	0.381	0.196	Yes	Segmental
GhPHD19	GhPHD25	0.081	0.447	0.180	Yes	Segmental
GhPHD19	GhPHD40	0.083	0.451	0.184	Yes	Segmental
GhPHD19	GhPHD73	0.026	0.053	0.492	Yes	Segmental

Duplicated gene 1	Duplicated gene 2	Ka	Ks	Ka/Ks	Purifying selection	Duplicate type
GhPHD19	GhPHD79	0.083	0.496	0.167	Yes	Segmental
GhPHD19	GhPHD93	0.085	0.451	0.188	Yes	Segmental
GhPHD20	GhPHD74	0.024	0.032	0.753	Yes	Segmental
GhPHD22	GhPHD51	0.080	0.398	0.200	Yes	Segmental
GhPHD22	GhPHD65	0.007	0.028	0.240	Yes	Segmental
GhPHD22	GhPHD104	0.079	0.391	0.201	Yes	Segmental
GhPHD28	GhPHD77	0.085	0.284	0.298	Yes	Segmental
GhPHD28	GhPHD82	0.042	0.037	1.121	No	Segmental
GhPHD24	GhPHD78	0.003	0.039	0.086	Yes	Segmental
GhPHD25	GhPHD73	0.057	0.434	0.131	Yes	Segmental
GhPHD25	GhPHD79	0.016	0.017	0.982	Yes	Segmental
GhPHD26	GhPHD44	0.204	0.368	0.553	Yes	Segmental
GhPHD26	GhPHD98	0.186	0.344	0.541	Yes	Segmental
GhPHD29	GhPHD83	0.019	0.041	0.473	Yes	Segmental
GhPHD31	GhPHD85	0.011	0.031	0.356	Yes	Segmental
GhPHD32	GhPHD86	0.015	0.027	0.554	Yes	Segmental
GhPHD34	GhPHD88	0.006	0.021	0.288	Yes	Segmental
GhPHD36	GhPHD89	0.015	0.031	0.499	Yes	Segmental
GhPHD39	GhPHD92	0.014	0.040	0.339	Yes	Segmental
GhPHD40	GhPHD73	0.051	0.442	0.116	Yes	Segmental
GhPHD40	GhPHD93	0.002	0.054	0.035	Yes	Segmental
GhPHD41	GhPHD94	0.013	0.031	0.416	Yes	Segmental
GhPHD44	GhPHD98	0.022	0.052	0.413	Yes	Segmental
GhPHD46	GhPHD54	0.127	0.418	0.304	Yes	Segmental
GhPHD46	GhPHD100	0.014	0.055	0.256	Yes	Segmental
GhPHD46	GhPHD107	0.101	0.395	0.256	Yes	Segmental
GhPHD47	GhPHD67	0.231	0.450	0.514	Yes	Segmental

Duplicated gene 1	Duplicated gene 2	Ka	Ks	Ka/Ks	Purifying selection	Duplicate type
GhPHD47	GhPHD101	0.016	0.030	0.530	Yes	Segmental
GhPHD49	GhPHD102	0.006	0.035	0.163	Yes	Segmental
GhPHD50	GhPHD103	0.011	0.048	0.233	Yes	Segmental
GhPHD51	GhPHD65	0.079	0.389	0.202	Yes	Segmental
GhPHD51	GhPHD104	0.010	0.036	0.270	Yes	Segmental
GhPHD52	GhPHD80	0.038	0.528	0.072	Yes	Segmental
GhPHD53	GhPHD62	0.039	0.606	0.064	Yes	Segmental
GhPHD54	GhPHD100	0.140	0.416	0.336	Yes	Segmental
GhPHD54	GhPHD107	0.136	0.405	0.335	Yes	Segmental
GhPHD55	GhPHD106	0.023	0.045	0.497	Yes	Segmental
GhPHD56	GhPHD76	0.236	0.639	0.369	Yes	Segmental
GhPHD56	GhPHD108	0.007	0.051	0.132	Yes	Segmental
GhPHD60	GhPHD95	0.125	0.397	0.314	Yes	Segmental
GhPHD65	GhPHD104	0.078	0.386	0.203	Yes	Segmental
GhPHD67	GhPHD101	0.225	0.456	0.492	Yes	Segmental
GhPHD71	GhPHD81	0.075	0.369	0.203	Yes	Segmental
GhPHD73	GhPHD79	0.054	0.486	0.111	Yes	Segmental
GhPHD73	GhPHD93	0.053	0.419	0.127	Yes	Segmental
GhPHD76	GhPHD108	0.237	0.630	0.376	Yes	Segmental
GhPHD91	GhPHD38	0.016	0.052	0.298	Yes	Segmental
GhPHD100	GhPHD107	0.104	0.379	0.276	Yes	Segmental
GhPHD2	GhPHD3	0.035	0.122	0.289	Yes	Tandem
GhPHD32	GhPHD33	0.031	0.076	0.407	Yes	Tandem
GhPHD58	GhPHD59	0.029	0.138	0.212	Yes	Tandem
GhPHD86	GhPHD87	0.027	0.062	0.428	Yes	Tandem

Gene structure and conserved motifs analysis

In order to better understand the similarity and diversity of different members of the PHD protein in G. hirsutum, we analyzed the phylogenetic tree, exon-intron structure, and conserved motif (Fig. 3). GhPHD genes in the branches with closer homology have similar gene structure and exon-intron number, and their corresponding PHD proteins have highly conserved motif, which is consistent with the conclusion in the phylogenetic tree. Among them, GhPHD49 has the longest genomic sequence and 26 exons, while GhPHD12 has the shortest genomic sequence and only has two exons (Fig. 3B and Table S3). A total of three conserved motifs were identified, named motif 1 to motif 3, respectively (Fig. 3C). Among them, 20 GhPHD proteins all have motif 1, 2, and 3, and 75 GhPHD proteins only have motif 1. Different GhPHD proteins have different number of motifs, but GhPHD proteins with the same clustering relationship have strong motif conservation.

The alignment of 108 GhPHD conserved domain was shown in Fig. S1. The GhPHD motifs have nine conserved sites (indicated with dark blue and pink) and they all have typical Cys4-His-Cys3 model. Conserved histidine (H) was separated by four amino acids from the fourth conserved cysteine (C), and was separated by two amino acids from the latter conserved cysteine C. The third and fourth conserved cysteines (C) before histidine (H) were separated by one or two amino acids, but the interval number between other conserved amino acids was uncertain. However, GhPHD17, GhPHD27, GhPHD71, and GhPHD81 were closely related, but they did not follow the rules about conserved domains mentioned above (Fig. S1 and Fig. 3).

Cis -acting element analysis

Many studies have showed that PHD genes were involved in various stress responses. To elucidate the putative function of GhPHD under different stresses, we identified putative stress-related and phytohormone-related cis-acting elements in the 1500 bp promoter sequence before the start transcription codon (ATG) of 108 GhPHD genes (Fig. 4 and Table S4). Ten kinds of plant hormone-related elements, ABRE (abscisic acid), TGA-element (auxin), AuxRR-core (auxin), ERE (ethylene), TATC-box (gibberellin), GARE-motif (gibberellin), P-box (gibberellin), CARE (gibberellin), CGTCA-motif (MeJA), TCA-element (salicylic acid); and three kinds of stress-responsive regulatory elements, TC-rich repeats (defense and stress responsiveness), MBS (drought), LTR (low-temperature) were identified in the GhPHD promoters. In addition, these GhPHD promoters contain circadian control elements and light-responsive elements (G-box). These results revealed that all GhPHD promoters contain at least one putative stress-responsive element.

In addition to GhPHD52 and GhPHD10, other GhPHD promoters contained elements that respond to single or multiple phytohormones (Fig. 4 and Table S5). There were 85, 67, 64, and 55 GhPHD promoters with ethylene, ABA, GA, and MeJA response elements, respectively. In particular, the 17 GhPHD promoters contain the response elements of four hormones including ethylene, ABA, GA, and MeJA. Further, there were 26 and 37 GhPHD promoters containing auxin and SA response elements, respectively. The promoter region of 27 GhPHD genes contained cis-acting elements in response to four hormones, and the promoter regions of eight GhPHDs contained cis-acting elements in response to five hormones. In

particular, GhPHD5, GhPHD47, GhPHD56, and GhPHD65 contained cis-acting elements in response to six hormones, indicating that these genes might be involved in multiple hormone signal pathways. Interestingly, cis-acting element analysis showed that GhPHD may be involved in light response and abiotic stress pathways, because there were 71, 38, and 36 GhPHD promoters with light, drought, and low temperature response elements, respectively. Additionally, some GhPHD genes promoters contain both low temperature and drought response elements. Coupled with these results, we speculated that GhPHDs might play important roles in light response, abiotic stress and hormone signaling pathways.

Tissue-specific expression patterns of GhPHD genes in cotton tissues

To predict the possible functions of GhPHD genes during cotton development, we investigated their expression profiles in 20 different tissues including root, stem, leaf, petal, stamen, pistil, ovule (-3, -1, 0, 1, 3, 5, 10, 20, 25 and 35 days ovule), and fiber (5, 10, 20, 25 DPA) using available transcriptomic data. According to the expression features (Fig. 5), 108 GhPHD genes were mainly clustered into four groups (A-D) based on a hierarchical clustering analysis. Among these, nine GhPHD genes in group A were highly expressed in each tissue and preferentially expressed in ovules, especially GhPHD23 and GhPHD77. The expression of 43 GhPHDs in group B were slightly lower than that of GhPHDs in group A. Groups C and D each contained 28 members and their expression were very low in all the tissues observed. In group D, three GhPHDs (GhPHD6, GhPHD10, and GhPHD63) were not expressed in root, stem, leaf, petal, stamen, and pistil. GhPHD6 was expressed only in 3 DPA ovule, while GhPHD103 was expressed only in root and fiber. GhPHD7 and GhPHD87 were not expressed in fibers, but they are low in other tissues. Different expression patterns of each GhPHD genes revealed that they have different physiological functions in cotton.

Identification of stress-related PHD genes in G. hirsutum

Cis-acting elements analysis presumed that GhPHDs may be involved in the stress responses. To verify this hypothesis, we analyzed the expression profile of GhPHD genes under heat, cold, salt, and drought treatments using the transcriptomic data. Under different stress treatments, each gene responds differently to stress. However, 66 GhPHD genes were highly expressed under stress treatment (Fig. S2). To identify GhPHD genes that play important roles in the stress responses, we verified the expression level of 12 GhPHD genes using qRT-PCR. Under the treatment of heat, cold, salt, and drought (Fig. 6), the relative expression level of GhPHD18 were up-regulated, indicating that GhPHD18 responded positively to all four stress treatments. GhPHD23 was also up-regulated after heat treatment, but was down-regulated after cold, salt, and drought treatment, as the expression level of GhPHD23 decreased dramatically after six hours of drought treatment (Fig. 6). Similarly, under heat and cold treatment, the expression patterns of the six GhPHD genes (GhPHD34, 40, 43, 80, 88, and 107) were consistent with that of GhPHD23, that is, their gene expression levels increased by 2.0-2.5 times under heat treatment, but decreased under cold treatment (Fig. 6).

Under salt treatment, GhPHD5, 24, 34, 40, 43, and 72 were up-regulated, while GhPHD80 and GhPHD88 were down-regulated. Under drought treatment, the relative expression levels of GhPHD5, 18, 80, and 88

increased, while GhPHD23, 40, 43, and 107 decreased (Fig. 6). The relative expression of GhPHD77 increased gradually at different times of heat treatment, but its expression level did not increase significantly under cold, salt, and drought treatments (Fig. 6). These results indicate that different GhPHDs respond differently to different abiotic stresses. These seven GhPHD (23, 34, 40, 43, 77, 80, and 107) showed positive responses to heat stress, while GhPHD23, GhPHD77, and GhPHD80 showed negative responses to cold, salt, and drought.

Identification of the GhPHD genes in response to phytohormone

It is an established fact that phytohormone act as chemical messengers in the physiological processes during the plant's life cycle, which simultaneously coordinate physiological responses to biotic and abiotic stresses. To further explore the potential functions of GhPHDs in response to stress-related plant hormones, we analyzed the expression pattern of six GhPHDs under GA, MeJA, IAA, SA, and BL treatment by qRT-PCR (Fig. 7). Among 12 GhPHDs mentioned above in response to abiotic stresses (Fig. 6), six GhPHD containing four or more hormone response elements (Fig. 4) in the promoter region were selected and their responses to different hormones were identified.

The GhPHD5 promoter contains responsive elements for six hormones (auxin, ethylene, GA, MeJA, SA, and ABA) (Fig. 4). However, qRT-PCR results showed that the relative expression level of GhPHD5 increased after treatment with MeJA, IAA, and BL. The expression level of GhPHD5 increased at 0.5 hour of SA treatment, while decreased significantly at 1, 3, and 5 hours of SA treatment. However, at different time points of GA treatment, the expression level of GhPHD5 fluctuated, but the difference was not significant. Similarly, after 0.5 hour of GA treatment, the expression level of GhPHD40 increased to 1.2, and then gradually decreased with time. After SA treatment, the expression level of GhPHD40 increased at different time points. However, its expression decreased at 0.5 hour of BL treatment, and then increased to 1.45 at one hour after BL treatment (Fig. 7).

Under the treatment of five hormones at different time points, the relative expression level of GhPHD43 gradually increased, which indicates that GhPHD43 has a strong response to BL and the weakest response to MeJA (Fig. 7). In contrast to GhPHD43, the expression level of GhPHD80 decreased after treatment with five hormones (Fig. 7). The relative expression level of GhPHD80 increased significantly to 2.9 at 0.5 hour after GA treatment, and then rapidly decreased to 0.2 at three hours of GA treatment. Interestingly, the response of GhPHD88 to GA treatment was the same as that of GhPHD80, that is, the relative expression level of GhPHD88 increased to 2.3 at 0.5 hour of GA treatment, and reached a minimum of 0.3 after three hours of GA treatment (Fig. 7). Under SA treatment, the relative expression level of GhPHD88 increased gradually, while under BL treatment, the relative expression level decreased to 0.3 at 0.5 hour, but increased to 1.7 at one hour, and decreased at three and six hours. Under GA and IAA treatments, the relative expression of GhPHD107 increased and reached the maximum at one hour. The relative expression of GhPHD107 decreased significantly at 0.5 hour of BL treatment, reached the maximum at one hour, and then decreased significantly at three hours and six hours after treatment

(Fig. 7). The above results indicated that GhPHDs showed diverse expression pattern under different hormone treatment and might play crucial roles in hormone signaling pathways.

Co-expression networks with functional modules for G. hirsutum and G. arboreum

GO ontology analysis of all 108 GhPHDs genes showed that protein binding and zinc ion binding were the most abundant functional terms compared to the regulation of transcription, histone binding, and chromatin binding (Fig. 8B). We then used ccNET software to analyze the co-expression networks of the genes identified above for stress and phytohormone response. After analyzing them one by one, we predicted many co-expressed genes and interacting proteins (Table S6). For example, GhPHD5 depicted a positive co-expression relationship with SLOMO (slow motion), which is a F-box protein required for auxin homeostasis and normal timing of lateral organ initiation at the shoot meristem [39]. GhPHD5 showed positive co-expression with a novel plant-specific DNA ligase, which is involved in seed germination and DNA repair (Table S6).

GhPHD18 may interact with highly hydrophilic proteins that are involved in positively regulating FLC expression. GhPHD18 have positive co-expression with a SHAGGY-related kinase involved in meristem organization, indicating that GhPHD18 play important roles in plant growth and development (Table S6). GhPHD34 was negatively co-expression with ERF subfamily B-1 of the ERF/AP2 transcription factor and participated in the ethylene signal pathway and abiotic stress response (Table S6). Similarly, GhPHD107 exhibited positive co-expression with ARF-GAP and ERF (ethylene response factor), suggesting that GhPHD107 may be involved in auxin and ethylene signaling pathways (Table S6). More interestingly, the physiological function of GhPHD88 is particularly powerful, and the genes predicted to interact with GhPHD88 are leucine-rich repeat protein kinase, late embryogenesis abundant protein (LEA), AP2/B3 transcription factor, R2R3 factor gene, DREB subfamily A-2 of ERF/AP2 transcription factor, cellulose synthase, gibberellin-regulated family protein, ethylene response factor, and other genes (Fig. 8 and Table S6)), which indicates that GhPHD88 may be involved in many physiological processes such as plant growth and development, hormone signal transduction and stress response. These results showed that GhPHDs were involved in the plant growth and development, and phytohormone may mediate the response of the GhPHD to abiotic stress, thereby increasing the plant's tolerance to adverse environmental stress.

Discussion

Identification, phylogenetic analysis, and evolution of PHD in cotton

We performed a genetic phylogenetic analysis of PHD proteins in cotton, rice, and Arabidopsis and divided into five subfamilies (Fig. 1). The genetic distance indicates the closeness of the homologous relationships. The members with closer homology relationship are clustered under the same branch, which also reveals that they may have similar genetic functions. Genetic evolution analysis showed that

the relationship between GhPHDs and AtPHDs were closer than that with OsPHDs, which were consistent with the evolutionary relationship between cotton, Arabidopsis, and rice.

In this study, we identified 55, 52, and 108 PHD proteins in G. raimondii, G. arboreum, and G. hirsutum, respectively (Table S1). Phylogenetic tree analysis revealed that the two orthologous GhPHD genes of G. hirsutum corresponded to that of G. arboreum and G. raimondii, respectively, and they were highly conserved in gene structure and motif and diverged from a common ancestor. This may be due to the allotetraploid cotton species G. hirsutum originated from an interspecific hybridization event between the progenitors G. raimondii (D5) and G. arboreum (A2). [40, 41]. The number of GhPHD genes is approximately equal to the sum of the number of GrPHD genes and GaPHD genes, indicating that whole-genome duplication was the major impetus, followed by varying degrees of gene loss in the evolution of allotetraploid cotton. Consistent with this, gene loss was accompanied by rapid arrangement of genomic sequences after hybridization and chromosome doubling during polyploidization [42]. Although the G. arboreum genome is about twice that of G. raimondii, the number of GrPHD genes is three more than the number of GaPHD genes, which may be explained by the insertion of long-term repeat retrotransposon into G. arboreum.

Duplication of PHD genes in G. hirsutum

Previous studies have reported that gene replication was the main cause of gene family expansion, including whole-genome duplication, segment duplication, tandem duplication, and transposition events [43, 44]. To understand the expansion mechanism of GhPHD gene, we analyzed their duplication events in G. hirsutum. A total of 77 duplicated gene pairs were identified in GhPHD gene family, including 73 segmental duplicated pairs and four tandem duplicated pairs (Table 2). This indicates that purifying selection dominated the expansion of GhPHD genes, which can eliminate harmful loss-of-function mutations at both duplicated loci, increase fixation and retain the function of a new duplicated gene [45].

After gene duplication, the coding region can gain new regulatory background by inserting or deleting tissue-specific enhancers or repressors, resulting in spatiotemporal change in the expression pattern of duplicated genes. Gene expression patterns in different tissues reflect difference in biological functions during plant growth and development. Most GhPHD duplicated gene pairs show similar expression patterns in different tissues, indicating that their biological functions are not much different, but some GhPHD duplicated gene pairs show different expression patterns, and their biological functions may have experienced divergence in order to quickly adapt to the new environment. However, the difference in tissue expression patterns and biological functions between GhPHD genes need further verification.

Identification of GhPHD respond to abiotic stresses and various hormones

Many studies demonstrated that PHD proteins are the main mediator of transcriptional regulation of plant developmental processes such as meiosis and postmeiotic events [46], germination [33], pollen maturation [47], embryo meristem initiation and root development [46, 48], and flowering time [49]. PHD

proteins affect the processes of growth and development by changing transcription and reading epigenetic histone modifications, but their functions in abiotic stress responses remain largely unclear. Transgenic Arabidopsis plants overexpressing the GmPHD2 showed salt tolerance when compared with the wild type plants [31, 50].

Analysis of cis-acting elements and transcriptomic data indicated that GhPHD may respond to both phytohormone and abiotic stress (Fig. 4 and S2). qRT-PCR identified GhPHD genes response to abiotic stress and different phytohormone (Figs. 6 and 7). GhPHD5 may regulate heat and drought through SA signal pathway because GhPHD5 depicted negative response to heat, drought, and SA. Further, GhPHD40, GhPHD43, and GhPHD80 were positively regulated by heat, but negatively regulated by cold, salt, and drought, and they may be involved in IAA, GA, and BL signaling pathways. GhPHD107 may be involved in GA and IAA pathways, which in turn has a significant positive response to heat. More interestingly, GhPHD88 may negatively respond to abiotic stresses (such as heat, cold, salt, and drought) through SA and BL pathways. The results of tissue specific expression pattern and co-expression network analysis indicated that GhPHDs play important roles in plant growth, development and phytohormone signal transduction pathways. However, the physiological function of GhPHDs in crosstalk between abiotic stress and phytohormone needs further study.

Conclusions

In this study, a total of 297 PHD proteins were identified based on the genome information of total five species including G. hirsutum, G. arboreum, G. raimondii, rice, and Arabidopsis. The PHD proteins were divided into five groups based on the phylogenetic analysis. Segmental duplication events were the main contributors to the expansion of PHD gene family in G. hirsutum. Moreover, duplicated gene pairs of PHD gene family in upland cotton might have experienced functional divergence, since their expression patterns were different in different tissues. Tissues specific expression patterns and co-expression network analysis showed that GhPHD were essential for plant growth and development and were involved in response to phytohormone and abiotic stress. We have identified GhPHD genes that may be involved in both abiotic stress and phytohormone signaling pathway in cotton. Taken together, our study provides key basic knowledge to understand the functional mechanisms of cotton growth and development, as well as candidate genes for cotton breeding that are resistant to abiotic stress and phytohormone stimulation.

Methods

Sequence retrieval, multiple sequence alignment, and phylogenetic analysis

The cotton (G. hirsutum, G. raimondii, and G. arboreum) genome sequences and information were acquired from the CottonFGD (https://cottonfgd.org/) [51]. HMMER (https://www.ebi.ac.uk/Tools/hmmer/) software with default parameters was used to search for

corresponding protein sequences, and the conserved PHD domain was used as a query. We used the BLAST program to further identify PHD sequences based on homology. The conserved domains of PHD proteins were predicted by Pfam [52] and SMART [53] software. Multiple sequence alignments of all identified PHD proteins were performed using Clustal X [54]. Phylogenetic tree of deduced amino acid sequences was constructed using the neighbor-joining (NJ) algorithm with default parameters and 1000 bootstrap replicates were used in MEGA 6.0 [55]. The molecular weight (Mw), isoelectric point (pI), and GRAVY values of the identified PHD proteins were predicted using ExPASy [56] and the subcellular localization of the GhPHD proteins were predicted through the CELLO v2.5 server [57].

Chromosomal location, gene structure, and conserved motif

The positional information for GhPHD genes were obtained from the general feature format (GFF) files downloaded from the CottonFGD website [51] and GhPHDs were mapped on the chromosomes using MapInspect (https://mapinspect.

software.informer.com/). For the exon-intron structural analysis of GhPHD genes, the coding sequences were used to align their genomic DNA sequences and the structure diagrams were drawn using the online Gene Structure Display Server (GSDS 2.0) program [58]. Conserved motifs of GhPHD proteins were investigated using the online toolkit Multiple Expectation maximization for Motif Elicitation (MEME 5.0.5) [59]. The optimized parameters of MEME were employed as follows: number of repetitions, any; maximum number of motifs, 50; and the optimum width of each motif, between 6 and 300 residues, and retaining only motifs associated with an E value < e^{-5}. The identified protein motifs were further annotated with TBtools [60].

Identification of cis-acting elements in the GhPHD promoter region and gene expression pattern

The promoter sequences of 1500 bp length before the transcriptional initiation site of GhPHD genes were downloaded from the CottonFGD website [51]. The cis-acting elements in GhPHD promoter regions were predicted using the website Plant Cis-Acting Regulatory Element [61]. The expression level of GhPHD genes were extracted from the transcriptome analysis data and heatmap were drawn by TBtools [60]. The ccNET software [62] was used to analyze the co-expression network relation of PHD genes.

Plant material, abiotic stress, and hormone treatments

Gossypium hirsutum L. acc. ZM24 is a short-season cotton variety selected by the Cotton Research Institute, Chinese Academy of Agricultural Sciences, and was approved by the National Crop Variety Examination and Approval Committee in 1997. The national examination and approval number is GS08001-1997. ZM24 was cultivated in the greenhouse of Zhengzhou Research Center, Cotton Research Institute of Chinese Academy of Agricultural Sciences. Firstly, ZM24 seeds were pre-germinated in a conical flask containing water for 48 hours at room temperature, and then transferred to a liquid medium with a 16 h light and 8 h dark cycle at 30 ℃. Four-week-old seedlings at the 3–4 leaf stage were treated with brassinolide (BL; 10 μM), gibberellins (GA; 100 μM), indole-3-acetic acid (IAA; 100 μM), salicylic acid

(SA; 10 μM), and methyl jasmonate (MeJA; 10 μM) for 0.5, 1, 3, and 6 hours. Similarity, the seedlings were treated with heat (38 ℃), cold (4 ℃), NaCl (300 nM), and polyethylene glycol (PEG) (20% mass fraction) for 1, 2, 4, and 6 hours. The collected tissues were frozen immediately in liquid nitrogen and stored at -80 ℃ for RNA extraction and qRT-PCR analysis. For stress and phytohormone treatment experiments, a total of 20 seedlings were used for each treatment with three biological replicates.

RNA extraction and qRT-PCR analysis

Total RNA of cotton seedlings were extracted using the RNAprep Pure Plant Kit (Polysaccharides & Polyphenolics-rich) (TianGen, Beijing, China). For synthesizing first-strand cDNA, the EasyScript All-in-One First-strand cDNA synthesis SuperMix for qRT-PCR (TransGen, Beijing, China) were used in accordance with the manufacturer's protocol, and the cDNAs were used as templates in subsequent qRT-PCR reactions. qRT-PCR was performed using TransStart Top Green qPCR SuperMix (TransGen, Beijing, China) in a LightCycler 480 (Roche, Basel, Switzerland) with a PCR program starting with 94 ℃ for 30 s, then 45 cycles of 94 ℃ for 5 s and 60 ℃ for 30 s. Each PCR reaction was performed in triplicate, and three biological replicates were quantified. GhHis3 (GenBank accession no. AF024716) was used as an internal control. Relative expression was calculated as described previously [63]. Primers used in this study for qPCR analysis were listed in Supplementary Table S7. For statistical analysis, the qRT-PCR data were considered to have a normal distribution and we conducted two-tailed Student's t-test in Microsoft Excel 2007.

Declarations

Ethics approval and consent to participate

Not applicable. Our research did not involved any human or animal subjects, material, or data. The plant materials used in this study were provided by the Institute of Cotton Research of Chinese Academy of Agricultural Sciences and are freely available for research purposes following institutional, national and international guidelines.

Consent to publication

Not applicable.

Availability of data and materials

The data used or analyzed during the current study has been included in this article and its Additional materials. And the genome sequence and annotation datasets that supported the findings of this study are available in:

1. *thaliana*: https://www.arabidopsis.org./
2. *sativa*: http://plants.ensembl.org/index.html
3. *hirsutum*, *G. arboreum*, and *G. raimondii*: https://cottonfgd.org/

Competing interests

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

Funding

This study was supported by the Major Research Plan of National Natural Science Foundation of China (grant number 31690093). The funder was not involved in the experimental design of the study, data collection and interpretation, and in writing the manuscript.

Authors' contributions

H.W. and G.Q. conceived and designed the experiments. H.W. and M.G. performed the experiment. H.W. and L.Z. analyzed the data. H.W. wrote the paper. Z.Y. and Z.W. revised the paper. All of the authors read and approved the final the manuscript.

Acknowledgements

None.

Abbreviations

G. hirsutum:Gossypium hirsutum; G. arboreum:Gossypium arboreum; G. raimondii:Gossypium raimondii; O. sativa:Oryza sativa; A. thaliana:Arabidopsis thaliana; DPA:Day post-anthesis; PHD:Plant homeodomain; NJ:Neighbor-joining; BL:Brassinolide; GA:Gibberellins; IAA:Indole-3-acetic acid; SA:Salicylic acid; MeJA:Methyl jasmonate; ABA:Abscisic acid; ET:Ethylene; CK:Cytokinin; SL:Strigolactone. PEG:Polyethylene glycol; qRT-PCR:quantitative real-time polymerase chain reaction; GO:Gene Ontology.

References

1. Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JI. Plant salt-tolerance mechanisms. Trends Plant Sci. 2014;19(6):371-379.
2. Hossain MA, Li Z-G, Hoque TS, Burritt DJ, Fujita M, Munné-Bosch S. Heat or cold priming-induced cross-tolerance to abiotic stresses in plants: key regulators and

possible mechanisms. Protoplasma. 2018;255(1):399-412.
3. Saeed M, Dahab A, Wangzhen G, Tianzhen Z. A cascade of recently discovered molecular mechanisms involved in abiotic stress tolerance of plants. Omics : a journal of integrative biology. 2012;16(4):188-199.
4. Bari R, Jones JD. Role of plant hormones in plant defence responses. Plant Mol Biol. 2009;69(4):473-488.
5. Nakashima K, Yamaguchi-Shinozaki K. ABA signaling in stress-response and seed development. Plant Cell Rep. 2013;32(7):959-970.
6. Loake G, Grant M. Salicylic acid in plant defence--the players and protagonists. Curr Opin Plant Biol. 2007;10(5):466-472.
7. Wasternack C, Hause B. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. Ann Bot. 2013;111(6):1021-1058.
8. Takatsuji H, Mori M, Benfey PN, Ren L, Chua NH. Characterization of a zinc finger DNA-binding protein expressed specifically in Petunia petals and seedlings. Embo j. 1992;11(1):241-249.
9. Sakai H, Medrano LJ, Meyerowitz EM. Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries. Nature. 1995;378(6553):199-203.
10. Omichinski JG, Clore GM, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl SJ, Gronenborn AM. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science. 1993;261(5120):438-446.
11. Daniel-Vedele F, Caboche M. A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi. Molecular & general genetics : MGG. 1993;240(3):365-373.
12. Yanagisawa S. Dof DNA-binding proteins contain a novel zinc finger motif. Trends in Plant Science. 1996;1(7):213-214.
13. Yanagisawa S, Izui K. Molecular cloning of two DNA-binding proteins of maize that are structurally different but interact with the same sequence motif. Journal of Biological Chemistry. 1993;268(21):16028-16036.
14. von Arnim AG, Deng XW. Ring finger motif of Arabidopsis thaliana COP1 defines a new class of zinc-binding domain. Journal of Biological Chemistry. 1993;268(26):19626-19631.
15. Schindler U, Beckmann H, Cashmore AR. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. Plant J. 1993;4(1):137-150.
16. Bellmann R, Werr W. Zmhox1a, the product of a novel maize homeobox gene, interacts with the Shrunken 26 bp feedback control element. Embo j. 1992;11(9):3367-3374.

17. Sanchez-Garcia I, Rabbitts TH. The LIM domain: a new structural motif found in zinc-finger-like proteins. Trends in genetics : TIG. 1994;10(9):315-320.
18. Baltz R, Domon C, Pillay DT, Steinmetz A. Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. The Plant journal : for cell and molecular biology. 1992;2(5):713-721.
19. Kaadige MR, Ayer DE. The polybasic region that follows the plant homeodomain zinc finger 1 of Pf1 is necessary and sufficient for specific phosphoinositide binding. J Biol Chem. 2006;281(39):28831-28836.
20. Bienz M. The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci. 2006;31(1):35-40.
21. Martin DG, Baetz K, Shi X, Walter KL, MacDonald VE, Wlodarski MJ, Gozani O, Hieter P, Howe L. The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. Molecular and cellular biology. 2006;26(21):7871-7879.
22. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P *et al*. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature. 2006;442(7098):86-90.
23. Reddy TV, Kaur J, Agashe B, Sundaresan V, Siddiqi I. The DUET gene is necessary for chromosome organization and progression during male meiosis in Arabidopsis and encodes a PHD finger protein. Development. 2003;130(24):5975-5987.
24. Yang X, Makaroff CA, Ma H. The Arabidopsis MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis. Plant Cell. 2003;15(6):1281-1295.
25. Woo HR, Pontes O, Pikaard CS, Richards EJ. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 2007;21(3):267-277.
26. Liang C, Zhang X, Song S, Tian C, Yin Y, Xing G, He F, Zhang L. Identification of UHRF1/2 as new N-methylpurine DNA glycosylase-interacting proteins. Biochemical and biophysical research communications. 2013;433(4):415-419.
27. Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, Koseki H, Wong J. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun. 2013;4:1563.
28. Maenohara S, Unoki M, Toh H, Ohishi H, Sharif J, Koseki H, Sasaki H. Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. PLoS Genet. 2017;13(10):e1007042.
29. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, Baird CL, Zhu H, Field SJ, Lessnick SL *et al*. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. Cell. 2003;114(1):99-111.

30. Coscoy L, Ganem D. PHD domains and E3 ubiquitin ligases: viruses make the connection. Trends Cell Biol. 2003;13(1):7-12.
31. Wei W, Huang J, Hao YJ, Zou HF, Wang HW, Zhao JY, Liu XY, Zhang WK, Ma B, Zhang JS *et al*. Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic Arabidopsis plants. PLoS One. 2009;4(9):e7209.
32. Kim DH, Sung S. Accelerated vernalization response by an altered PHD-finger protein in Arabidopsis. Plant Signal Behav. 2017;12(5):e1308619.
33. Molitor AM, Bu Z, Yu Y, Shen WH. Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. PLoS Genet. 2014;10(1):e1004091.
34. Ye Y, Gong Z, Lu X, Miao D, Shi J, Lu J, Zhao Y. Germostatin resistance locus 1 encodes a PHD finger protein involved in auxin-mediated seed dormancy and germination. Plant J. 2016;85(1):3-15.
35. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM *et al*. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature biotechnology. 2015;33(5):531-537.
36. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C *et al*. Genome sequence of the cultivated cotton Gossypium arboreum. Nature genetics. 2014;46(6):567-572.
37. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J *et al*. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423-427.
38. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151-1155.
39. Lohmann D, Stacey N, Breuninger H, Jikumaru Y, Muller D, Sicard A, Leyser O, Yamaguchi S, Lenhard M. SLOW MOTION is required for within-plant auxin homeostasis and normal timing of lateral organ initiation at the shoot meristem in Arabidopsis. Plant Cell. 2010;22(2):335-348.
40. Cronn RC, Small RL, Wendel JF. Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci U S A. 1999;96(25):14406-14411.
41. Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM: Evolution and Natural History of the Cotton Genus. In: *Genetics and Genomics of Cotton.* Edited by Paterson AH. New York, NY: Springer US; 2009: 3-22.
42. Paterson AH, Bowers JE, Chapman BA. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A. 2004;101(26):9903-9908.
43. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell. 2004;16(7):1667-1678.

44. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183(3):557-564.
45. Tanaka KM, Takahasi KR, Takano-Shimizu T. Enhanced fixation and preservation of a newly arisen duplicate gene by masking deleterious loss-of-function mutations. Genetics research. 2009;91(4):267-280.
46. Sebastian J, Ravi M, Andreuzza S, Panoli AP, Marimuthu MP, Siddiqi I. The plant adherin AtSCC2 is required for embryogenesis and sister-chromatid cohesion during meiosis in Arabidopsis. Plant J. 2009;59(1):1-13.
47. Li H, Yuan Z, Vizcay-Barrena G, Yang C, Liang W, Zong J, Wilson ZA, Zhang D. PERSISTENT TAPETAL CELL1 encodes a PHD-finger protein that is required for tapetal cell death and pollen development in rice. Plant Physiol. 2011;156(2):615-630.
48. Schlereth A, Moller B, Liu W, Kientz M, Flipse J, Rademacher EH, Schmid M, Jurgens G, Weijers D. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature. 2010;464(7290):913-916.
49. Lopez-Gonzalez L, Mouriz A, Narro-Diego L, Bustos R, Martinez-Zapater JM, Jarillo JA, Pineiro M. Chromatin-dependent repression of the Arabidopsis floral integrator genes involves plant specific PHD-containing proteins. Plant Cell. 2014;26(10):3922-3938.
50. Wei W, Zhang YQ, Tao JJ, Chen HW, Li QT, Zhang WK, Ma B, Lin Q, Zhang JS, Chen SY. The Alfin-like homeodomain finger protein AL5 suppresses multiple negative factors to confer abiotic stress tolerance in Arabidopsis. Plant J. 2015;81(6):871-883.
51. Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R. CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol. 2017;17(1):101.
52. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 1998;26(1):320-322.
53. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493-d496.
54. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al*. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-2948.
55. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-2729.
56. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E *et al*. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40(Web Server issue):W597-W603.
57. Yu C-S, Lin C-J, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004;13(5):1402-1406.

58. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296-1297.
59. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(Web Server issue):W202-W208.
60. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. bioRxiv. 2018:289660.
61. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325-327.
62. You Q, Xu W, Zhang K, Zhang L, Yi X, Yao D, Wang C, Zhang X, Zhao X, Provart NJ *et al*. ccNET: Database of co-expression networks with functional modules for diploid and polyploid Gossypium. Nucleic Acids Res. 2017;45(D1):D1090-d1099.
63. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-408.

Supplementary Files Legend

Additional file 1: Fig. S1.tif Alignment results from the conserved domain of 108 GhPHD proteins and PHD motifs have nine conserved amino acids with a typical C_4HC_3 model.

Additional file 2: Fig. S2.tif Expression profiles of the *GhPHD* genes under cold, hot, salt, and drought. The expression characteristics of 108 *GhPHD* genes under these four stress treatments were investigated using available transcriptomic data. 1 h, 3 h, 6 h, and 12 h indicate hours after different stress treatments. Gene names and the subfamilies are shown on the right. Blocks with colors represent the relative expression levels of GhPHDs.

Additional file 3: Table S1.xlsx The PHD members from *G. hirsutum*, *G. raimondii*, G. *arboreum*, *A. thaliana* , and *O. sativa*.

Additional file 4: Table S2.docx Chromosomal location and gene annotation of *GhPHD* genes in *G. hirsutum* genome.

Additional file 5: Table S3.docx Transcript-features of the 108 *GhPHD* genes.

Additional file 6: Table S4.xlsx Distribution of major stress-related and plant hormone-related *cis*-acing elements in the promoter regions of *GhPHD* genes.

Additional file 7: Table S5.xlsx Number of *cis*-acting elements in the promoters of *GhPHD* genes.

Additional file 8: Table S6.xlsx Co-expression network analysis results.

Additional file 9: Table S7.docx Primers for qRT-PCR in this study.

Figures

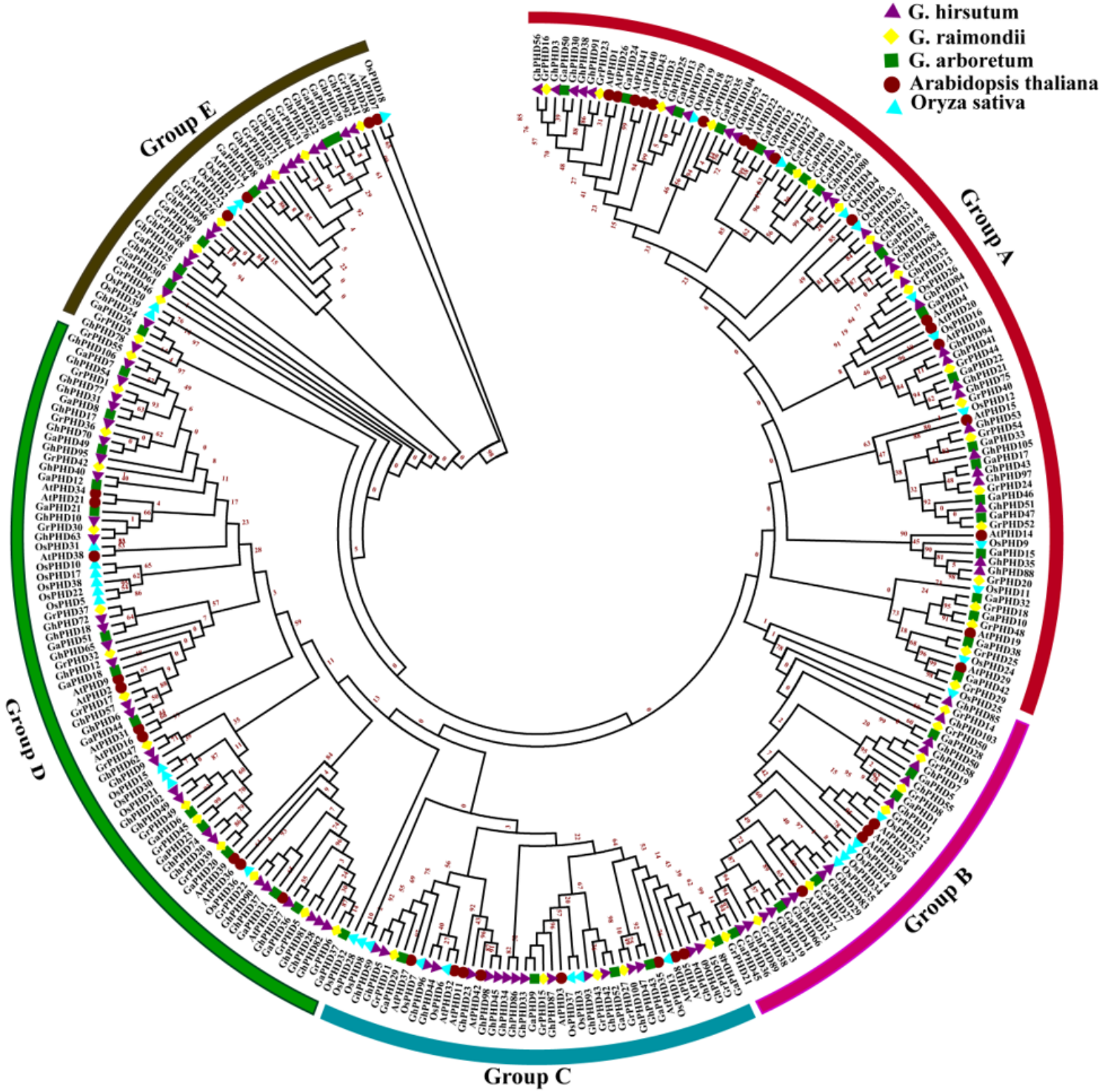

Figure 1

Phylogenetic tree displaying relationships between 108 G. hirsutum, 52 G. arboreum, 55 G. raimondii, 39 O. sativa and 43 A. thaliana PHD proteins. The phylogenetic tree was constructed in MEGA 6.0 using the neighbor-joining method. The bootstrap test was performed with 1000 iterations. The five subgroups are shown with different colors. At, Arabidopsis thaliana; Ga, Gossypium arboreum; Gr, Gossypium raimondii; Gh, Gossypium hirsutum; Os, Oryza sativa.

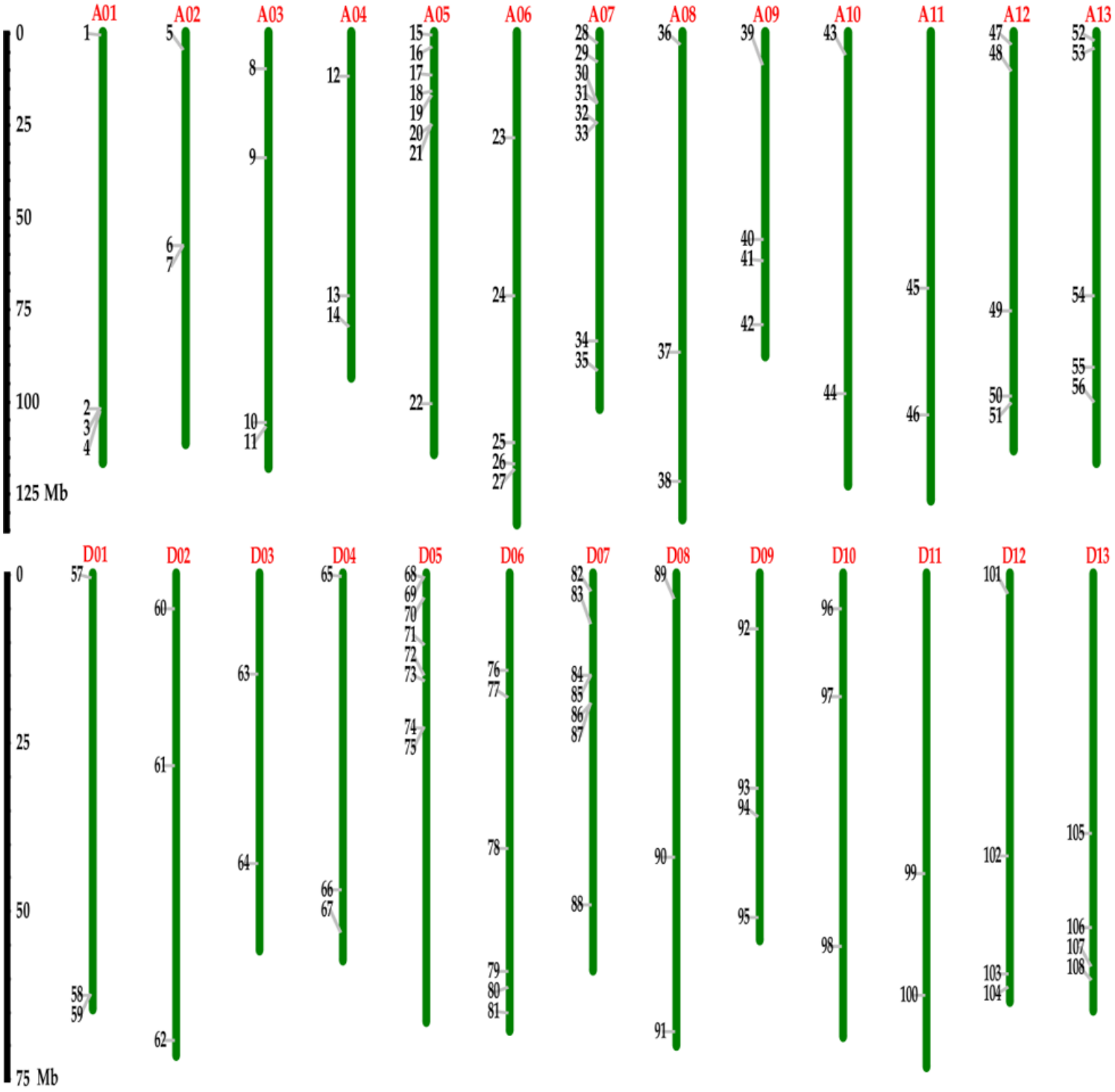

Figure 2

Chromosomal location of PHD genes on 26 chromosomes in G. hirsutum. The chromosome numbers were shown on the top of each chromosome. The scale bar indicated the length in megabases (Mb).

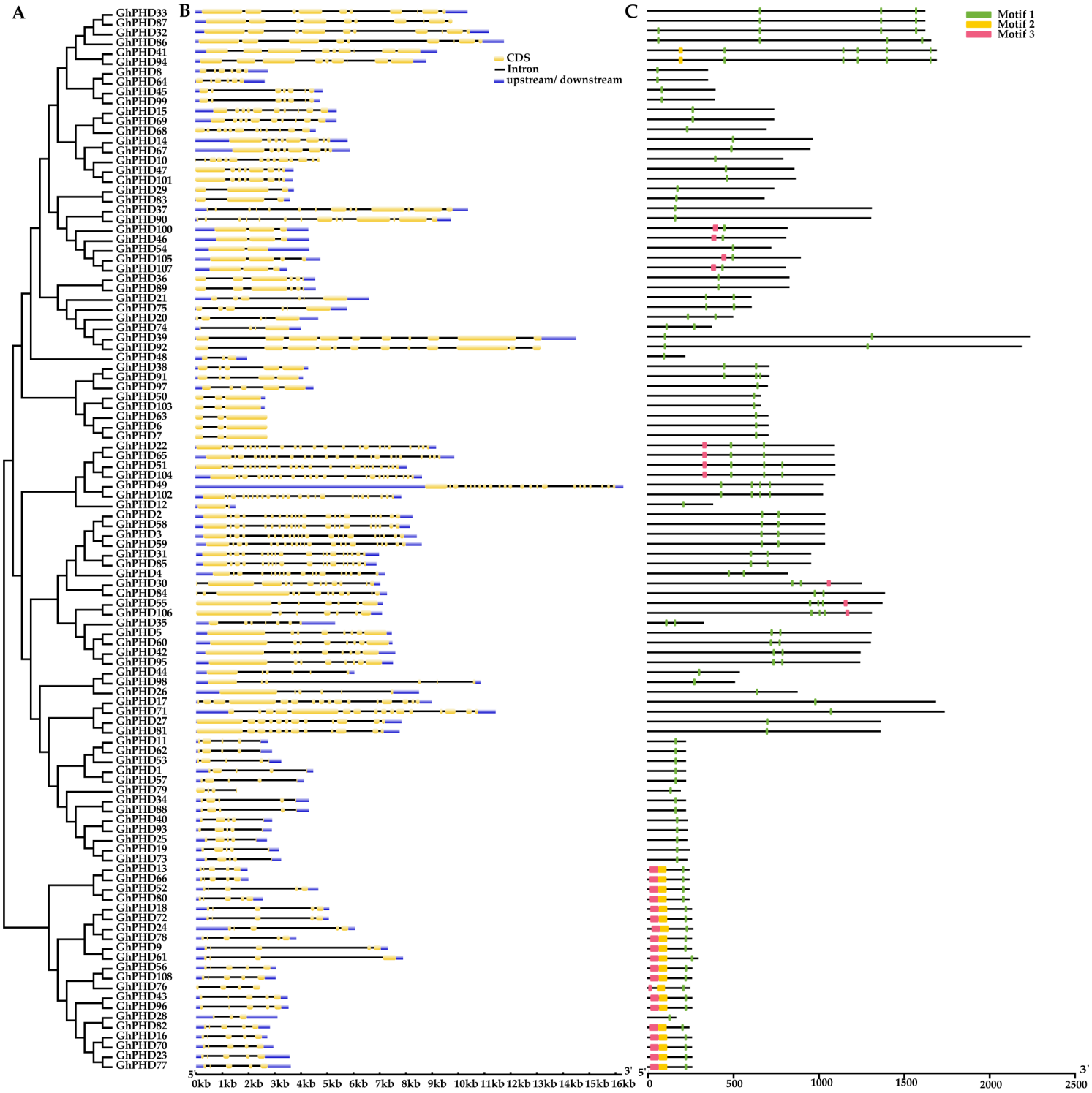

Figure 3

Phylogenetic tree, gene structure, and conserved motif analysis of GhPHD proteins. A An unrooted phylogenetic tree was generated in MEGA 6.0 by neighbor-joining method. B Exon-intron structure of GhPHD genes. The yellow boxes represent exons, black lines represent introns, and blue boxes represent the upstream/downstream UTRs. The sizes of exons and introns can be estimated using the scale bar at the bottom. C Motifs distribution of GhPHD proteins and different motif boxes are in different colors (motif 1 to 3). Motif 1 in (c) is the PHD domain.

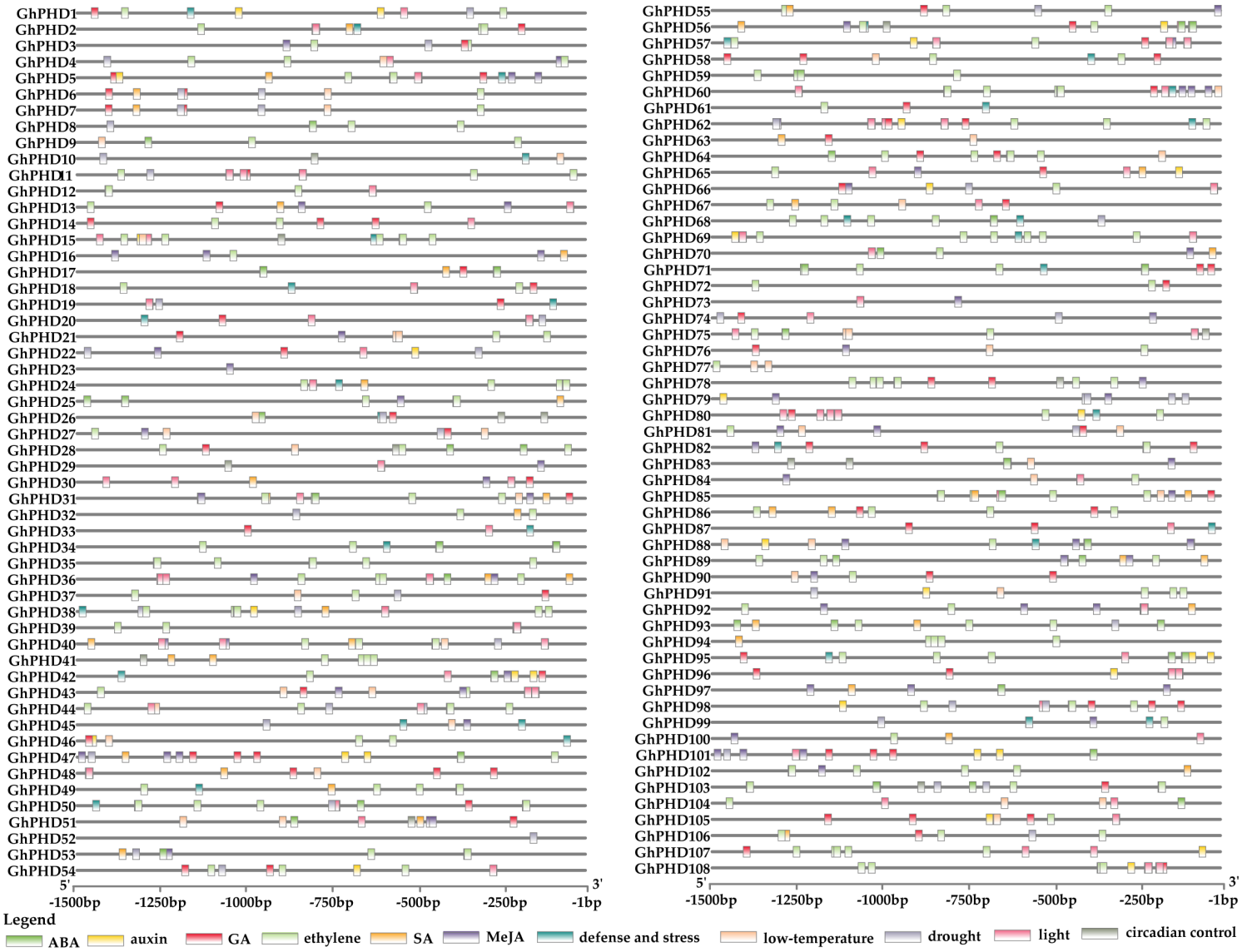

Figure 4

Distribution of major stress-related and plant hormone-related cis-acting elements in the promoter regions of GhPHD genes. The locations of these cis-acting elements were confirmed using the PlantCARE database. Different cis-acting elements were represented by different color boxes.

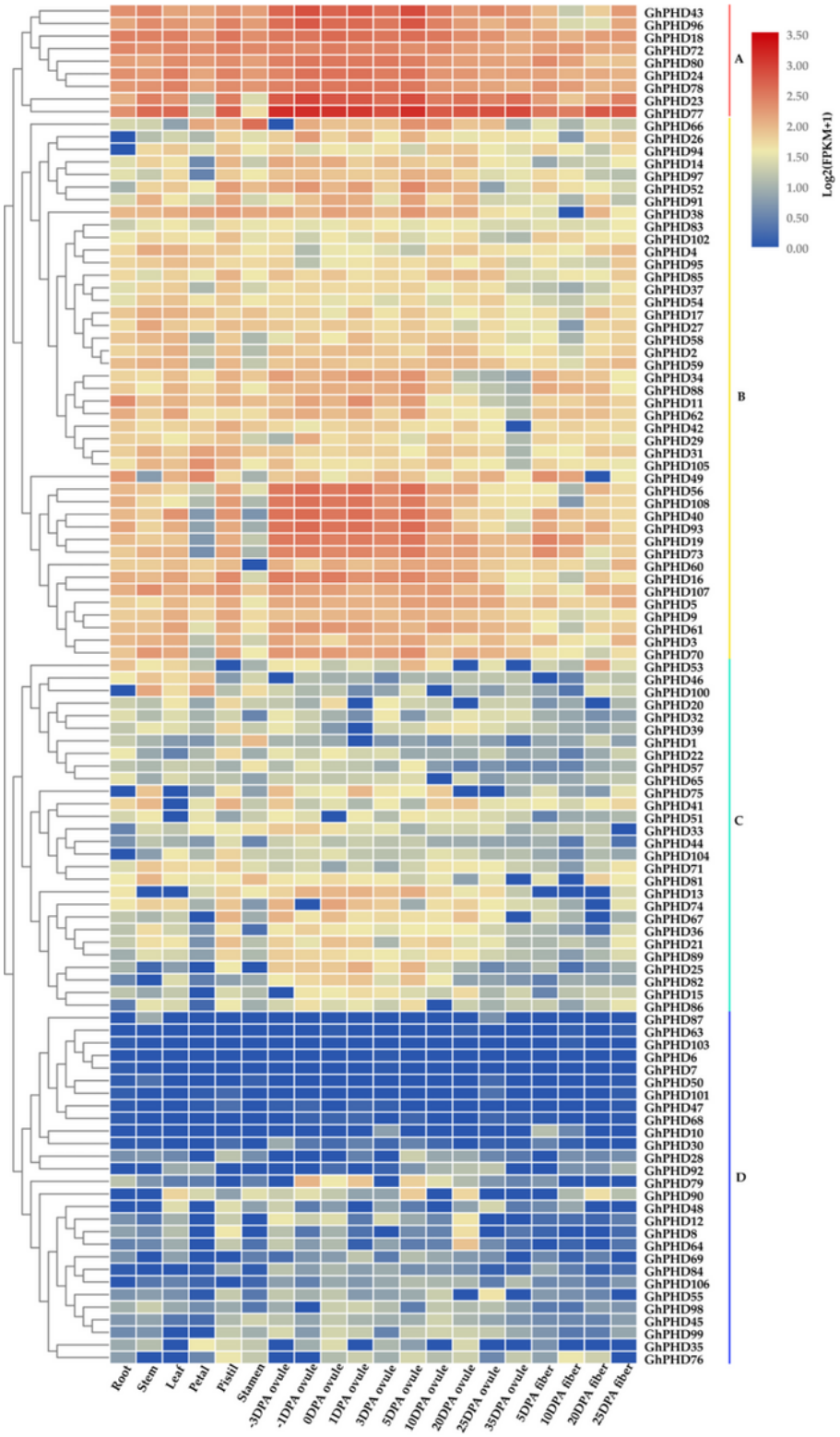

Figure 5

Expression patterns of GhPHD genes in cotton tissues. A heatmap indicates the clustering of 108 GhPHD genes in eight tissues (shown on the bottom). DPA is days post anthesis. Gene names are shown on the right. Scale bars at the top show Log2 (FPKM+1) values of each gene.

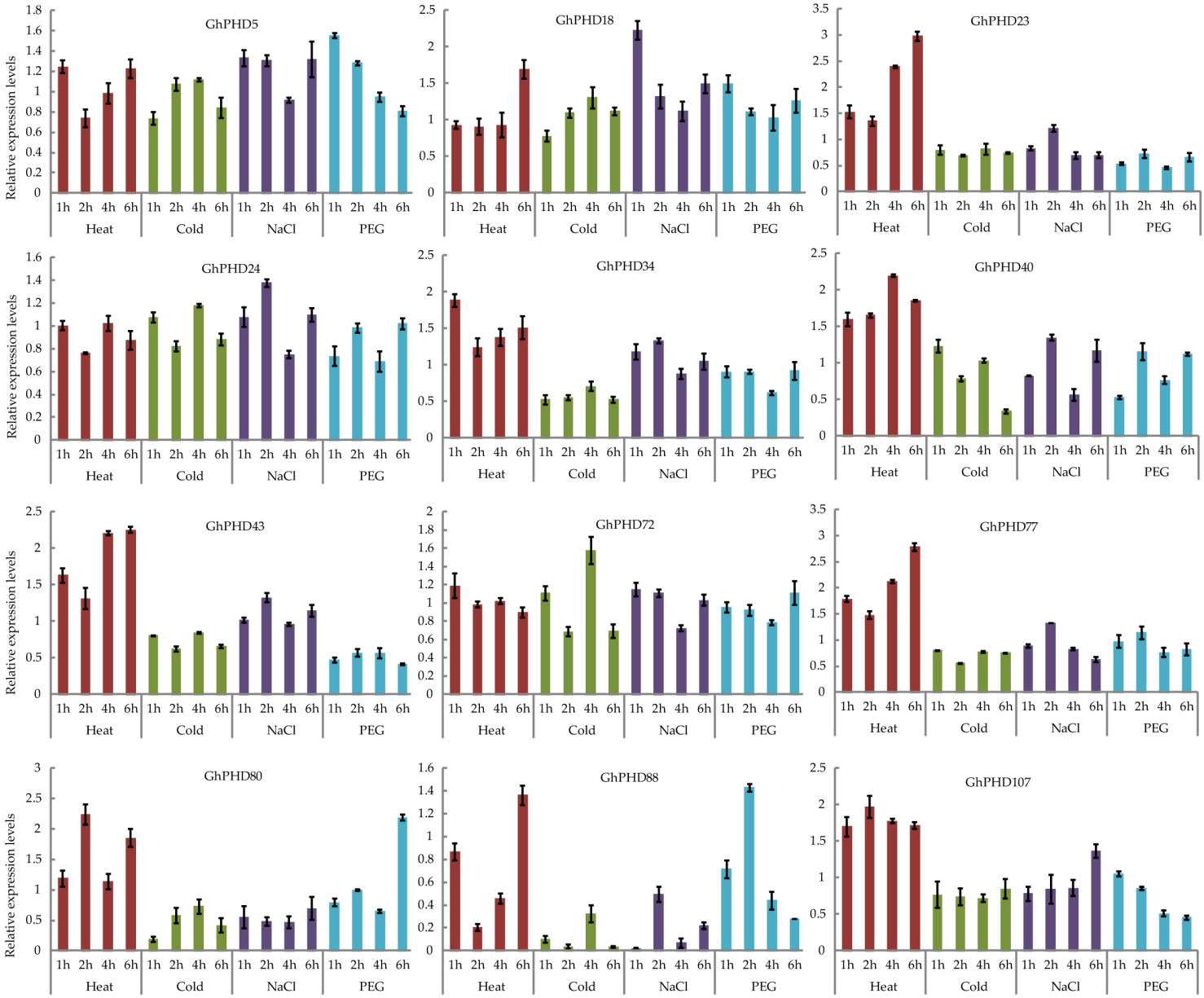

Figure 6

The relative expression levels of 12 GhPHD genes under heat, cold, salt, and drought. GhHis3 was used as an internal reference gene. The error bars represent the standard deviations of three experiments.

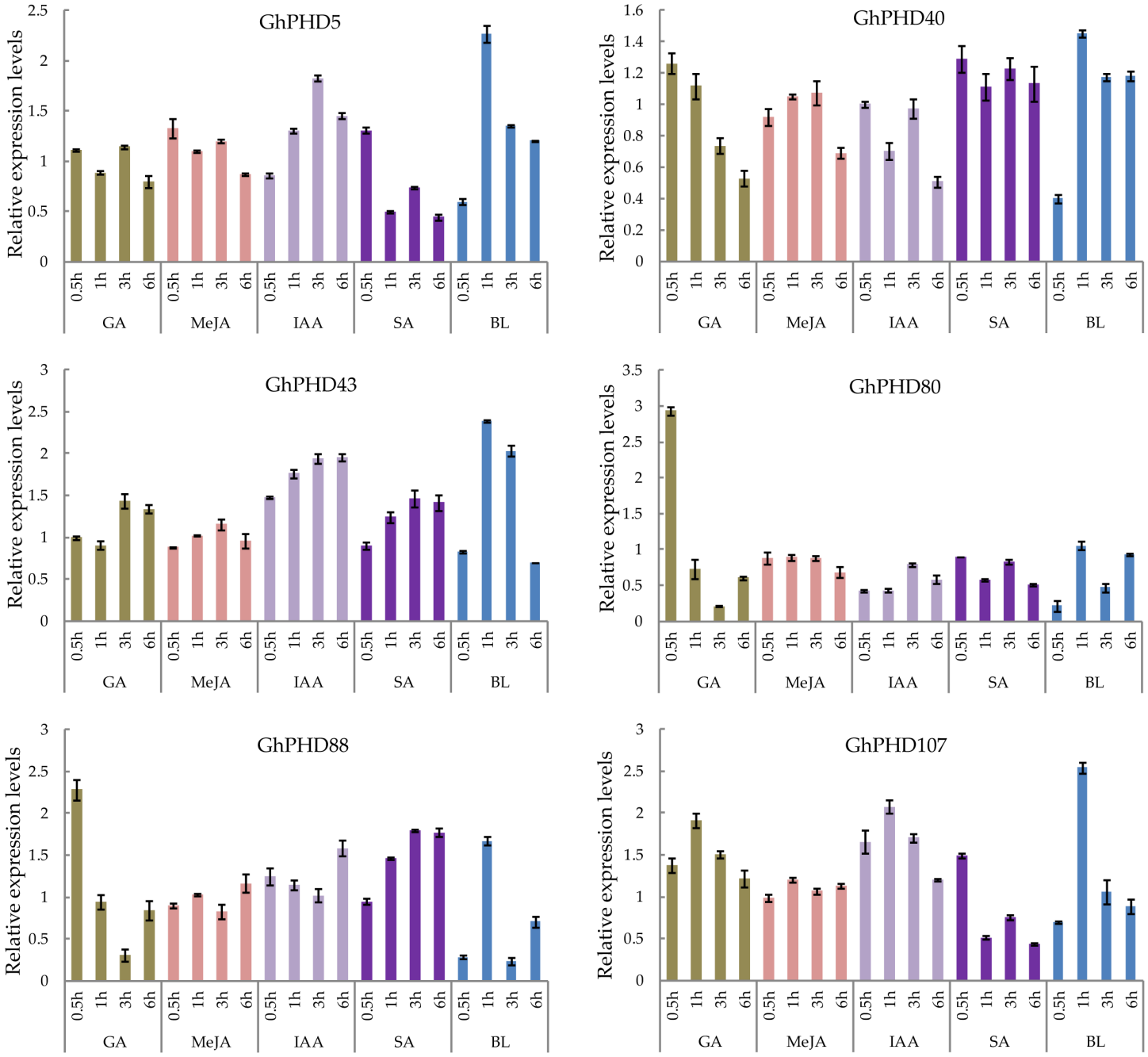

Figure 7

Expression pattern analysis of GhPHD genes under GA, MeJA, IAA, SA, and BL treatments by qRT-PCR. The expression levels of six GhPHD genes were estimated by qRT-PCR. 0.5 h, 1 h, 3 h, and 6 h indicate hours after hormone treatments. The error bars show the standard deviation of three biological replicates.

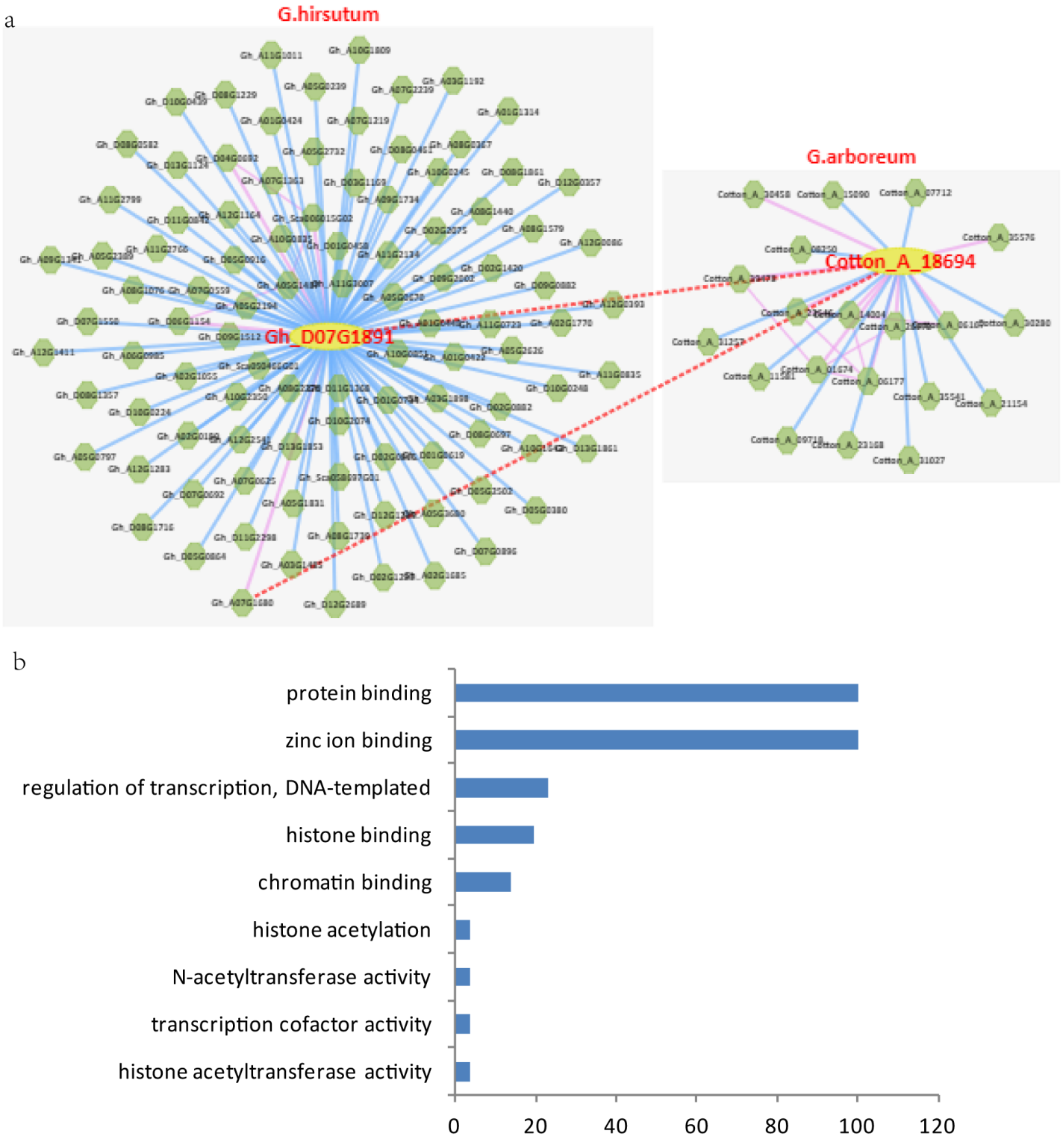

Figure 8

Co-expression networks analysis of GhPHD88 and GO enrichment analysis of 108 GhPHDs. A Co-expression network analysis of GhPHD88 with functional modules for G. hirsutum and G. arboreum. Yellow and green color indicates that query protein and interaction proteins, respectively. There are four interaction lines, red lines indicates ortholog gene pairs in G. hirsutum and G. arboreum; pink lines and blue lines indicates proteins own interaction and positive/negative co-expression relationship with target

protein; orange lines indicates proteins own interaction and protein-protein relationship with target protein. B GO enrichment analysis of all GhPHD genes.

Supplementary Files

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

- FigureS1.tif
- TableS5.xlsx
- TableS7.docx
- TableS6.xlsx
- TableS1.xlsx
- TableS3.docx
- TableS2.docx
- TableS4.xlsx
- FigureS2.tif