Genome-wide identification of *Gossypium INDETERMINATE DOMAIN* genes and their expression profiles in ovule development and abiotic stress responses

ALI Faiza[1,2], QANMBER Ghulam[1,2], LI Yonghui[1,2], MA Shuya[1,2], LU Lili[1,2], YANG Zuoren[1,2], WANG Zhi[1,2]* and LI Fuguang[1,2]*

Abstract

Background: *INDETERMINATE DOMAIN (IDD)* transcription factors form one of the largest and most conserved gene families in plant kingdom and play important roles in various processes of plant growth and development, such as flower induction in term of flowering control. Till date, systematic and functional analysis of *IDD* genes remained infancy in cotton.

Results: In this study, we identified total of 162 *IDD* genes from eight different plant species including 65 *IDD* genes in *Gossypium hirsutum*. Phylogenetic analysis divided *IDDs* genes into seven well distinct groups. The gene structures and conserved motifs of *GhIDD* genes depicted highly conserved exon-intron and protein motif distribution patterns. Gene duplication analysis revealed that among 142 orthologous gene pairs, 54 pairs have been derived by segmental duplication events and four pairs by tandem duplication events. Further, *Ka/Ks* values of most of orthologous/paralogous gene pairs were less than one suggested the purifying selection pressure during evolution. Spatiotemporal expression pattern by qRT-PCR revealed that most of the investigated *GhIDD* genes showed higher transcript levels in ovule of seven days post anthesis, and upregulated response under the treatments of multiple abiotic stresses.

Conclusions: Evolutionary analysis revealed that *IDD* gene family was highly conserved in plant during the rapid phase of evolution. Whole genome duplication, segmental as well as tandem duplication significantly contributed to the expansion of *IDD* gene family in upland cotton. Some distinct genes evolved into special subfamily and indicated potential role in the allotetraploidy *Gossypium hisutum* evolution and development. High transcript levels of *GhIDD* genes in ovules illustrated their potential roles in seed and fiber development. Further, upregulated responses of *GhIDD* genes under the treatments of various abiotic stresses suggested them as important genetic regulators to improve stress resistance in cotton breeding.

Keywords: Upland cotton, *IDD* transcription factor, Gene duplication, Collinearity, Spatiotemporal expression, Abiotic stresses

* Correspondence: wangzhi.12@163.com; aylifug@163.com
[1]State Key Laboratory of Cotton Biology, Institute of Cotton Research,
Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China
Full list of author information is available at the end of the article

Background

Transcription factors containing DNA binding domains play an important role in many biological processes in almost all living organisms. They function as either repressors or activators, depending on whether they inhibit or stimulate the transcription of target genes. Transcription factors of the same family generally have distinct actions because of differences in their domains and protein regions that tend to diverge from one another (Eveland et al. 2014).

According to the quantity and arrangement of cysteine (C) and histidine (H) residues, the transcription factors containing zinc fingers fall into five classes (C2H2, C3H, C2C2 (GATA finger), C3HC4 (RING finger), and C2HC5 (LIM finger)) (Moreno-Risueno et al. 2015). As one of the largest transcription factor families, C2H2 zinc-finger transcription factors are structurally characterized by the amino acid sequence F/Y-X-C-X2–5-C-X3-F/Y-X5-Ψ-X2-H-X3–5-H, where X is any amino acid while Ψ represents a hydrophobic residue (Fan et al. 2017). Two cysteine (C) and histidine (H) residues coordinate a zinc ion and interact with the major groove of DNA by folding two β-sheets and one α-helix (Lee et al. 1989; Parraga et al. 1988). INDETERMINATE (IDD) (Riddick and Simmons 2014) gene family encoding transcription factors containing a C2H2 (Cys2His2) zinc-finger domain (Colasanti et al. 2006) have been investigated to involve in animals (Riechmann et al. 2000; Takatsuji 1998). Previously, it has been reported that zinc-finger family was only 19% conserved among other eukaryotes except plants (Englbrecht et al. 2004; Pabo and Sauer 1992) suggesting that extensive duplication resulted in the expansion of zinc-finger gene family in plants (Coelho et al. 2018).

It's known that IDD proteins have multiple functions in plant development. In maize (Zea mays), three IDDs have been characterized. ID1 gene was first reported to induce phase transition from vegetative to reproductive growth in maize (Colasanti et al. 1998). In rice, OsID1/Ehd2/RID1 has also been found to play an important role in mediating flower initiation besides vegetative to reproductive growth phase transition (Colasanti et al. 2006; Matsubara et al. 2008; Park et al. 2008; Wong and Colasanti 2007; Wu et al. 2008). Furthermore, OsIDD10 is involved in ammonium absorption and nitrogen metabolism (Xuan et al. 2013). In Arabidopsis, 16 IDD genes were identified (Colasanti et al. 2006). Among them, AtIDD8 and AtIDD14 play an important role in sugar and starch metabolism (Ingkasuwan et al. 2012). AtIDD8 is phosphorylated by AKIN10 and its loss of function mutant idd8–3 exhibited later flowering in Arabidopsis. Moreover, SnRK1 interacts with AtIDD8 to control sugar metabolism during the flowering transition (Jeong et al. 2015). Similarly, AtIDD15 has been reported to participate in sugar and starch metabolism (Tanimoto

et al. 2008), as well as in gravitropic response, while AtIDD3 and AtIDD8 are involved in root development (Ingkasuwan et al. 2012). AtIDD10 (JKD) is essential for the precise expression of GL2 (GLABRA2), CPC(CAPRICE), and WER(WEREWOLF) and has been proposed that JKD acts in the cortex to define root hair cells in the epidermis (Hassan et al. 2010). Moreover, AtIDD9 plays a role in epidermal cell fate specification (Long et al. 2015a; Long et al. 2015b). Additionally, AtIDD3 binds to the SCL3 promoter to control plant development, and regulate the expression of downstream genes in gibberellin (GA) signaling dependent manner (Yoshida et al. 2014). AtIDD14, AtIDD15, and AtIDD16 regulate the expression of genes involved in auxin biosynthesis, thereby influencing organ morphogenesis (Cui et al. 2013).

Cotton (Gossypium hirsutum L.) is the preeminent source of natural fiber and is cultivated worldwide (Rinehart et al. 1996). It provides important raw material for textile industry. However, low fiber quality and yield are the main limiting factors affecting its overall world contribution and consumption. Cotton faces several environmental and abiotic stresses that restrict its growth and productivity. The roles of IDDs have been well-described in the growth and development of model plants like Arabidopsis, rice and maize. However, investigation of IDD genes in upland cotton remained elusive. Present study shows the systematic analysis of IDD genes in G. hirsutum using a genome-wide structure depiction, spatiotemporal expression patterns and stress responses investigations. Total of 65 GhIDD gene family members were identified and further characterized to explore the phylogenetic relationships, chromosome locations, gene duplication, gene structures, conserved motifs and spatiotemporal expression patterns and responses of GhIDD genes under various abiotic stresses. This study will help to understand the evolution of GhIDD genes and provide the foundation to explore the functional mechanism of GhIDD genes in plant growth, fiber development and abiotic stress tolerance in cotton.

Methods

Identification and chemical characterization of IDD family members

The protein sequences of 16 IDD genes from Arabidopsis thaliana were used as queries for the computational identification of IDD genes in Gossypium arboreum (ICR, version 1.0), G. hirsutum (NAU, version 1.1), G. raimondii (JGI, version 2.0), Oryza sativa (version 7.0), Zea mays (version 1.1), Physcomitrella patens (moss) (version 3.3), Selaginella moellendorffii (fern) (version 1.0), Theobroma cacao (version 1.1) and Chlamydomonas reinhardtii (algae) (version 1.0). The genome databases were downloaded from Phytozome (version 11) (https://phytozome.jgi.doe.gov/pz/portal.html) for all species except for G.

arboreum, G. hirsutum, G. raimondii and *A. thaliana.* The *G. arboreum* genome was downloaded from a publicly available online resource (ftp://bioinfo.ayit.edu.cn/downloads/), while the *G. hirsutum* and *G. raimondii* databases were downloaded from COTTONGEN (https://www.cottongen.org/). The *A. thaliana* database was downloaded from TAIR 10 (http://www.arabidopsis.org). The putative *IDD* protein sequences retrieved by Local BLASTP were further confirmed by using SMART (Letunic et al. 2015) (http://smart.emblheidelberg.de/), and InterProScan 63.0 program (http://www.ebi.ac.uk/InterProScan/) and Hidden Markov model (HMM) (Jones et al. 2014). Gene IDs and names were listed or given according to the positions on chromosomes (Additional file 1: Table S1). ExPASy ProtParam tool (http://us.expasy.org/tools/protparam.html) was employed to predict the biophysical characteristics and protein localization of all *GhIDDs*.

Phylogenetic tree construction and conserved *IDD* sequences analyses

Full length protein sequences of *IDD* genes from eight species (*G. hirsutum, G. arboreum, G. raimondii, T. cacao, A. thaliana, O. sativa, P. patens,* and *S. moellendorffii*) were aligned to test a phylogenetic tree by MEGA 7.0 program using ML (Maximum Likely hood) method (Kumar et al. 2016). To test the tree, bootstrap method with 1 000 repeats and 50% cutoff values were used. Further, two other phylogenetic trees of 110 *IDD* genes from three cotton species (*G. hirsutum, G. arboreum, G. raimondii*) and 65 *IDD* genes from *G. hirsutum* were also constructed using NJ (neighbor-joining) method (Kumar et al. 2016) by MEGA 7.0 program. Next, for conserved sequence logos analysis, multiple sequence alignment of IDD proteins of *A. thaliana,* rice, and upland cotton (*G. hirsutum*) was performed with Clustal X 2.0, and the results were subjected to WEB-LOG online program (Crooks et al. 2004) to visualize conserved amino acid sequence logos.

Analyses of gene structures and conserved motifs

We performed an exon–intron structural and conserved motif analysis of 65 *IDD* gene of *G. hirsutum.* Sequences were first aligned using Clustal X 2.0, and then a phylogenetic tree was constructed using the NJ method by MEGA 7.0 program. To examine gene structures, the BED file along with the data from the NJ phylogenetic tree were subjected to GSDS 2.0 (Gene Structure Display Server 2.0) online tool (Hu et al. 2015) (http://gsds.cbi.pku.edu.cn/index.php). Motifs were examined by submitting full-length protein sequences to the MEME online program (Bailey et al. 2006) (http://memesuite.org/tools/meme), with parameters as described previously (Li et al. 2019).

Chromosomal mapping, gene duplication and *Ka/Ks* values

The chromosomal positions of *GhIDDs* were obtained from cotton genome annotation file (ftp://ftp.bioinfo.wsu.edu/species/Gossypium_hirsutum/NAUNBI_G), and gff3-file was extracted. The physical localization of *GhIDD* genes was mapped by using MapInspect program (Jia et al. 2018) (http://www.plantbreeding.wur.nl/UK/software_mapinspect.html) to visualize the distribution of the *GhIDD* genes on corresponding chromosomes. Orthologous and paralogous gene pairs of the *GhIDD* genes were obtained by all-versus-all BLASTP searches (Altschul et al. 1990). The blastp results were then analyzed by MCscan, which generated collinearity blocks for the cotton *IDD* genes between and within At and Dt sub-genomes of upland cotton. The collinear pairs of *IDD* genes generated by MCscan were used to construct a collinearity map of *IDD* genes using CIRCOS software (Krzywinski et al. 2009). To estimate *Ka/Ks* values, the amino acid sequences of orthologous gene pairs were first aligned by Clustal X2.0 and then converted to cDNA sequences using PAL2NAL program (Suyama et al. 2006) (http://www.bork.embl.de/pal2nal/). Further, non-synonymous (Ka) and synonymous (Ks) divergence level values were calculated by CODEML program of the PAML package (Yang 2007).

RNA-seq data analysis of *GhIDD* genes

To determine the expression patterns of the *GhIDD* genes in 22 different tissues (vegetation, reproduction and fiber) of cotton, we used publicly available high-throughput microarray data (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4482290/). TopHat and cufflinks were used to analyze the RNA-seq expression and the gene expressions were uniformed in fragments per kilobase million (FPKM) (Trapnell et al. 2012). The *IDDs* expression values were extracted from the expression data. Genesis software was used to generate the heat map (Sturn et al. 2002) of *IDDs* expression in various tissues and responses to abiotic stresses including cold, hot, salt (300 mmol·L^{-1} NaCl) and 10% PEG 6000.

Plant material and treatments

Cotton seeds of CCRI24 were obtained from the Institute of Cotton Research of Chinese Academy of Agricultural Sciences. To analyze spatial and temporal expression patterns of genes, the different plant tissues such as root, stem, leaf, flower, ovules of 1, 3, 5, 7, 10, 15 and 20 DPA (day post anthesis) as well as fiber tissues of 7, 10, 15 and 20 DPA were collected for the RNA preparation from cotton plants, grown under field conditions (Zhengzhou, China). To investigate the expression of *GhIDD* genes under abiotic stresses, seeds were germinated on a wet filter paper for 3 days at 28 °C, and

seedlings were transferred to a liquid culture medium (Yang et al. 2014). At the 3-leaf stage, the seedlings were treated at 4 °C and 38 °C for cold and heat stress, and with 10% PEG 6000 and 300 mmol·L^{-1} NaCl, respectively; the true leaves were sampled at 0, 1, 2, 4, and 6 h of the treatments. The total RNA was extracted using RNAprep Pure Plant Kit (TIANGEN, Beijing, China), as per the manufacturer's instructions. The first strand cDNA was synthesized using a Prime Script® RT reagent kit (Takara, Dalian, China). SYBR Premix Ex Taq™ II (Takara) was used for PCR amplifications. Premix Ex Taq™ II (Takara) was used along with the Light Cycler 480 system (Roche Diagnostics, Mannheim, Germany) for Real-time PCR. For each analysis, qRT-PCR assays had three biological replicates, each consisting of three technical replicates. Histone 3 from cotton (GeneBank, accession number AF024716) was used as an internal control (Wan et al. 2016). The relative fold difference value (N) was calculated as follows: $N = 2 - \Delta\Delta C_t = 2 - (\Delta C_{t\ treated} - \Delta C_{t\ control})$, where $\Delta\Delta C_t = \Delta C_t$ of the treated sample $- \Delta C_t$ of the untreated control sample. The primers used in this study were enlisted in Additional file 1: Table S2.

Results

Genome-wide identification of *IDD* genes

We identified total of 162 genes in 8 investigated plant species including monocots (*O. sativa*), dicots (*A. thaliana, G. hirsutum, G. arboreum, G. raimondii,* and *T. cacao*), ferns and moss. However, no *IDD* gene family member was identified in algae. Among these, 65 *IDD* genes were confirmed in *G. hirsutum,* 22 in *G. arboreum,* 23 in *G. raimondii,* 15 in *T. cacao,* 12 in *O. sativa,* 7 in moss, and 2 in fern. Higher number of *IDD* genes was identified in *G. hirsutum* than that in *G. arboreum, G. raimondii, T. cacao*, rice, moss, fern and *Arabidopsis* indicating polyploidization and duplication effect on *GhIDD* genes in *G. hirsutum.*
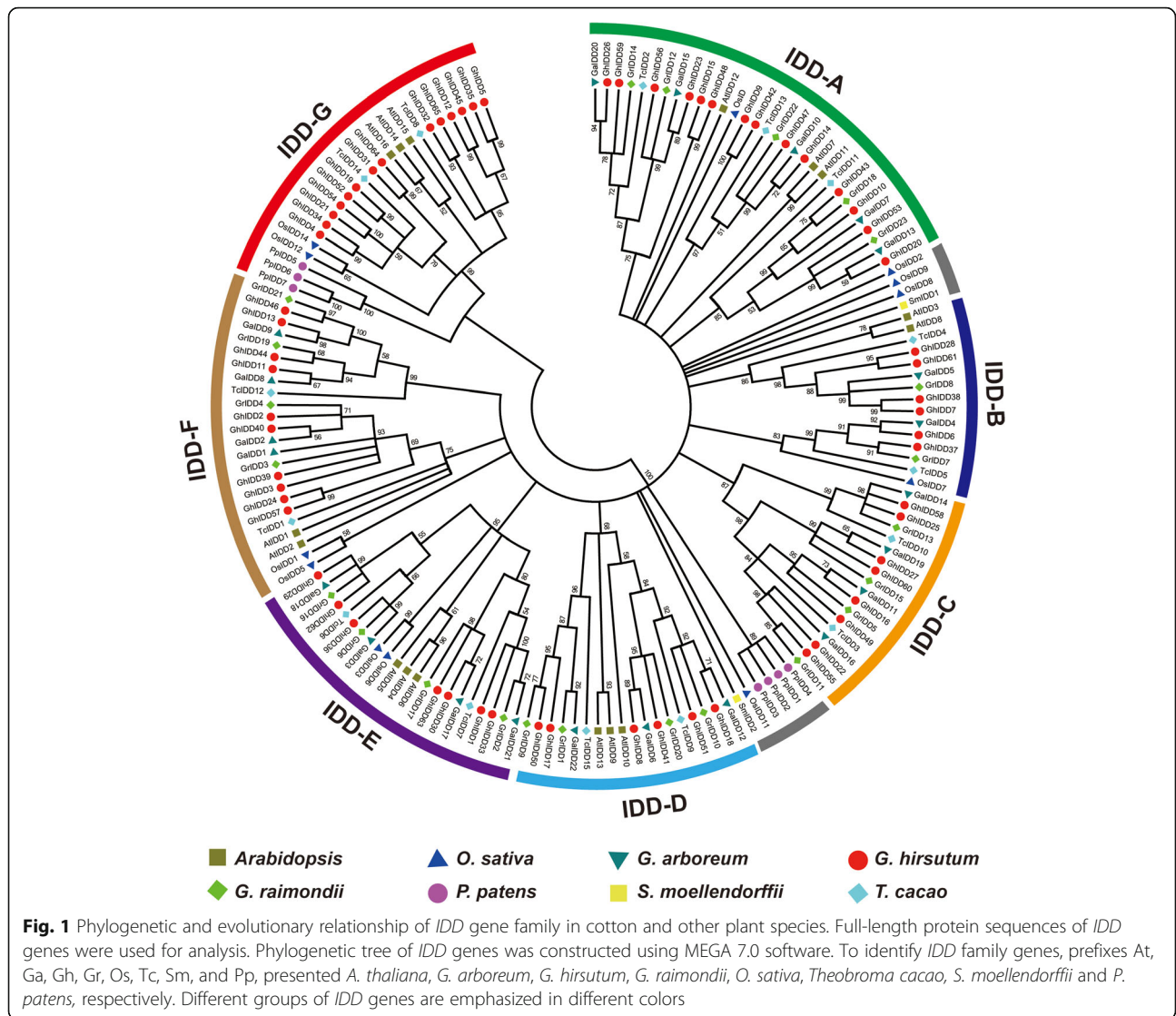
Phylogenetic analysis of *IDD* gene family

To determine the phylogenetic relationships among *IDDs* and explore both conserved and diversified functions of this TF family, a phylogenetic tree by ML method using MEGA 7.0 software was constructed among 162 *IDD* genes. To indicate the *IDD* genes from *A. thaliana, G. arboreum, G. hirsutum, G. raimondii, O. sativa, S. moellendorffii, P. patens* and *T. cacao,* the prefixes At, Ga, Gh, Gr, Os, Sm, Pp, and Tc were used, respectively. The phylogenetic analysis divided the 162 *IDD* genes into seven well distinct groups (Fig. 1). Group IDD-A contained the maximum number of *IDD* genes (31 genes) from all species while group IDD-B have the minimum number of *IDD* genes (15 genes). Groups IDD-A, IDD-B, IDD-C, IDD-D, IDD-E, and IDD-F

contained *IDD* genes from monocot and dicot but not from moss and fern, indicating that these groups might be evolved after separation of ferns and moss from monocot and dicot plant species. Group IDD-F contained *IDD* genes from monocot, dicot and fern but lack moss *IDD* genes illustrating the divergence of these *IDD* genes after the division of moss from monocots, dicots and ferns. However nine *IDD* genes (*OsIDD2, OsIDD8, OsIDD9, OsIDD11, SmIDD1, PpIDD1, PpIDD2, PpIDD3,* and *PpIDD4*) from *O. sativa, S. moellendorffii* and *P. patens* did not fall in any group, indicating their potential special functions in the associated species evolution and development. *S. moellendorffii* and *P. patens* are resurrection plants which can tolerate extreme dehydration. *O. sativa* is a kind of semi-aquatic crop. All these indicated that the nine ungrouped genes may play some especial roles in the evolution from aquatic to terrestrial organisms.

Moreover, the phylogenetic tree results depicted the close relationship among cotton and cacao *IDD* genes, as the genes from these two species were found to be closely clustered to each other in different groups and subgroups of phylogenetic tree (Fig. 1). However, the number and distribution of *IDD* genes in cacao and cotton were different in all groups. For instance, in group IDD-G, 14 *GhIDD* genes showed a close relationship with two cacao *IDD* genes (*TcIDD8* and *TcIDD14*), also supporting the hypothesis that cacao and cotton were closely related and probably derived from the same ancestors (Li et al. 2014).

To further investigate the evolutionary relationship of cotton *IDD* genes from *G. hirsutum, G. arboreum,* and *G. raimondii,* a phylogenetic tree within three cotton species using NJ method was generated (Fig. 2). The phylogenetic tree divided all *IDD* genes of three cotton species into four groups. Group IDD-b contained more *IDD* genes (38) while group IDD-d depicted less *IDD* gene family members (14). In group IDD-a, IDD-b and IDD-c, all paralogous and orthologous genes from the allotetraploid and corresponding diploid cotton clustered together. Group IDD-d exhibited 14 *IDD* genes only from *G. hirsutum,* which showed that it is far away from its two ancestor species (*G. arboreum* and *G. raimondii*) and may come from the new gene duplication and genome polyploidization, reconfirming the results that these *GhIDD* genes might be evolved after divergence from the common ancestors of cotton and cacao (Fig. 2).

Furthermore, to explore the evolutionary relationship and potential function catalogue among *G. hirsutum IDD* genes, another phylogenetic tree was constructed by NJ method. Total of 65 *GhIDD* genes were divided into five (IDD-a, IDD-b, IDD-c, IDD-d, and IDD-e) groups (Additional file 2: Figure S1). Group IDD-a was the biggest group with 21 *GhIDD* genes, however group

Fig. 1 Phylogenetic and evolutionary relationship of *IDD* gene family in cotton and other plant species. Full-length protein sequences of *IDD* genes were used for analysis. Phylogenetic tree of *IDD* genes was constructed using MEGA 7.0 software. To identify *IDD* family genes, prefixes At, Ga, Gh, Gr, Os, Tc, Sm, and Pp, presented *A. thaliana*, *G. arboreum*, *G. hirsutum*, *G. raimondii*, *O. sativa*, *Theobroma cacao*, *S. moellendorffii* and *P. patens*, respectively. Different groups of *IDD* genes are emphasized in different colors

IDD-b was the smallest with 6 *GhIDD* genes in it. Group *IDD-c* and *IDD-d* contained 16 and 8 genes, respectively. In group *IDD-e*, all (14) *GhIDD* genes are same with that in *IDD-d* of Fig. 2, which showed consistency in our analysis and strengthened the hypothesis that these *IDD* genes might originate from common ancestors of cotton and cacao.
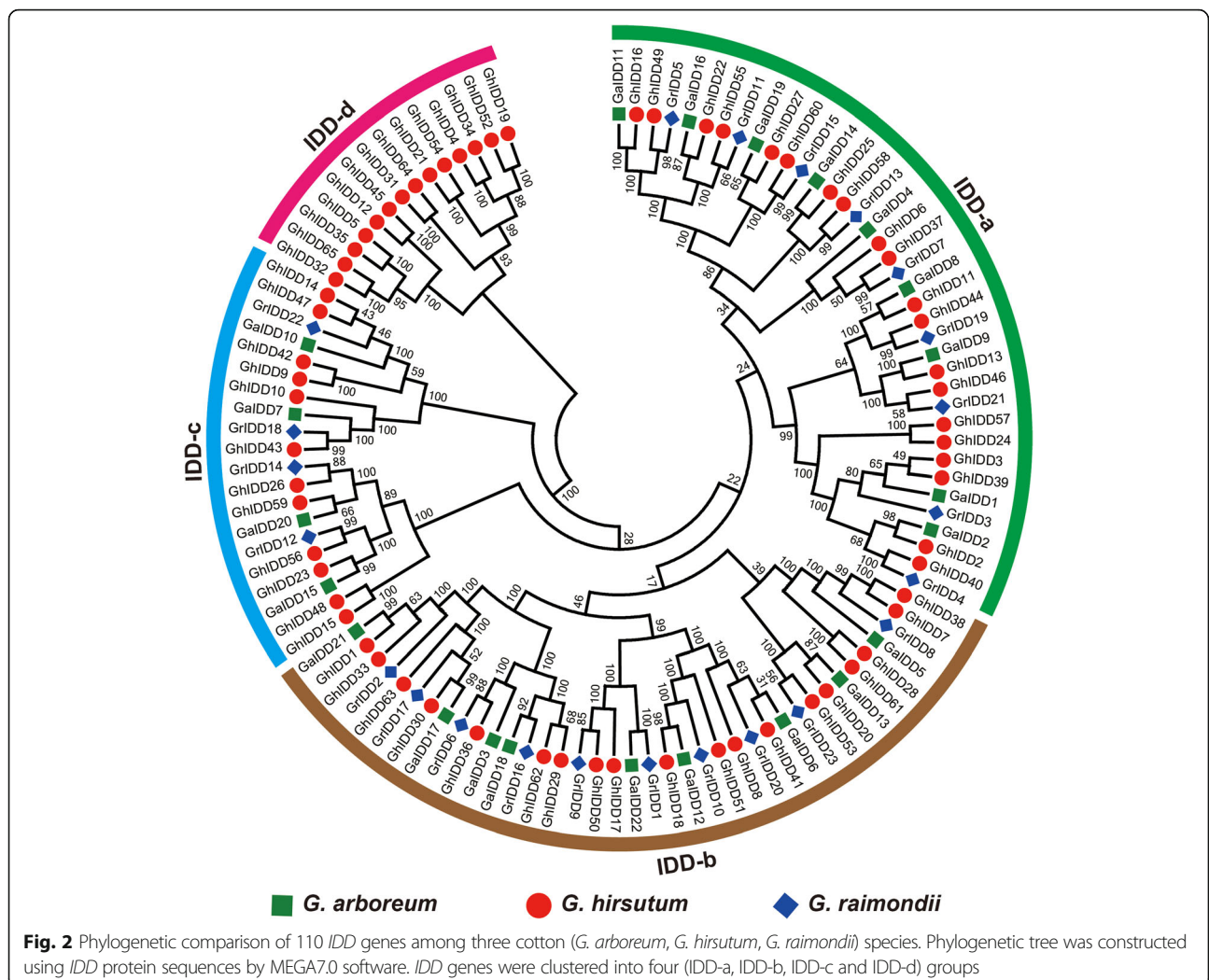
Biophysical characteristics of *GhIDD* genes

We predicted the biophysical characteristics of all the members of *GhIDD* gene family in *G. hirsutum*. The details of biophysical properties including chromosomal position (start and end points), coding sequence (CDS), number of amino acids (protein length), molecular weight (MW), isoelectric point (pI), and grand average of hydropathicity (GRAVY) of *GhIDD* genes are provided in Additional file 1: Table S3.

The results indicated that *GhIDD* coding sequence ranged from 1 140 bp to 2 418 bp for *GhIDD37* and *GhIDD42*, respectively. Similarly, the numbers of amino acids in the predicted protein sequences of *GhIDD* genes ranged from 379 to 805 for same genes. Molecular weights ranged from 41 310.77 to 89 465.69 kDa for *GhIDD42* and *GhIDD13*, respectively. Isoelectric point of *GhIDD41* was the highest (9.68) and that of *GhIDD60* was the lowest of 8.37. The grand averages of hydropathicity values of all *GhIDD* genes were less than 0, ranging from − 0.843 to − 0.62 for *GhIDD64* and *GhIDD18*, respectively. In addition, the predicted subcellular localization of the *G. hirsutum* IDD proteins were all in nuclear (Additional file 1: Table S3).

Gene structure and conserved motif analysis

To deeply understand the phylogenetic relationships, gene variation and potential protein function among *G.*
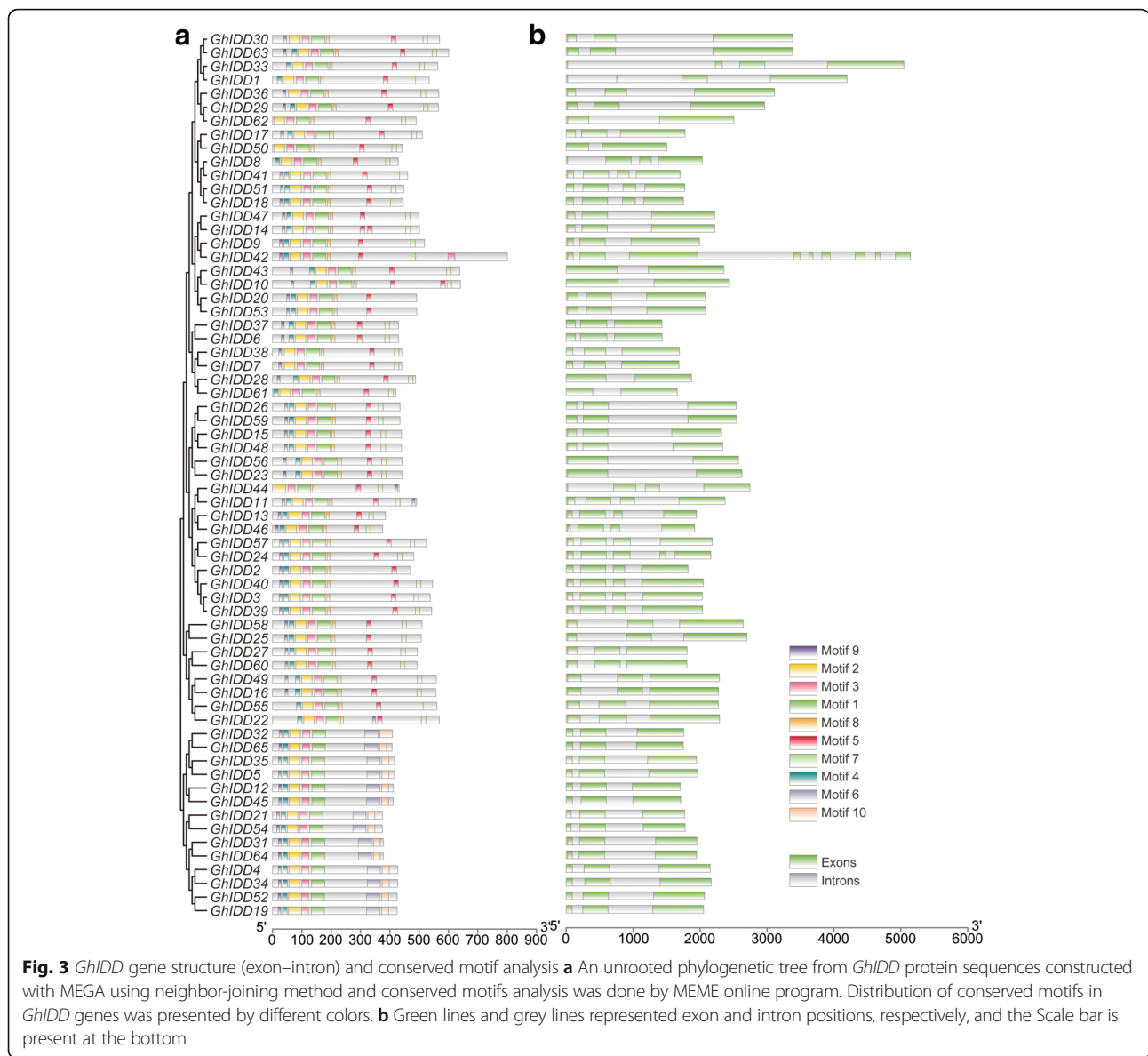
Fig. 2 Phylogenetic comparison of 110 *IDD* genes among three cotton (*G. arboreum*, *G. hirsutum*, *G. raimondii*) species. Phylogenetic tree was constructed using *IDD* protein sequences by MEGA7.0 software. *IDD* genes were clustered into four (IDD-a, IDD-b, IDD-c and IDD-d) groups

hirsutum IDD genes, the intron–exon structure and conserved motifs analysis were performed (Fig. 3). It was observed that *GhIDD* genes showing similar intron–exon numbers and distribution patterns were clustered into the same group. The numbers of introns in *GhIDD* genes ranged from one to eight. Here, the genes with one intron accounted for 12% of the total *GhIDD* genes whereas only one gene (*GhIDD42*) had eight introns (Fig. 3a). To investigate the conserved motif distribution pattern of *G. hirsutum IDD* genes, another unrooted tree was constructed coupled with MEME program. The results illustrated that most of the *GhIDD* proteins displayed similar motifs distribution pattern, as motif 1, 2, and 3 were present in almost all proteins (Fig. 3b). Motif 6 and 10 were only present in the 14 proteins of group *GhIDD-e* of Additional file 2: Figure S1, however, in which motif 5, 7, 8 were absent. Moreover, motif 4 was not identified in *GhIDD7, GhIDD30, GhIDD36, GhIDD38, GhIDD44, GhIDD50,* and *GhIDD62*. Motif 9 was present in all *GhIDD* genes except in seven (GhIDD1, GhIDD8,

GhIDD33, GhIDD50, GhIDD55, GhIDD61, and GhIDD62) proteins. In general, *GhIDD* genes with similar motif distribution pattern occupied the position in same group or subgroup of phylogenetic tree.

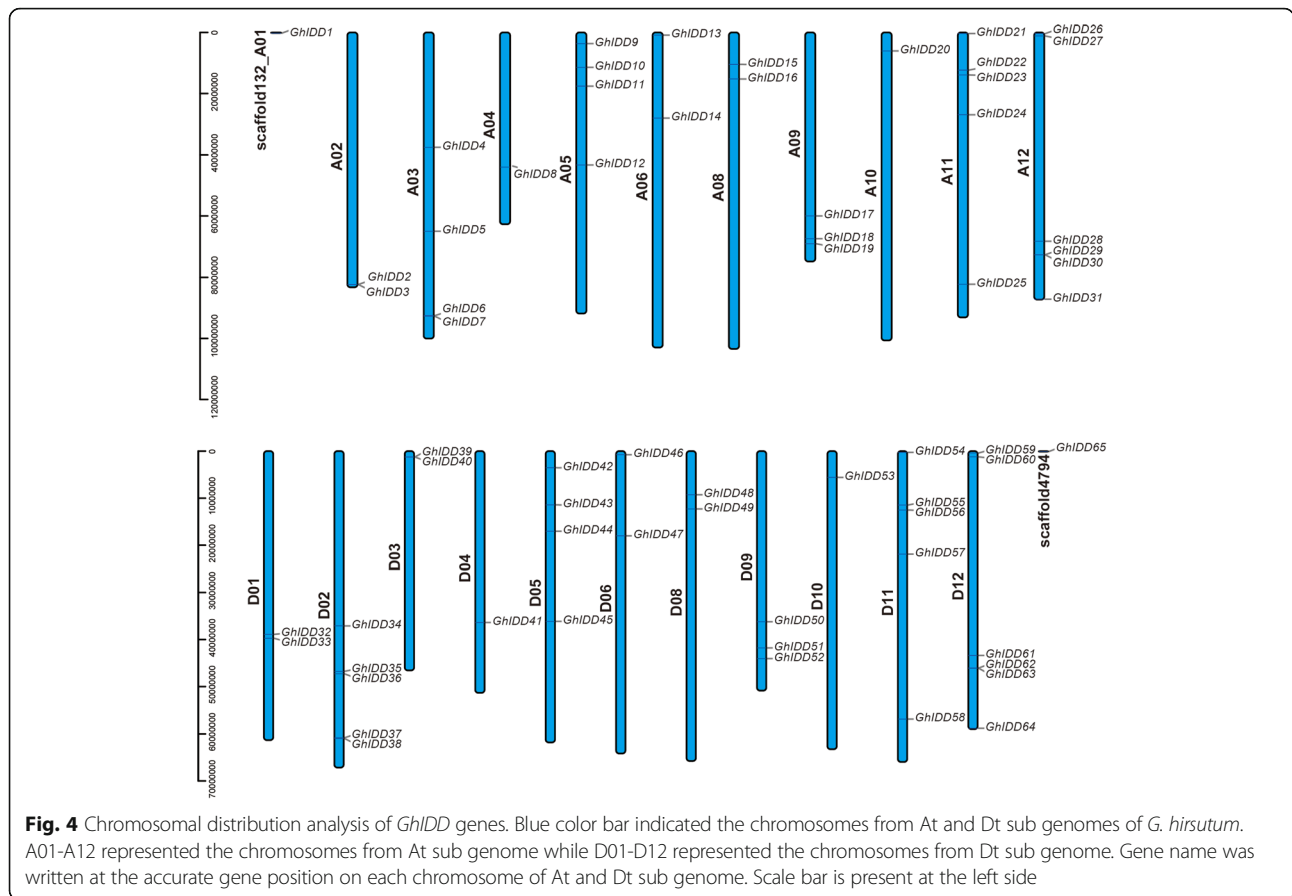## Chromosomal distribution, gene duplication and synteny analysis

The chromosomal distribution of *GhIDD* genes on their corresponding chromosomes (At and Dt sub-genome chromosomes of *G. hirsutum*) were employed. The 65 *GhIDD* genes were unevenly distributed on 21 chromosomes, including 30 genes on At sub-genome chromosomes, 33 genes on Dt sub-genome chromosomes while 2 genes were allotted on scaffolds (Fig. 4). The maximum numbers of genes (six genes on each) were found to be located on A12 and its orthologous chromosome D12 of At and Dt sub-genomes, respectively. We found that the distribution of genes was uneven within each chromosome, and most of the orthologues from the At and Dt sub-genomes were located on homologous

Fig. 3 *GhIDD* gene structure (exon–intron) and conserved motif analysis **a** An unrooted phylogenetic tree from *GhIDD* protein sequences constructed with MEGA using neighbor-joining method and conserved motifs analysis was done by MEME online program. Distribution of conserved motifs in *GhIDD* genes was presented by different colors. **b** Green lines and grey lines represented exon and intron positions, respectively, and the Scale bar is present at the bottom

chromosomes, however two orthologous genes were found on heterozygous chromosomes from the At and Dt sub-genomes. Four chromosomes out of 21 contained one *GhIDD* gene; seven chromosomes contained two genes; and two chromosomes contained three and six genes and three chromosomes contained four and five genes (Fig. 4). We did not found any gene on chromosome one and seven of At sub-genome as well as chromosome seven of Dt sub-genome, which showed that the gene duplications diversified from the diploid cotton species to the allotetroploid species, and these variety also result in the favorable economic characters in *G. hisutum*.

To study the locus relationship of orthologs/paralogous gene pairs between the At and Dt sub-genomes, we investigated the gene locus on chromosome and

performed synteny analysis. The synteny analysis revealed that most of the *IDD* loci were highly conserved between the At and Dt sub-genomes (Fig. 5). Tandem duplication, segmental duplication, and whole-genome duplication played an important role for gene family expansion (Xu et al. 2012). To understand the expansion of *GhIDD* gene family in cotton genome, we performed the gene duplication analysis within and between At and Dt sub-genomes of *G. hirsutum* (Additional file 1: Table S4). A total of 142 duplicated gene pairs were investigated, and among them 84 orthologous gene pairs were observed as a result of whole genome duplication, whereas 54 paralogous gene pairs contributed by segmental duplication and four duplicated gene pairs depicting tandem duplication event (two each sub-genome) were observed.

Fig. 4 Chromosomal distribution analysis of *GhIDD* genes. Blue color bar indicated the chromosomes from At and Dt sub genomes of *G. hirsutum*. A01-A12 represented the chromosomes from At sub genome while D01-D12 represented the chromosomes from Dt sub genome. Gene name was written at the accurate gene position on each chromosome of At and Dt sub genome. Scale bar is present at the left side

According to the Darwinian theory of natural selection, we investigated the non-synonymous divergence levels (*Ka*) versus synonymous divergence levels (*Ks*) for 142 duplicated gene pairs. It is found that 125 duplicated gene pairs showed *Ka/Ks* value less than 0.5, while 15 duplicated gene pairs *Ka/Ks* value was between 0.5 and 1 (Additional file 1: Table S4). However, only two duplicated gene pairs (*GhIDD15-GhIDD48* and *GhIDD23-- GhIDD56*) showed *Ka/Ks* value greater than 1. From above, the *Ka/Ks* values of most of duplicated gene pairs were less than 1 indicating that the upland cotton *IDD* gene family underwent a strong purifying selection pressure with limited functional divergence. That might be occurred after segmental and whole genome duplication (WGD) event during polyploidization followed by hybridization in the evolutionary history.
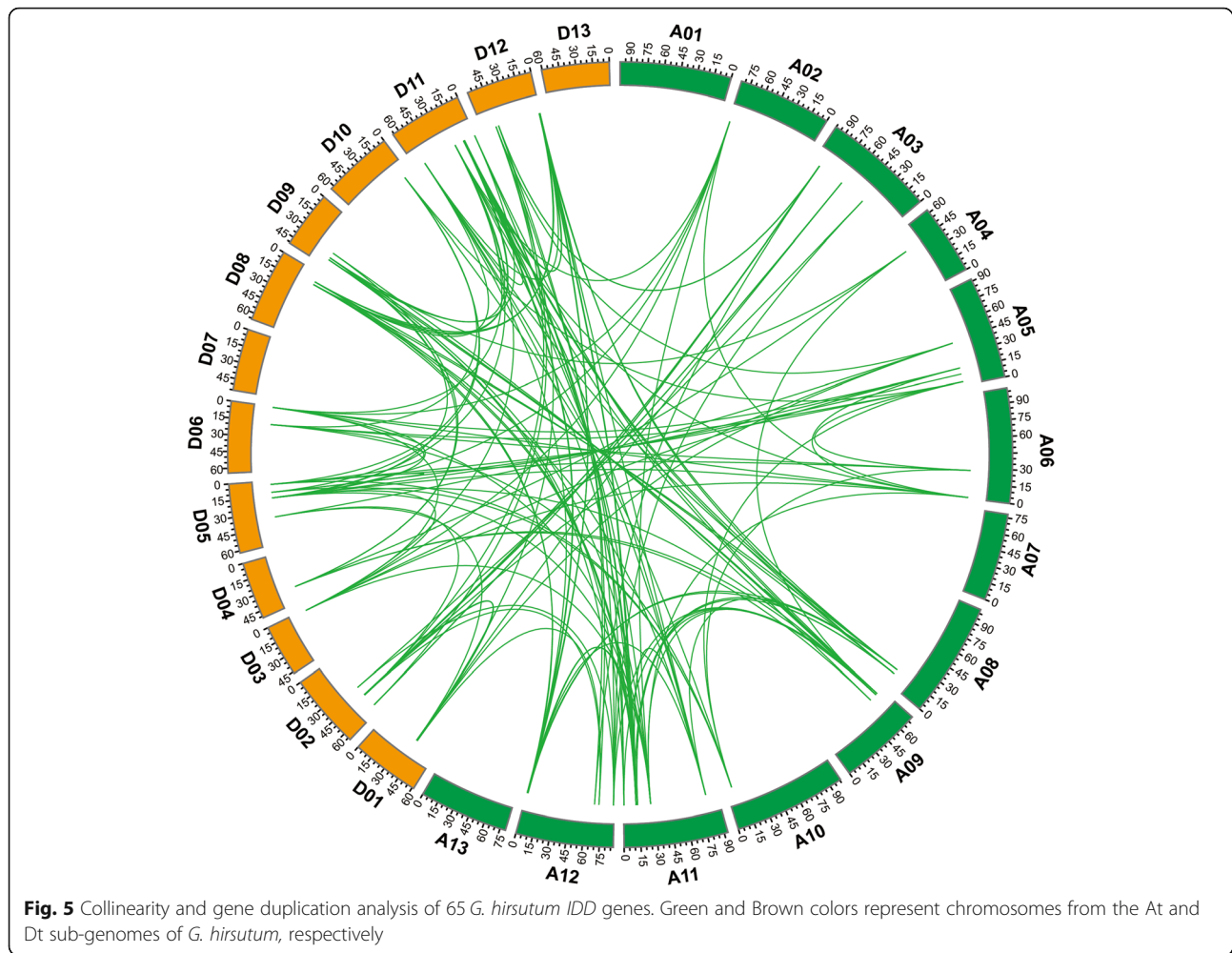
Conserved amino acid residues

IDD gene family is characterized by the presence of three conserved zinc finger C2H2 domains in their protein sequence. Protein sequence alignment of Arabidopsis, rice, and upland cotton (*G. hirsutum*) was performed to generate sequence logos of the zinc finger C2H2 domains, so as to investigate the homologous domain sequences and the conservation of each residue in the zinc finger C2H2 domains (Fig. 6). Results illustrated that conserved amino acid residues such as C [3], C [6], H [19], H [23], C [39], C [44], H [46], H [47], C [74], C [77], H [90] and H [116] were sequentially distributed throughout the conserved domain. However, among three C2H2 domains in this conserved region, two C2H2 domains occupied their positions in N terminal while one C2H2 domain was present in C terminal of that in all observed plant species, which showed the enrichment of C2H2 domain in N terminal of conserved domain across monocots and dicots plant species. From the results of conserved amino acid residues analysis, we deduced that the amino acid residues distribution in the *IDD* domain was highly conserved among dicot and monocot plant species. The results also indicated that most of the IDD proteins may bear similar biochemical function and target similar elements in the downstream gene regulation.

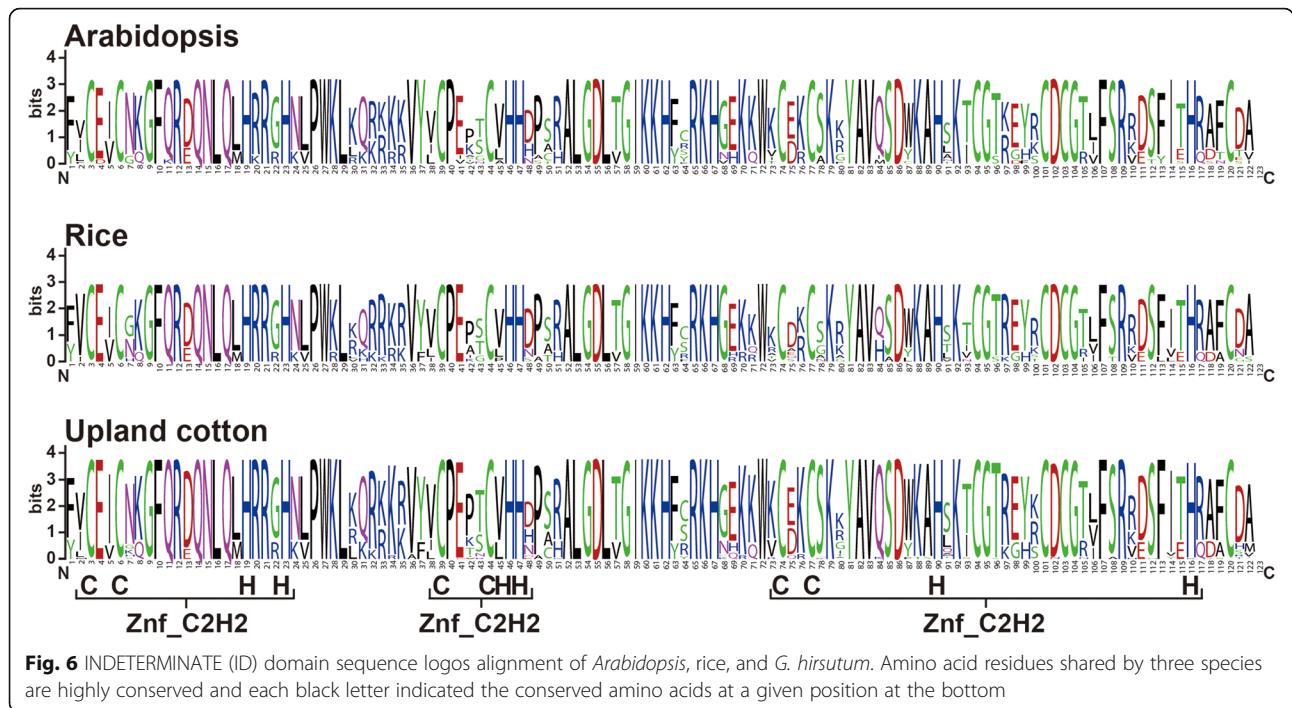### Spatial and temporal expression pattern of *GhIDD* genes

Plant *IDD* gene family has an important role in plant growth and development such as root development (Ingkasuwan et al. 2012; Yoshida et al. 2014), sugar and starch metabolism during flower transition in maize, rice and *Arabidopsis* (Colasanti et al. 2006; Ingkasuwan et al.

Fig. 5 Collinearity and gene duplication analysis of 65 *G. hirsutum IDD* genes. Green and Brown colors represent chromosomes from the At and Dt sub-genomes of *G. hirsutum*, respectively

2012; Matsubara et al. 2008; Park et al. 2008; Wong and Colasanti 2007; Wu et al. 2008). Spatiotemporal expression of transcript is tightly correlated with the biological function of a specific gene. To investigate the tissue specific expression patterns of different *IDD* genes, RNA-seq data were downloaded from NCBI to generate heat map. We noted that all the genes were clustered according to their expression patterns in the vegetative organs (root, stem, and leaf), reproductive organs (torus, petal, stamen, pistil, and calycle), ovule (− 3, − 1, 0, 1, 3, 5, 10, 20, 25 and 35 DPA) and fiber (5, 10, 20, and 25 DPA) (Additional file 3: Figure S2). Heat map displayed that most *GhIDD* genes showed a ubiquitous expression pattern in different observed tissues and minority showed much lower expression level. Only *GhIDD7* and *GhIDD38* displayed the specific expression in stamen.

Afterwards, qRT-PCR was performed by using root, stem, leaf, flower, 1, 3, 5, 7, 10, 15, and 20 DPA ovule tissues as well as 7, 10, 15, and 20 DPA fiber tissues to confirm the expression pattern obtained from the microarray data (Fig. 7). We selected the 12 segmentally

duplicated *GhIDD* genes (selecting one gene from each pair of segmentally duplicated genes) exhibiting higher expression pattern in different tissues and proceeded for qRT-PCR analysis. qRT-PCR results indicated that eight genes (*GhIDD2, GhIDD7, GhIDD9, GhIDD11, GhIDD15, GhIDD21, GhIDD39* and *GhIDD42*) represented clearly increased transcript levels in tissues of 7 DPA ovules indicating their potential roles in earlier development process of seed or fiber elongation. While *GhIDD4* and *GhIDD32* depicted peak transcript levels in stem tissues. *GhIDD48* showed significantly increased transcript level in flowers whereas *GhIDD33* had preferable expression in roots (Fig. 7). Moreover, four *GhIDD* genes (*GhIDD9, GhIDD15, GhIDD21,* and *GhIDD32*) displayed notable up-regulation only in vegetation and ovule tissues but not in fiber tissues, indicating that these *GhIDD* genes are important for vegetative as well as seed development. Overall, *GhIDD2* was distinctly expressed at almost all vegetative and reproductive organs, ovule and fiber tissues, suggested that *GhIDD2* may play multiple functions in growth and development. In a word, above results

Fig. 6 INDETERMINATE (ID) domain sequence logos alignment of *Arabidopsis*, rice, and *G. hirsutum*. Amino acid residues shared by three species are highly conserved and each black letter indicated the conserved amino acids at a given position at the bottom

revealed that the *GhIDD* genes in cotton have experienced functional deviation, because the segmentally duplicated genes showed different expression patterns in different tissues.

Responses of *GhIDD* genes under various abiotic stresses

Plant often faces the various stresses such as heat, cold, drought, and high salinity which influence the plant growth and productivity. These stresses induce or repress the expression of various genes 'effect on' genes functions related to plant growth and development. To investigate the responses of *GhIDD* genes under different abiotic stresses, RNA-seq data were downloaded from NCBI and a heat map depicting different responses was constructed (Additional file 4: Figure S3). RNA-seq data revealed that all the genes were clustered according to their different responses under specific abiotic stresses, which indicated the positive and negative regulating roles of *GhIDD* genes under different abiotic stresses.

To verify the results of RNA-seq data, qRT-PCR analysis was performed by treating the plants with different abiotic stresses such as cold, heat, salt (NaCl) and drought (PEG). We found that the *GhIDD15, GhIDD21, GhIDD32, GhIDD33, GhIDD42,* and *GhIDD48* were up-regulated in response to all stresses indicating that these genes might play an important positive role in abiotic stress response, while *GhIDD2* might play negative role with down-regulated under all abiotic stress treatments (Fig. 8). Further, the expression levels of *GhIDD4, GhIDD7, GhIDD11,* and *GhIDD21* were upregulated
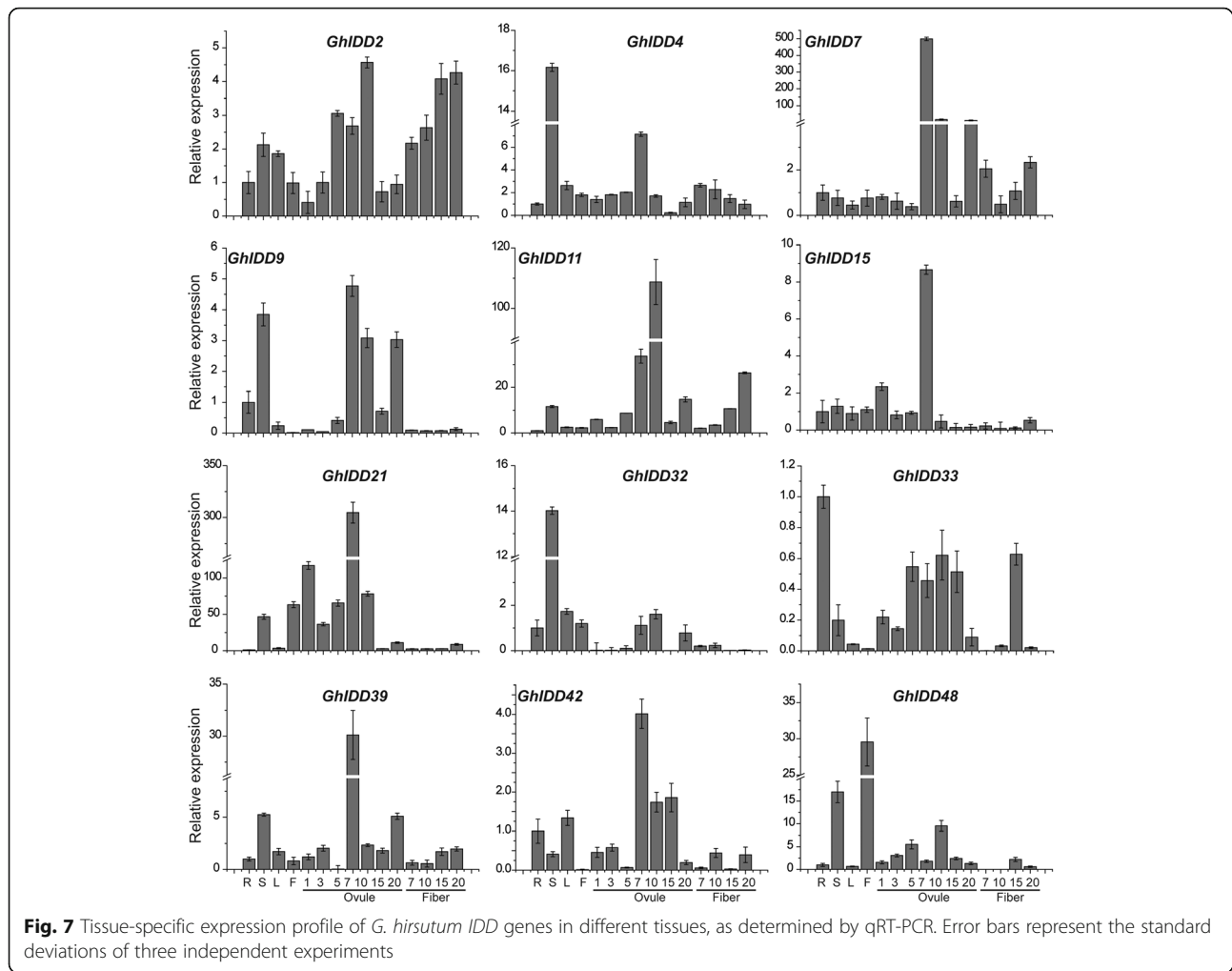
significantly in response to 6 h PEG treatment, however *GhIDD39* was clearly upregulated after 1 h treatment of PEG.

Discussion

The C2H2 transcription factors family, encoded by *IDD* genes, is one of the biggest plant gene families, and plays an important function in plant development and growth. In previous studies, identification of *IDD* gene family in rice, maize, *Arabidopsis*, and apple have been performed (Colasanti et al. 2006; Fan et al. 2017). But the genome-wide identification and analysis of *IDD* genes have not been performed in cotton till now. In present, a comprehensive identification and analysis of *IDD* genes in *G. hirsutum, G. arboreum, G. raimondii, T. cacao, A. thaliana, O. sativa, P. patens* (moss), and *S. moellendorffii* (fern) were performed. The *IDD* genes in allotetraploid cotton *G. hirsutum* were focused to understand the roles of *IDD* gene family in cotton development.

Phylogenetic analysis

A phylogenetic analysis was applied to determine the evolutionary relationship from eight species. No *IDD* gene family member was identified in algae indicating that the first *IDD* gene was originated in a moss, which is agreed with the result of a previous study (Wu et al. 2016). A total of 162 *IDD* genes were divided into seven different groups (IDD-A, IDD-B, IDD-C, IDD-D IDD-E, IDD-F, and IDD-G), which revealed that most

Fig. 7 Tissue-specific expression profile of *G. hirsutum IDD* genes in different tissues, as determined by qRT-PCR. Error bars represent the standard deviations of three independent experiments

of cotton *IDD* genes showed more close relationship with cacao *IDD* genes and predicted that cotton and cacao are evolved from common ancestors. Additionally, another phylogenetic tree was constructed from two diploid and an allotetraploid cotton species to confirm the evolutionary relationship among them. Phylogenetic tree among three cotton species divided *IDD* genes into four groups from IDD-a to d. Among these, group IDD-d had only 14 *GhIDD* genes that might be the result of introgression during the hybridization and polyploidization. Further, these results also strengthen the previous findings that *G. hirsutum* was evolved from the hybridization of A and D genomes cotton (*G. arboreum* and *G. raimondii,* respectively) as most of *IDD* genes from all three cotton species were closely distributed in phylogenetic tree (Li et al. 2014). To deeply understand the evolutionary history of *GhIDD* genes, another phylogenetic tree was constructed among *GhIDD* genes and 65 *GhIDD* genes were distributed into five groups. Consistent with our findings, group IDD-e contained the same 14 *GhIDD* genes as

previous phylogenetic analysis (Additional file 2: Figure S1). All these indicated that some *IDD* genes (14 genes in IDD-e of Fig. 2) are very ancient and important in the plant evolution and development, which may comprise the core gene resources in the plant.

Biophysical characteristics and chromosomal location
The prediction of the biophysical characteristics of all *GhIDD* gene family members provided valuable information to us too. Biophysical characteristics of all 65 *GhIDD* genes identified in *G. hirsutum* predicted that *GhIDD* genes were all located in nuclear. The values of isoelectric and grand average of hydropathicity (GRAVY) of the 65 *GhIDDs* suggested that all *IDD* proteins were alkaline and hydrophilic. These results were in accordance with previous genome wide study of *IDDs* in apple, which depicted alkaline and hydrophilic nature of all identified *IDD* genes with isoelectric point values more than 7 and grand average of hydropathicity (GRAVY) values less than 0 (Fan et al. 2017).
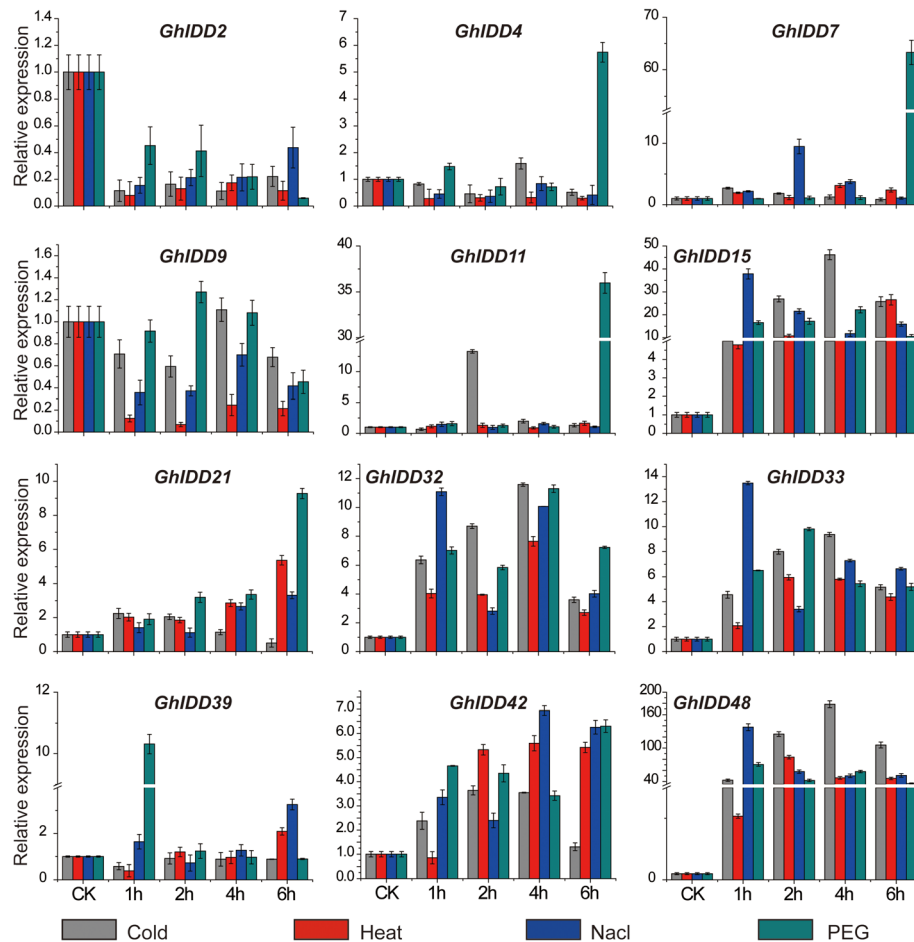
Fig. 8 Confirmation, the expression of the selected *GhIDD* genes in response to abiotic stresses using qRT-PCR. The mean expression values were calculated from three independent replicates. 0, 1, 2, 4, and 6 h indicate the hours after treatment. Mean values and standard errors were calculated from three replicates

Furthermore, the 65 identified *GhIDD* genes were distributed on 21 At and Dt sub-genome chromosomes of upland cotton, and didn't display obvious sub-genome bias. Where 30 *GhIDD* genes out of 65 were noticed to be located on 10 At chromosomes (A2, A3, A4, A5, A6, A8, A9, A10, A11, and A12) and 33 on 11 Dt chromosomes (D1, D2, D3, D4, D5, D6, D8, D9, D10, D11 and D12). The remaining two genes (*GhIDD1* and *GhIDD65*) were distributed on two unoriented scaffolds. The reason for uneven distribution of *GhIDD* genes on 21 chromosomes of At and Dt sub-genome of *G. hirsutum* is the addition or loss of genes during long evolutionary history of *G. hirsutum*.

Conserved amino acid residues, protein motifs and gene structure analysis

Furthermore, conserved amino acid residues analysis of *IDD* conserved domain from *O. sativa*, *A. thaliana*, and *G. hirsutum* revealed that the *IDD* domain was highly conserved in monocotyledons and dicotyledons during

the phase of evolution. In addition, a total of 10 motifs were identified which indicated that *IDD* proteins may function in divergent physiological pathways associated with different co-factors. Motifs distribution of *IDD* proteins suggested that *IDD* proteins motif distribution was relative conserved, and minor differences among the proteins from different groups might be associated with particular functions related to growth, development and stress tolerance in cotton. In detail, the motif 5 and 7 are conserved in the IDD-a, b, c and d subfamilies but no IDD-e subfamily, while the motif 6 and 10 only distributed in the proteins of IDD-e subfamily (Additional file 2: Figure S1), indicating that the gene evolution or duplication is correlated with the gene function variation.

Gene structure (exon–intron) is important that might be contributed by insertion/deletion events (Lecharny et al. 2003). Several genome-wide studies proved that the loss or gain of introns during eukaryotic diversification was extensive (Rogozin et al. 2003; Roy and Penny 2007). Gene structure analysis showed that duplicated

genes have similar gene structure, while intron length varies among genes indicating that intron length might play major roles in the functional diversification of *GhIDD* genes.

It is reported that introns play a vital role for the evolution of different plant species (Roy and Gilbert 2006). Here, we found that the number of introns varies from one to eight, however most genes showed two to three introns in their gene structure indicating that *G. hirsutum* is a newly evolved species with less number of introns, which supported previous study that large number of introns decreased over time during an early expansion stage (Roy and Penny 2007), and suggested that newly evolved species have less number of introns as compared with their primitive species (Roy and Gilbert 2006).

Gene duplication and selection pressure

We identified 65 *GhIDD* genes in the upland cotton genome, which were more in numbers than that previously identified in *Arabidopsis*, maize, rice, and apple. The main reason for larger number of *IDD* genes is that the upland cotton experienced polyploidization. Polyploidization was an important event for the evolution of cotton and contributed to gene duplication (Paterson et al. 2004). *G. hirsutum* is an allotetraploid cotton which is evolved as the result of hybridization of *G. arboreum* (A2 genome) and *G. raimondii* (D5 genome), and an important plant species for studying polyploidization (Wendel and Cronn 2003). The At and Dt sub-genome donors (*G. arboreum* and *G. raimondii,* respectively) of upland cotton are close relatives sharing the same number of orthologs, and resulted in duplication and doubling numbers of *GhIDD* genes in upland cotton. Accordingly, the numbers of *IDD* genes in *G. arboreum* and *G. raimondii* are 22 and 23, respectively, less than one half of *G. hirsutum.*

In previous studies, it is clear that gene duplication and diversification played an important role in the evolution. The gene duplications were always found in many plants and usually consisted of tandem, segmental, and whole genome duplications (Xu et al. 2012). Tandem duplication event occurred when two or more genes located on same chromosome, while segmental duplication event occurred between different chromosomes (He et al. 2012). Many transcription factor gene families including *AP2, WOX, YABBY, RH2FE3,* and *GRAS* genes underwent segmental duplication and attributed the gene family expansion and functional divergence in cotton (Liu and Zhang 2017; Qanmber et al. 2018; Yang et al. 2017; Yang et al. 2018; Zhang et al. 2018). In our study, 54 out of 142 duplicated gene pairs were associated with segmental duplication while four with tandem duplication contributed to the expansion of *GhIDDs* besides the diversification of *GhIDD* gene structure and function (Additional file 1: Table S4).

Many gene families have expanded too much higher numbers in plants than in other eukaryotes, suggesting that these expansions correlate with environmental pressure and selection pressure. To estimate the environmental pressures and selection pressure, non-synonymous (Ka) and synonymous substitution (Ks) rates of substitution (Ka/Ks) was calculated. Generally, $Ka/Ks > 1$, $Ka/Ks = 1$, and $Ka/Ks < 1$ indicate positive selection, neutral evolution, and purifying selection, respectively. In this study, we found that most Ka/Ks values of the *GhIDD* genes were smaller than 1.0 indicating that *GhIDD* gene family underwent a strong purifying selection pressure.

GhIDD genes expression in specific tissues and under different stresses

It has been reported that the *IDD* genes had essential functions in plant growth and development. The *AtIDD9* and *AtIDD10* interacted with DELLA which were used as scaffolds to mediate GA signaling pathways (Hassan et al. 2010). *AtIDD9* also plays an important role in epidermal cell fate specification in root (Long et al. 2015a; Long et al. 2015b). Moreover, It has been noted that *AtIDD10* acts upstream of root hair to regulate the accurate alternate pattern of N and H cells around cortex cells (Hassan et al. 2010). The phylogenetic analysis of *IDD* genes in apple revealed that *IDD* genes mediated flower induction (Fan et al. 2017). In rice, the *IDD* homolog LOOSE PLANT ARCHITECTURE1 (*LPA1/OsIDD16/IDD18*) also affects shoot response to gravity by modulating auxin flux in a brassinosteroid-dependent manner (Wu et al. 2013; Xuan et al. 2013).

Here, we analyzed the spatiotemporal expression of *GhIDD* genes in different tissues by Q-PCR. The results showed that most genes expressed peak in the ovule of 7 DPA, indicating their potential roles in the fiber elongation stage. *GhIDD2* may be as a constitutive regulator with its ubiquitous expression pattern. *GhIDD4, GhIDD32, GhIDD33,* and *GhIDD48* may play different roles in cotton vegetative and reproductive development with their distinct expression patterns. Thus, our results indicated that *GhIDD* genes demonstrated substantial functional diversity during cotton development and suggested that *GhIDD* genes are playing important function in seed or fiber development.

Previously, it has been reported that *IDD* genes functions were related to flower transition and epidermal cell development, but there is no report of *IDD* genes function under different abiotic stresses. Thus, to find whether they might play some roles in stress response, the responses of *GhIDD* genes under various abiotic stresses were determined. In our study, we found that the *GhIDD15, GhIDD21, GhIDD32, GhIDD33, GhIDD42,* and *GhIDD48*

expressions were upregulated under all treated abiotic stresses suggesting that these genes might play positive and important roles under the exposure of different abiotic stresses. In contrast, *GhIDD2* gene expression was down-regulated under all abiotic stresses indicating the negative response for abiotic stress. Further, *GhIDD4* and *GhIDD9* were down-regulated in response to heat and NaCl indicating that these genes might play negative role in response to heat and NaCl. However, *GhIDD11* is up-regulated in response to 2 h cold and 6 h PEG treatments. Whereas the expression level of *GhIDD4, GhIDD7, GhIDD11,* and *GhIDD21* were upregulated in response to PEG treatment at 6 h, indicating that these genes might play a role in a type of long-term dehydration tolerance and not as the instant sensors for abiotic stress signaling. In a word, most of the *IDD* genes were induced by different abiotic stresses, indicating that *GhIDD* genes might meditate the abiotic stress responses. Although *IDD* genes showed different expression levels under different stresses, there is no study on the function of cotton *IDDs* in stress. Therefore, there is need to investigate the functions of *IDD* genes under abiotic stresses in future studies. In short, our results showed that *GhIDD* genes may play an important role in plant vegetative development, seed and fiber development and might be proved important regulator in abiotic stresses tolerance of cotton.

Conclusions

IDD gene family plays significant role in plant growth and development. We identified 65 *IDD* genes in upland cotton genome that were deliberately investigated in gene phylogenetic evolution, gene structure variation, transcriptional expression pattern, prediction of protein motifs, subcellular localization and other characteristics. The phylogenetic analysis of *IDD* genes confirmed the close relationship of cotton and cacao, as the cotton and cacao were derived from the common ancestors. Collinearity analysis verified the expansion and evolution of *GhIDD* genes. Furthermore, the spatial and temporal expression patterns in different tissues revealed their diverse functions in cotton development along with their essential roles in ovule and fiber development. Most *GhIDD* genes transcript levels were high in 7 DPA ovule tissues indicating the potential pivotal roles in seed development and fiber elongation. Moreover, most of *IDD* gene family members showed positive responses under various tested abiotic stresses suggesting that *GhIDD* genes are involved in mediating abiotic stress response. Our study puts light on cotton *GhIDD* genes and provides basic information which will not only help to understand the evolutionary history of cotton *IDD* genes, but also be helpful to provide excellent candidate genes for genetic engineering to improve abiotic stress tolerance and fiber quality in cotton.

Additional files

Additional file 1: Table S1. Gene ID and name for *IDD* genes of *Arabidopsis,* rice, *G. hirsutum, G. arborium, G. raimondii, T. cacao, S. moellendorffii* and *P. patens.* Table S2. List of qPCR primers used in this study. Table S3. Biophysical properties of all *GhIDD* gene family members. Table S4. Orthologs/paralogs *IDD* gene pairs within and between At and Dt sub-genomes of *G. hirsutum* along with their types of duplication and *Ka/Ks* values. (XLSX 37 kb)

Additional file 2: Figure S1. Phylogenetic analysis of 65 *IDD* genes showing evolutionary relationships of *IDD* genes in *G. hirsutum.* Phylogenetic analysis divided the *GhIDD* genes into five groups differentiated with different colors. (TIF 19406 kb)

Additional file 3: Figure S2. Expression profiles of *GhIDD* genes from RNA-Seq data in various tissues of *G. hirsutum.* Gene expression levels are described with different colors on the scale. The log10-transformed FPKM values were used to construct the scale bars. Blue and red colors considered as low and high expression, respectively. (TIF 33463 kb)

Additional file 4: Figure S3. Analysis of Expression patterns of *GhIDD* genes under various stresses. Heat map of the expression level of *GhIDD* genes under different abiotic stresses including cold (a), heat (b), salt (c, 300 mmol·L^{-1} NaCl) and PEG600 (d) based on RNA-seq data. Scale bars represent log10 of the RPKM values. (TIF 33559 kb)

Abbreviations
BLAST: Basic Local Alignment Search Tool; DPA: Day(s) post-anthesis; IDD: INDETERMINATE DOMAIN; NaCl: Sodium chloride; PEG: Polyethylene glycol; qRT-PCR: Quantitative real time polymerase chain reaction; WGD: Whole genome duplication

Availability of data and materials
All data generated or analyzed in this study included in published article and additional files.

Authors' contributions
Li FG and Wang Z conceived and designed the study; Ali F, Qanmber G and Li YH carried out the experiments; Ma SY, Lu LL, Yang ZR analyzed and interpreted the data; Ali F and Wang Z prepared the manuscript. All the authors have read, edited, and approved the current version of the manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China. [2]Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 450001, China.

References

Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

Bailey TL, Williams N, Misleh C, et al. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006;34:W369–73.

Coelho CP, Huang P, Lee DY, et al. Making roots, shoots, and seeds: IDD gene family diversification in plants. Trends Plant Sci. 2018;23:66–78. https://doi.org/10.1016/j.tplants.2017.09.008.

Colasanti J, Tremblay R, Wong AY, et al. The maize *INDETERMINATE1* flowering time regulator defines a highly conserved zinc finger protein family in higher plants. BMC Genomics. 2006;7:158.

Colasanti J, Yuan Z, Sundaresan V. The *indeterminate* gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. Cell. 1998;93:593–603.

Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

Cui D, Zhao J, Jing Y, et al. The *Arabidopsis* IDD14, IDD15, and IDD16 cooperatively regulate lateral organ morphogenesis and gravitropism by promoting auxin biosynthesis and transport. PLoS Genet. 2013;9:e1003759. https://doi.org/10.1371/journal.pgen.1003759.

Englbrecht CC, Schoof H, Bohm S. Conservation, diversification and expansion of C2H2 zinc finger proteins in the *Arabidopsis thaliana* genome. BMC Genomics. 2004;5:39.

Eveland AL, Goldshmidt A, Pautler M, et al. Regulatory modules controlling maize inflorescence architecture. Genome Res. 2014;24:431–43. https://doi.org/10.1101/gr.166397.113.

Fan S, Zhang D, Xing L, et al. Phylogenetic analysis of *IDD* gene family and characterization of its expression in response to flower induction in *malus*. Mol Gen Genomics. 2017;292:755–71. https://doi.org/10.1007/s00438-017-1306-4.

Hassan H, Scheres B, Blilou I. JACKDAW controls epidermal patterning in the *Arabidopsis* root meristem through a non-cell-autonomous mechanism. Development. 2010;137:1523–9.

He HS, Dong Q, Shao YS, et al. Genome-wide survey and characterization of the gene family in *Populus trichocarpa*. Plant Cell Rep. 2012;31:1199–217. https://doi.org/10.1007/s00299-012-1241-0.

Hu B, Jin J, Guo AY, et al. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7. https://doi.org/10.1093/bioinformatics/btu817.

Ingkasuwan P, Netrphan S, Prasitwattanaseree S, et al. Inferring transcriptional gene regulation network of starch metabolism in *Arabidopsis thaliana* leaves using graphical Gaussian model. BMC Syst Biol. 2012;6:100. https://doi.org/10.1186/1752-0509-6-100.

Jeong EY, Seo PJ, Woo JC, et al. AKIN10 delays flowering by inactivating IDD8 transcription factor through protein phosphorylation in *Arabidopsis*. BMC Plant Biol. 2015;15:110. https://doi.org/10.1186/s12870-015-0503-8.

Jia J, Zhao P, Cheng L, et al. MADS-box family genes in sheepgrass and their involvement in abiotic stress responses. BMC Plant Biol. 2018;18:42. https://doi.org/10.1186/s12870-018-1259-8.

Jones P, Binns D, Chang HY, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40. https://doi.org/10.1093/bioinformatics/btu031

Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4. https://doi.org/10.1093/molbev/msw054.

Lecharny A, Boudet N, Gy I, et al. Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure. In: Meyer A, Van de Peer Y, eds. Genome Evolution: Gene and Genome Duplications and the Origin of Novel Gene Functions. Dordrecht: Springer Netherlands. 2003;111–6.

Lee MS, Gippert GP, Soman KV, et al. Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science. 1989;245:635–7.

Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43:D257–60. https://doi.org/10.1093/nar/gku949.

Li FG, Fan GY, Wang KB, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat Genet. 2014;46:567–72. https://doi.org/10.1038/ng.2987.

Li J, Yu DQ, Qanmber G, et al. *GhKLCR1*, a kinesin light chain-related gene, induces drought-stress sensitivity in *Arabidopsis*. Sci China Life Sci. 2019;62:63. https://doi.org/10.1007/s11427-018-9307-y.

Liu CX, Zhang TZ. Expansion and stress responses of the AP2/EREBP superfamily in cotton. BMC Genomics. 2017;18:118. https://doi.org/10.1186/s12864-017-3517-9.

Long Y, Goedhart J, Schneijderberg M, et al. SCARECROW-LIKE23 and SCARECROW jointly specify endodermal cell fate but distinctly control SHORT-ROOT movement. Plant J. 2015a;84:773–84. https://doi.org/10.1111/tpj.13038.

Long Y, Smet W, Cruz-Ramirez A, et al. *Arabidopsis* BIRD zinc finger proteins jointly stabilize tissue boundaries by confining the cell fate regulator SHORT-ROOT and contributing to fate specification. Plant Cell. 2015b;27:1185–99. https://doi.org/10.1105/tpc.114.132407.

Matsubara K, Yamanouchi U, Wang ZX, et al. *Ehd2*, a rice ortholog of the maize *INDETERMINATE1* gene, promotes flowering by up-regulating *Ehd1*. Plant Physiol. 2008;148:1425–35.

Moreno-Risueno MA, Sozzani R, Yardimci GG, et al. Transcriptional control of tissue formation throughout root development. Science. 2015;350:426–30. https://doi.org/10.1126/science.aad1171.

Pabo CO, Sauer RT. Transcription factors: structural families and principles of DNA recognition. Annu Rev Biochem. 1992;61:1053–95.

Park SJ, Kim SL, Lee S, et al. Rice *Indeterminate 1* (*OsId1*) is necessary for the expression of *Ehd1* (*early heading date 1*) regardless of photoperiod. Plant J. 2008;56:1018–29.

Parraga G, Horvath SJ, Eisen A, et al. Zinc-dependent structure of a single-finger domain of yeast ADR1. Science. 1988;241:1489–92.

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;101:9903–8.

Qanmber G, YU DQ, LI J, et al. Genome-wide identification and expression analysis of *Gossypium* RING-H2 finger E3 ligase genes revealed their roles in fiber development, and phytohormone and abiotic stress responses. Journal of Cotton Research. 2018;1:1. https://doi.org/10.1186/s42397-018-0004-z.

Riddick EW, Simmons AM. Do plant trichomes cause more harm than good to predatory insects? Pest Manag Sci. 2014;70:1655–65. https://doi.org/10.1002/ps.3772.

Riechmann JL, Heard J, Martin G, et al. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science. 2000;290:2105–10.

Rinehart JA, Petersen MW, John ME. Tissue-specific and developmental regulation of cotton gene *FbL2A*. Demonstration of promoter activity in transgenic plants. Plant Physiol. 1996;112:1331–41.

Rogozin IB, Wolf YI, Sorokin AV, et al. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol. 2003;13:1512–7.

Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet. 2006;7:211–21.

Roy SW, Penny D. Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of *O. sativa* and *A. thaliana*. Mol Biol Evol. 2007;24:171–81.

Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. Bioinformatics. 2002;18:207–8.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12.

Takatsuji H. Zinc-finger transcription factors in plants. Cell Mol Life Sci. 1998;54:582–96.

Tanimoto M, Tremblay R, Colasanti J. Altered gravitropic response, amyloplast sedimentation and circumnutation in the *Arabidopsis shoot gravitropism 5* mutant are associated with reduced starch levels. Plant Mol Biol. 2008;67:57–69.

Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc. 2012;7:562–78. https://doi.org/10.1038/nprot.2012.016.

Wan Q, Guan XY, Yang NN, et al. Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. New Phytol. 2016;210:1298–310. https://doi.org/10.1111/nph.13860.

Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. Adv Agron. 2003;78:139–86.

Wong AY, Colasanti J. Maize floral regulator protein INDETERMINATE1 is localized to developing leaves and is not altered by light or the sink/source transition. J Exp Bot. 2007;58:403–14. https://doi.org/10.1093/jxb/erl206.

Wu C, You C, Li C, et al. *RID1*, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. Proc Natl Acad Sci U S A. 2008;105:12915–20.

Wu P, Song XM, Wang Z, et al. Genome-wide analysis of the BES1 transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Plant Growth Regul. 2016;80:291–301. https://doi.org/10.1007/s10725-016-0166-y.

Wu XR, Tang D, Li M, et al. Loose Plant Architecture1, an INDETERMINATE DOMAIN protein involved in shoot gravitropism, regulates plant architecture in rice. Plant Physiol. 2013;161:317–29. https://doi.org/10.1104/pp.112.208496.

Xu GX, Guo C, Shan HY, et al. Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci U S A. 2012;109:1187–92. https://doi.org/10.1073/pnas.1109047109.

Xuan YH, Priatama RA, Huang J, et al. Indeterminate domain 10 regulates ammonium-mediated gene expression in rice roots. New Phytol. 2013;197:791–804. https://doi.org/10.1111/nph.12075.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

Yang ZE, Gong Q, Qin WQ, et al. Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. BMC Plant Biol. 2017;17:113. https://doi.org/10.1186/s12870-017-1065-8.

Yang ZE, Gong Q, Wang LL, et al. Genome-wide study of YABBY genes in upland cotton and their expression patterns under different stresses. Front Genet. 2018;9:33. https://doi.org/10.3389/fgene.2018.00033.

Yang ZR, Zhang CJ, Yang XJ, et al. *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytol. 2014;203:437–48. https://doi.org/10.1111/nph.12824.

Yoshida H, Hirano K, Sato T, et al. DELLA protein functions as a transcriptional activator through the DNA binding of the indeterminate domain family proteins. Proc Natl Acad Sci U S A. PNAS. 2014;111(21):7861–7866. https://doi.org/10.1073/pnas.1321669111.

Zhang B, Liu J, Yang ZE, et al. Genome-wide analysis of GRAS transcription factor gene family in *Gossypium hirsutum* L. BMC Genomics. 2018;19:348. https://doi.org/10.1186/s12864-018-4722-x.