SCIENCE CHINA
Life Sciences

THEMATIC ISSUE: Cotton genome plus
• RESEARCH PAPER •

February 2016 Vol.59 No.2: 130–141
doi: 10.1007/s11427-015-4916-2

Genome-wide analysis and expression profiling of the phospholipase D gene family in *Gossypium arboreum*

Kai TANG [1], Chunjuan DONG [2] & Jinyuan LIU [1*]

[1] *Laboratory of Plant Molecular Biology, Center for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China;*
[2] *Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

Received May 27, 2015; accepted June 16, 2015; published online December 29, 2015

The plant phospholipase D (PLD) plays versatile functions in multiple aspects of plant growth, development, and stress responses. However, until now, our knowledge concerning the PLD gene family members and their expression patterns in cotton has been limited. In this study, we performed for the first time the genome-wide analysis and expression profiling of PLD gene family in *Gossypium arboretum*, and finally, a total of 19 non-redundant *PLD* genes (*GaPLDs*) were identified. Based on the phylogenetic analysis, they were divided into six well-supported clades (α, β/γ, δ, ε, ζ and φ). Most of the *GaPLD* genes within the same clade showed the similar exon-intron organization and highly conserved motif structures. Additionally, the chromosomal distribution pattern revealed that *GaPLD* genes were unevenly distributed across 10 of the 13 cotton chromosomes. Segmental duplication is the major contributor to the expansion of *GaPLD* gene family and estimated to have occurred from 19.61 to 20.44 million years ago when a recent large-scale genome duplication occurred in cotton. Moreover, the expression profiling provides the functional divergence of *GaPLD* genes in cotton and provides some new light on the molecular mechanisms of *GaPLDα1* and *GaPLDδ2* in fiber development.

***Gossypium arboreum*, phospholipase D, gene family, expression patterns, fiber development**

Citation: Tang, K., Dong, C., Liu, J. (2016). Genome-wide analysis and expression profiling of the phospholipase D gene family in *Gossypium arboreum*. Sci China Life Sci 59, 130−141, doi: 10.1007/s11427-015-4916-2

INTRODUCTION

In recent years, phospholipids have gradually been recognized as the rich sources for generating mediators and the active participants in cell regulation. Phospholipase D (PLD, EC 3.1.4.4) constitutes a major phospholipase gene family that hydrolyzes the terminal phosphodiester bond of phospholipids to generate the phosphatidic acid (PA) and a free head group (Jacobs et al., 2010; Mendonsa and Engebrecht, 2009; Peng and Frohman, 2012; Wang et al., 2012). In plant, PLD gene family contains various types with distinguishable biochemical, structural and catalytic characteristics (Chen et al., 2011). It is now known that PLD is a ubiquitous enzyme participating in the regulation of numerous physiological processes, including plant growth, development, and multiple stress responses, such as drought, salinity, nutrient deficiency, freezing, wounding and pathogen infection (Hong et al., 2010; Wang, 2005). The diverse functions of PLD are often carried out through its enzymatic product, PA, which is regarded as a universal lipid signaling molecule (Pleskot et al., 2013; Wang et al., 2006).

PLD was identified as an enzyme involved in lipid metabolism and membrane reconstruction as early as in 1940s (Hanahan and Chaikoff, 1947). However, until 1994, the first PLD encoding cDNA was isolated from castor bean (Wang et al., 1994). And since then, numerous PLD genes were cloned from other plants, including rice (Li et al.,

*Corresponding author (email: liujy@mail.tsinghua.edu.cn)

2007; Ueki et al., 1995), *Arabidopsis* (Hong et al., 2009; Hong et al., 2008; Qin et al., 1997), maize (Qin et al., 1997), cabbage (Kim et al., 1999), tobacco (Lein and Saalbach, 2001), tomato (Whitaker et al., 2001), longan (You et al., 2014). Recently, the genomic resources facilitate us to characterize the plant PLD gene families at a genome-wide level. Up to date, the PLD gene families have been successively identified in *Arabidopsis* (Elias et al., 2002), rice (Li et al., 2007), poplar (Liu et al., 2010), grape (Liu et al., 2010), soybean (Zhao et al., 2012), apple (Du et al., 2013), strawberry (Du et al., 2013) and Chinese plum (Du et al., 2013). All these plants contain more than ten *PLD* genes. According to the sequence characteristics, the PLD gene family can be generally classified into six or seven types, including PLDα, PLDβ, PLDγ, PLDδ, PLDε, PLDζ and/or PLDφ (Elias et al., 2002; Li et al., 2007). The diverse PLD types differentiated depending on their requirements and affinities for Ca^{2+}, phosphatidylinositol 4,5-bisphosphate (PIP_2) and free fatty acids (Qin and Wang, 2002; Qin et al., 1997). The exception was PLDφ, which was poorly characterized as of yet (Li et al., 2007).

All of the PLD proteins are featured by two highly conserved catalytic motifs called HKD domain, which is named after three invariant residues H, K and D in "HxKxxxxD" motif. Two HKD domains are far from each other in the primary structure, but they interact with each other to form the active site which promotes the lipase activity (Qin and Wang, 2002). Based on the domains nearby the N-terminus, PLD proteins can be divided into three subfamilies, C2-PLD, PX/PH-PLD and SP-PLD. The C2-PLD subfamily proteins contain a Ca^{2+}-dependent phospholipid-binding protein kinase C-conserved 2 domain (C2 domain) and require Ca^{2+} for their enzymatic activity (Wang et al., 2000). The PX/PH-PLDs contain a phox homology (PX) domain and a pleckstrin homology (PH) domain at the N-terminus and are independent of Ca^{2+} for activity (Wang et al., 2000). In the third subfamily, SP-PLD, only a signal peptide (SP) can be detected (Li et al., 2007). PLDα, PLDβ, PLDγ, PLDδ, and PLDε belong to C2-PLD, PLDζ belongs to PX/PH-PLD and PLDφ belongs to SP-PLD.

Cotton is an important crop worldwide. It is a major resource of natural fiber in the textile industry, and also can provide a significant amount of oilseed for food and biofuel (Cao, 2015; Pei, 2015). In the previous proteomic studies, PLDs were proposed to be involved in cotton fiber development. The PLD proteins began to accumulate during the rapid fiber elongation, and reached up to a peak simultaneously with the initiation of second cell wall synthesis (Hu et al., 2013; Yang et al., 2008b). Moreover, consistent with the PLD protein expression profile, the PA content was low in the elongating fibers and high in the mature ones (Wanjie et al., 2005). Additionally, PLD is also known to function during the recovery from freezing temperatures (Kargiotidou et al., 2010). However, up to now, no genome-wide characterization of PLD gene family has been performed in cotton. The availability of the genome sequence of *Gossypium arboreum*, a diploid cotton species, provides us with a great opportunity to identify the PLD gene family in cotton genome (Li et al., 2014).

In the present study, we performed for the first time the comprehensive analysis of PLD gene family in *G. arboreum*. A total of 19 non-redundant *PLD* encoding genes (*GaPLD*s) were identified and subjected to a systematic analysis, including their chromosomal location, gene structure, protein domain, phylogenetic relationship, gene duplication events, and expression profiling. It was remarkable that the expansion of *GaPLD* gene family might be caused mainly by gene segmental duplications but not associated with tandem duplication events. Also, based on their expression profiles, *GaPLDα1* and *GaPLDδ2* might functionally relate with fiber development. In a word, our genome-wide analysis provides a comprehensive characterization of *PLDs* in *G. arboreum* and further elucidates the potential function of the *PLD* genes in fiber development.

RESULTS AND DISCUSSION

PLD gene family in *G. arboreum*

To identify the *PLD* genes, the HHM profile of HKD domain (PF00614) and corresponding *PLD* gene sequences from *Arabidopsis* and rice were employed as query to perform multiple searches of the *G. arboreum* genome database using the blastp and tblastn algorithms and HMMER search (Altschul et al., 1997; Eddy, 2009; Li et al., 2014). Subsequently, in order to verify the reliability of the initially identified results, a survey was conducted to confirm the existence of two conserved HKD domains with the InterproScan program (Quevillon et al., 2005). Following this strategy, we identified 19 non-redundant *PLD* genes (*GaPLD*s). According to the high sequence identity with *Arabidopsis* and rice *PLD*s, *GaPLD*s were classified into seven types of *PLD*s, including four *PLDα*s (*GaPLDα1~GaPLDα4*), two *PLDβ*s (*GaPLDβ1 and GaPLDβ2*), one *PLDγ* (*GaPLDγ*), five *PLDδ*s (*GaPLDδ1~GaPLDδ5*), one *PLDε* (*GaPLDε*), four *PLDζ*s (*GaPLDζ1~GaPLDζ4*) and two *PLDφ*s (*GaPLDφ1* and *GaPLDφ2*) (Tables 1 and S1).

The members within the individual type of *GaPLD*s showed high sequence identities in both nucleotide and protein levels (Tables S2 and S3). The lengths of the deduced GaPLD proteins varied from 494 to 1,142 amino acids (aa) with putative molecular weights (MW) ranging from 55.71 to 129.80 kD and isoelectric points (pI) ranging

from 5.50 to 8.64 (Table 1). In addition, 12 of the 19 GaPLD proteins were predicted to be located in the cytoplasm, four GaPLDs were predicted to reside in the nucleus, and the others were mainly in the plasma membrane and lysosome (Table 1). The cotton PLD gene family had largest size and more isostatic type distribution in comparison with the plants reported (Table S4). Thus, it made the cotton PLD gene family as the good model to shed light on the molecular heterogeneity of plant PLD gene families.

Chromosomal distribution and gene duplication

Mapping the *GaPLD* genes to the *G. arboreum* genome, 18 of the 19 *GaPLD* genes were dispersed on 10 chromosomes, and the remaining one (*GaPLDα1*) was anchored on an unmapped scaffold (Table 1, Figure 1). On each chromosome, the numbers of *GaPLD* genes range from one to three. The chromosomes 2, 6, 7, 8 and 12 contained one *GaPLD* gene, followed by two on Chr9 and Chr11, and three on Chr1, Chr3 and Chr4 (Table 1, Figure 1).

Table 1 The *PLD* genes in *G. arboreum* and properties of the deduced proteins[a)]

Gene name	Gene ID	Chromosomal location ChrNo.(Orientation): start-end*	ORF (bp)	Protein**			
				Size (aa)	MW (kD)	pI	Subcellular localization***
GaPLDα1	Cotton_A_40825	Scaffold5309(+): 1363-4130	2,388	795	90.41	5.53	C
GaPLDα2	Cotton_A_28309	Ghr11(+): 55908865-55912228	2,418	805	91.56	5.50	C
GaPLDα3	Cotton_A_07260	Chr1(–): 48369550-48365993	2,415	804	91.50	6.41	C
GaPLDα4	Cotton_A_01221	Chr3(+): 33063206-33066188	2,454	817	92.77	6.71	C
GaPLDβ1	Cotton_A_10944	Chr3(–): 104157709-104151762	3,300	1,099	122.92	6.98	N
GaPLDβ2	Cotton_A_03378	Chr2(+): 24758836-24763297	3,168	1,055	119.89	7.62	N
GaPLDγ	Cotton_A_09962	Chr7(–): 14736009-14732463	2,547	848	95.51	8.31	C
GaPLDδ1	Cotton_A_11389	Chr9(+): 77533497-77538982	2,745	914	104.09	7.03	C
GaPLDδ2	Cotton_A_01522	Chr1(–): 134896399-134891836	2,571	856	97.05	6.79	C
GaPLDδ3	Cotton_A_28560	Chr4(+): 59372475-59376722	2,538	845	95.68	6.64	C
GaPLDδ4	Cotton_A_17331	Chr6(+): 8745800-8749613	2,523	840	95.69	8.64	PM
GaPLDδ5	Cotton_A_10805	Chr9(–): 66245264-66240400	2,547	848	96.05	6.80	C
GaPLDε	Cotton_A_28518	Chr12(+): 41803809-41806638	2,307	768	87.82	6.59	C
GaPLDζ1	Cotton_A_17807	Chr4(+): 107626173-107633930	3,321	1,106	126.00	6.16	C
GaPLDζ2	Cotton_A_19886	Chr8(–): 36407783-36399285	3,429	1,142	129.80	6.19	N
GaPLDζ3	Cotton_A_04839	Chr4(+): 42934890-42943049	3,252	1,083	123.00	6.12	N
GaPLDζ4	Cotton_A_10711	Chr3(+): 72779783-72785341	3,318	1,105	126.00	5.57	C
GaPLDφ1	Cotton_A_19212	Chr1(+): 49253802-49256135	1,545	514	57.95	7.63	L
GaPLDφ2	Cotton_A_32589	Chr11(+): 9553235-9555567	1,485	494	55.71	8.11	PM

a) *: ChrNo.(Orientation): start-end: "+" and "–" indicated the forward and reverse orientation, respectively; **: The theoretical molecular weight (MW) and isoelectric point (pI) were calculated by ExPASy (http://cn.expasy.org/tools). Subcellular localization was analyzed using the CELLO v2.5 server (http://cello.life.nctu.edu.tw/). The Nucleic acid and deduced amino acid sequences of each *GaPLD* gene are listed in Table S1; ***: C, cytoplasm; N, nucleus; PM, plasma membrane; L, lysosom.

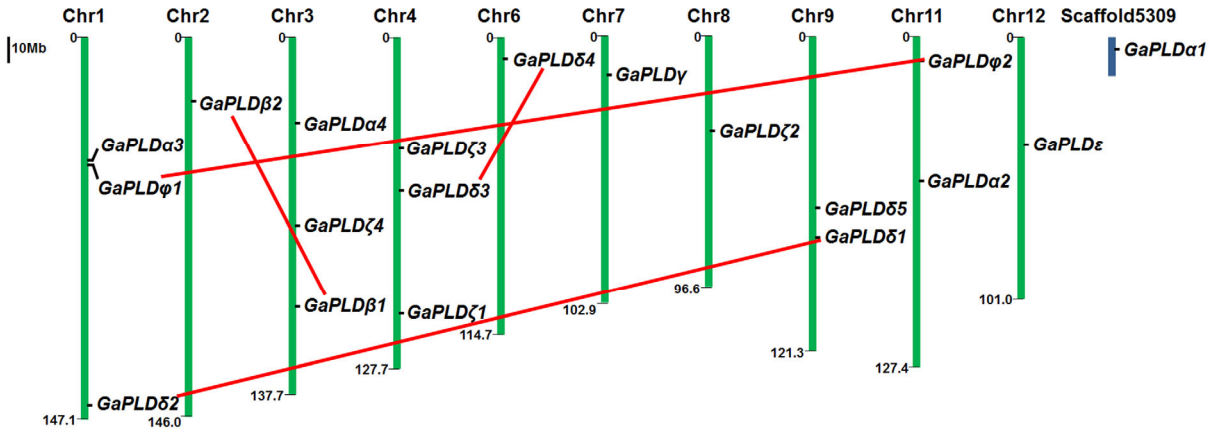

Figure 1 The chromosome distribution and synteny of *GaPLD* genes. Scale represents a 10 Mb chromosomal distance, but the Scaffold5309 does not follow the scale. The duplicated genes are connected with red lines.

To determine the roles of gene duplication in the expansion of *GaPLD* gene family, potential duplication events were analyzed. We firstly selected seven putative paralogous gene pairs (*GaPLDα1-GaPLDα2, GaPLDα3-GaPLDα4, GaPLDβ1-GaPLDβ2, GaPLDδ1-GaPLDδ2, GaPLDδ3-GaPLDδ4, GaPLDζ1-GaPLDζ2, GaPLDφ1-GaPLDφ2*) with high degree of protein sequence identities and subsequently explored the degree to whether their flanking genes were conserved. The results showed that four pairs of *GaPLD* genes were identified (*GaPLDβ1-GaPLDβ2, GaPLDδ1-GaPLDδ2, GaPLDδ3-GaPLDδ4, GaPLDφ1-GaPLDφ2*) (Figure 1), accounting for about 42% of the entire *GaPLD* gene family and thereby supporting the hypothesis that putative gene duplication events were the main causes of the expansion of the *GaPLD* gene family. All four gene pairs appeared randomly scattered throughout the genomes, suggesting that these eight genes originate from four segmental duplication events. During *GaPLD* gene family expansion, no significant tandem duplication event was detected, and this was quite different from the PLD gene families in *Arabidopsis* and rice (Elias et al., 2002; Li et al., 2007). For the gene pair *GaPLDα1-GaPLDα2*, no traceable duplication events could be determined, partially for the reason that *GaPLDα1* was anchored on unmapped scaffolds (Table 1).

It is generally assumed that the level of synonymous substitutions (*Ks*) between two homologous genes increases approximately linearly with time (Blanc and Wolfe, 2004). Thus, we can estimate the evolutionary dates of the segmental duplication events by *Ks* calculation. The protein-coding genes flanking the four pairs of segmental duplicated genes had consistent mean *Ks* values (from 0.59 to 0.61), suggesting that the segmental duplication events occurred from 19.61 to 20.44 million years ago (mya) in *G. arboreum* (Table 2). This time period was later than the division between cotton and *Arabidopsis*, circa 83~86 mya, and was consistent with the time (20~40 mya) when a recent large-scale genome duplication event occurred in cotton. The doubling of gene contents in newly formed polyploids is often thought to relax selection on individual genes, thus creating opportunities for novel gene evolution and expression patterns (Adams et al., 2003; Desai et al., 2006).

Phylogenetic relationships

To get a better understanding of the evolutionary history and phylogenetic relationships of GaPLD gene family, an NJ phylogenetic tree was constructed. The PLD protein sequences from *G. arboreum, Arabidopsis* and rice were compared with the sequences of cacao, which is a close relative of cotton in the *Malvaceae* family and whose genome has been sequenced (Figure 2) (Argout et al., 2011). The bootstrap values for some nodes of the tree were low as a result of relatively large number of sequences. The phylogenetic trees reconstructed with Maximum likelihood and Minimal Evolution methods, were almost identical with only minor differences at some branches, suggesting that the three methods were highly consistent with each other.

According to the NJ phylogenetic tree, 19 GaPLDs formed six well-supported clades (α, β/γ, δ, ε, ζ and φ) (Figure 2). The interspersed PLDs distribution indicated that different PLD types had existed before the divergence of the evolutionary lineages of these plant sepecies. PLDs from each of the clades were found in all four higher plants examined with the exception of members of the two small clades, ε and φ, which were absent in the rice and *Arabidopsis* genomes, respectively (Figure 2). Among these clades, the α type constituted the largest clade containing 19 members, and the second largest one, β/γ-clade, was the only bi-type clade comprised of seven PLDβs and five PLDγs, while in the smallest one, ε-clade, only three PLDεs were included. As expected, in each clade with all four plants, dicot PLDs (cotton, cacao and *Arabidopsis* PLDs) were more closely related to each other than to monocot PLDs (rice PLDs), and *Malvaceae* PLDs (cotton and cacao PLDs) distributed more similarly than the others (*Arabidopsis* and rice PLDs). Notablely, in β/γ- and δ-clades, *TcPLDβ*, *TcPLDδ1* and *TcPLDδ2* individually had two counterparts in cotton, matching that *GaPLD* gene segmental duplication occurred later than the split of cotton and cacao in the *Malvaceae* family (Paterson et al., 2012). Additionally, ζ-clade and φ-clade were far from the other four clades in evolutionary distance, suggesting that their sequence characteristics might be quite different from the others.

Exon-intron organization and domain architecture of the GaPLD gene family

To gain the further insights into the evolutionary relationships among *GaPLD* genes, we investigated the gene structures of individual *GaPLD* genes. As shown in Figure 3A,

Table 2 Duplicated *GaPLD* genes and the number of conserved protein-coding genes flanking them[a)]

Duplicated *PLD* gene 1	Duplicated *PLD* gene 2	Number of flanking protein-coding genes	Mean *Ks*	SD *Ks*	Mini *Ks*	Max *Ks*	Date (mya)
GaPLDβ1	*GaPLDβ2*	5	0.589,2	0.026,2	0.556,0	0.619,6	19.64
GaPLDδ1	*GaPLDδ2*	2	0.588,4	0.014,7	0.578,0	0.598,8	19.61
GaPLDδ3	*GaPLDδ4*	2	0.613,2	0.019,7	0.599,3	0.627,2	20.44
GaPLDφ1	*GaPLDφ2*	3	0.610,7	0.184,0	0.484,7	0.821,9	20.35

a) *Ks*: synonymous substitution rates; SD *Ks*: standard deviation *Ks*; Mini *Ks*: minimum *Ks*; Max *Ks*: maximum *Ks*; mya: million years ago.

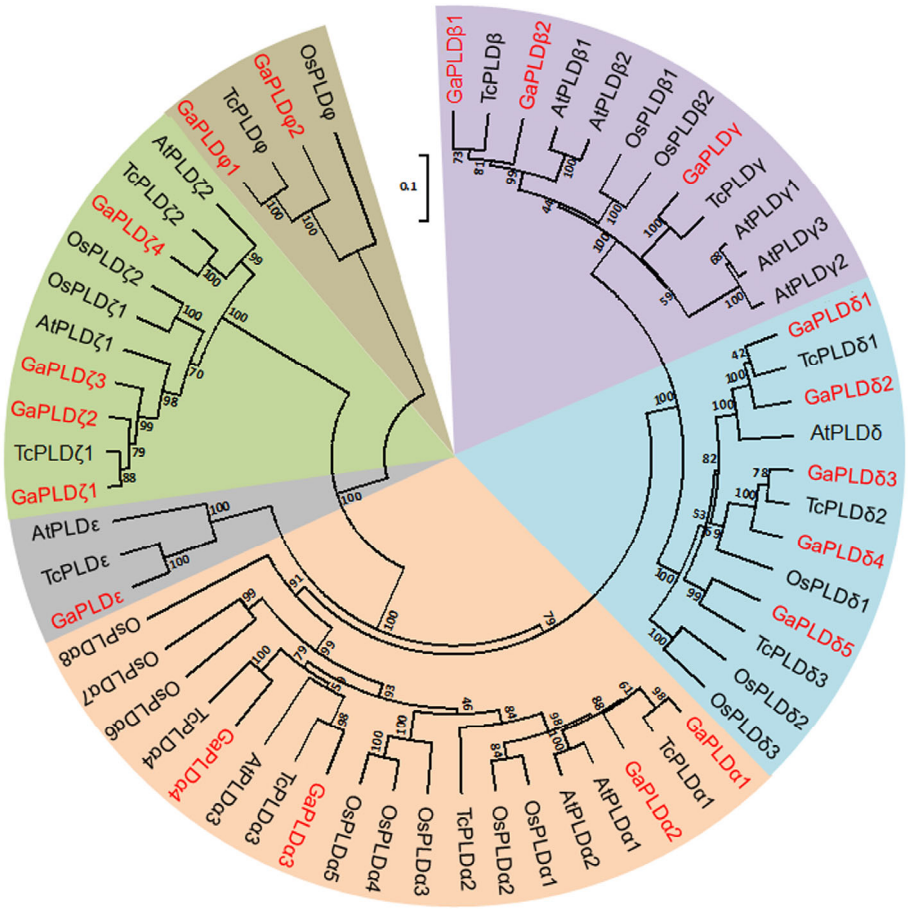

Figure 2 Phylogenetic analyses of the plant PLD proteins. The conserved PLD proteins from *G. arboreum* (GaPLD), cacao (TcPLD), *Arabidopsis* (AtPLD) and rice (OsPLD) were aligned using Clustal W 2.0, and the phylogenic tree was constructed using the NJ method with bootstrapping analysis (1,000 replicates). The numbers beside the branches indicate the bootstrap values that support the adjacent node. GaPLDs were highlighted in red.

the *GaPLD* members in the same clade of phylogenetic tree shared similar exon-intron organization. For example, the members from β/γ- and δ-clades were close in phylogenetic relationships and had 9~10 exons. Similarly, the α- and ε-clades members possessed 2~4 exons. Members of ζ-clade which were distinct from the others possessed 19~20 exons. The rest members from φ-clade had 7~8 exons (Figure 3A). Despite the exon numbers, most exons within members of the same clade had highly similar lengths. For instance, the lengths of the last exon in δ-clade members were about 294 (Figure 3A). However, the intron lengths were not so conserved (Figure 3A). The synthesis of sequence identities, phylogenetic relationships and gene structures of *GaPLDs* suggested that β/γ- and δ-clades were originated from a common ancestor in cotton, so were α- and ε-clades, whereas ζ- and φ-clades respectively originated via the independent evolutionary paths separate to the others, just like the analysis in poplar and grape (Liu et al., 2010).

We further performed the domain analysis of GaPLD proteins by Hmmpfam program to identify the major ones. Here, the analysis showed that all the GaPLDs possessed two conserved HKD domains (Figure 3B). Both the HKD1 and HKD2 domains contained three highly conserved amino acids (H6, K8 and D13; Figure 3C and 3D). However, when the flanking sequences were compared, the HKD1 sequence was relatively more diverse than that of the HKD2 domain (Figure 3C and 3D). Meanwhile, other four kinds of conserved domains (C2 domain, PX domain, PH domain and SP) were also identified in the N-terminal region of GaPLD proteins, which were noted previously in other plant PLD proteins (Li et al., 2007; Liu et al., 2010; Qin and Wang, 2002). Based on their domain architectures, the GaPLD proteins could be divided into three subfamilies: C2-PLD, PX/PH-PLD and SP-PLD. In detail, the α, β, γ, δ and ε-clades of PLDs were classified in the C2-PLD subfamily, and the PLDζs and PLDφs were classified in the PX/PH-PLD subfamily and the SP-PLD subfamily, respectively (Figure 3B), consistent with the findings in other plants (Li et al., 2007; Liu et al., 2010; Qin et al., 1997).

To obtain more insights into the diversity of motif compositions among different GaPLD proteins, we further searched for the conserved motifs by MEME program. As shown in Figure 4, a total of 20 conserved motifs designated

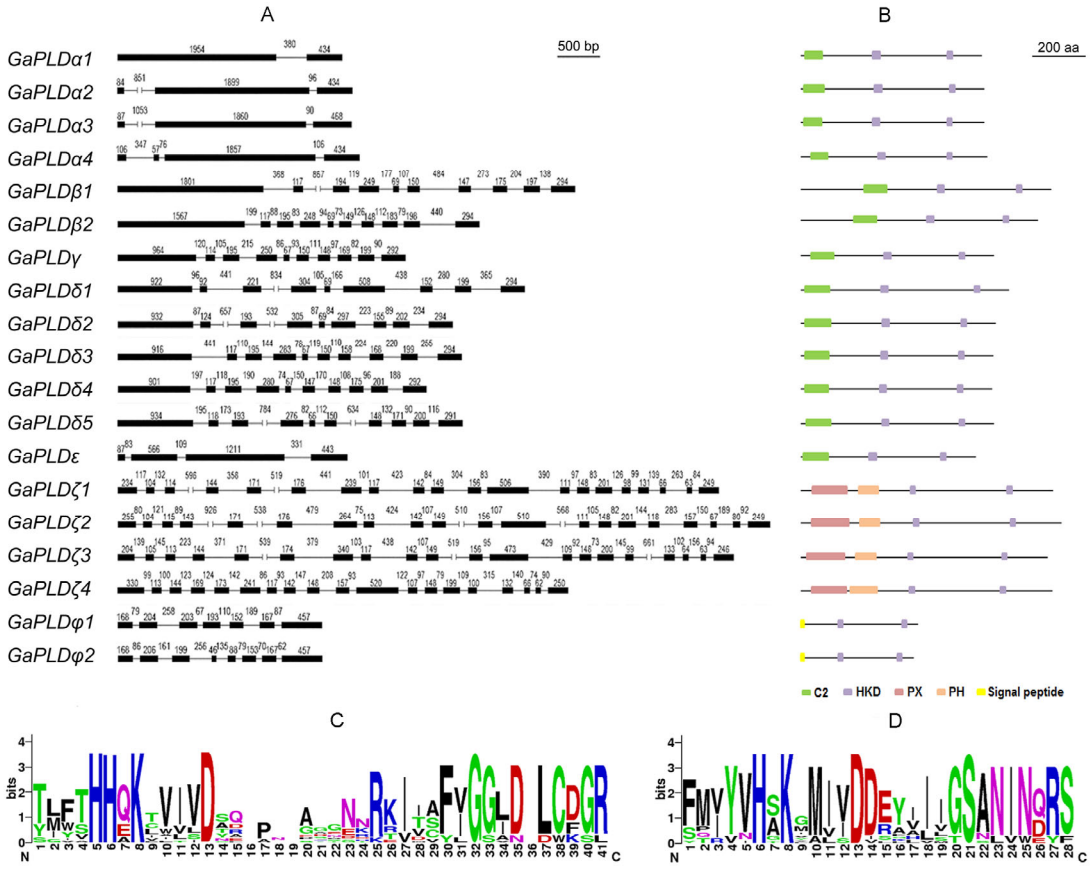

Figure 3 Exon-intron organization of *GaPLD* genes and the deduced protein domain analysis. A, Schematic diagram for the exon/intron organization of *GaPLD* genes. The *boxes* and *lines* indicate the exons and introns, respectively, and their lengths are indicated in base pairs. B, Schematic diagram for domain structures of GaPLD proteins. The C2, PX, PH and HKD domain and signal peptide are represented by several *rectangles* with different colours. C and D, Sequence logos for the two HKD domains (C: HKD1, D: HKD2). Numbers on the *x*-axis represent the sequence positions of respect HKD domain. The *y*-axis represents the information content measured in bits.

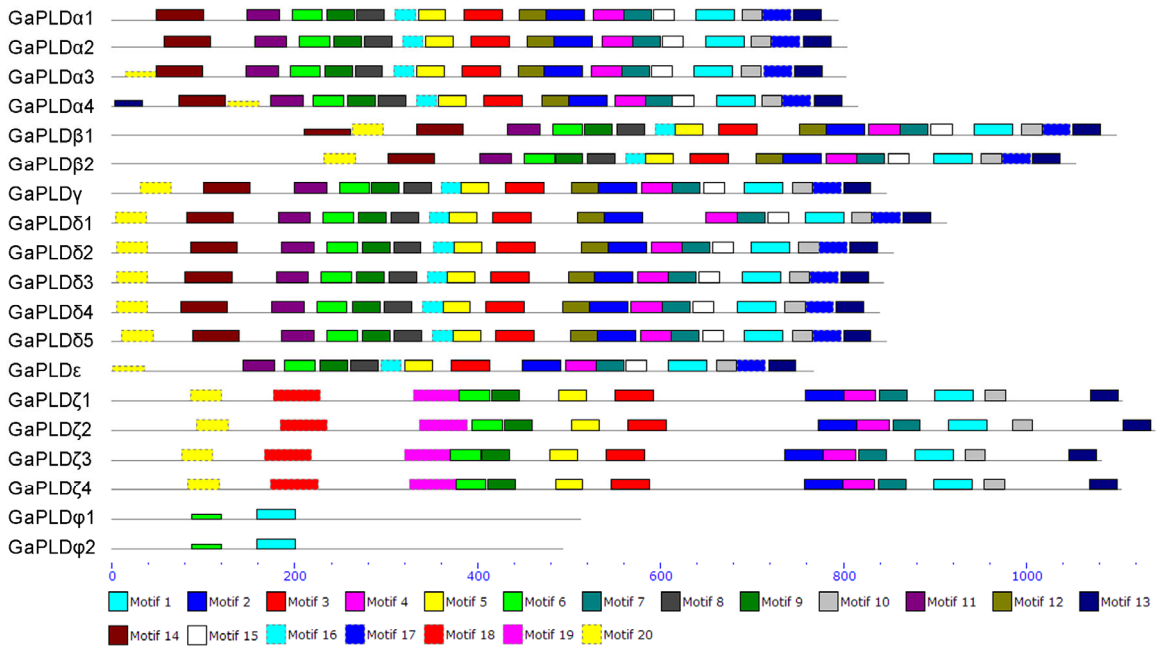

Figure 4 MEME/MAST motif analysis of GaPLD proteins. Different motifs are indicated by different colors. The regular expression sequences of the motifs 1~20 are listed in Table S5.

as motif 1 to motif 20 were identified. The GaPLD proteins within the same subfamily shared similar motif composition. Most of the C2-PLDs possessed the motifs 1~17 and 20, while all PX/PH-PLD members contained the motifs 1~7, 9, 10, 13, and 18~20. By comparison, the SP-PLD subfamily (GaPLDφ1 and GaPLDφ2) only shared motifs 1 and 6 with the other two subfamilies (Figure 4). Moreover, the predicted motifs were annotated by ScanProsite. However, only five motifs (1~5) could be matched to the annotated motifs in the database. Motif 1 and 5 were annotated as the conserved HKD domains, especially motif 1 was uniformly observed in all PLD proteins (Table S5, Figures S1 and S2). Motif 2 contained a regular-expression sequence "IYIENQ[F/Y]F". The seventh amino acid of this sequence, Phenylalanine (F), appeared in all PX/PH-PLDs, but was often substituted by Tyrosine (Y) in the C2-PLDs (Table S5, Figure S3). This short sequence "IYIENQ[F/Y]F" was only found in the PLD gene family members, and has been postulated to increase the rate of catalysis and ensure substrate specificity (McDermott et al., 2004). Motif 3 was considered as the binding site of PIP_2, and the variations in the sequence of this motif exhibited different PIP_2 binding affinity (Table S5, Figure S4) (Pappan et al., 1997). Motif 4 contained a highly conserved core triad "ERF" in the C2-PLDs (Table S5, Figure S5), and was reported to be able to bind to the α subunit of the heterotrimeric G protein (Zhao and Wang, 2004).

Expression profiles of *GaPLD* genes

To investigate the tissue-specific expression profiles of *GaPLD* genes, the quantitative RT-PCR analysis was performed for seven different tissues, including root, stem, leaf, hypocotyl, petal, anther and fiber. As indicated in Figure 5A, *GaPLDα1* and *GaPLDδ2* were constitutively expressed, and the expression levels were very high, implying that these two *GaPLDs* might play important roles at multiple developmental stages (Figure 5A). However, the expression of *GaPLDα3, GaPLDα4, GaPLDβ2, GaPLDδ3, GaPLDδ4, GaPLDδ5, GaPLDζ1* and *GaPLDφ2* was very low, or even could not be detected in some tissues (Figure 5A), although they might be primarily expressed in other organs not tested or under some special conditions. For *GaPLDα2* and *GaPLDβ1*, their expression was relatively

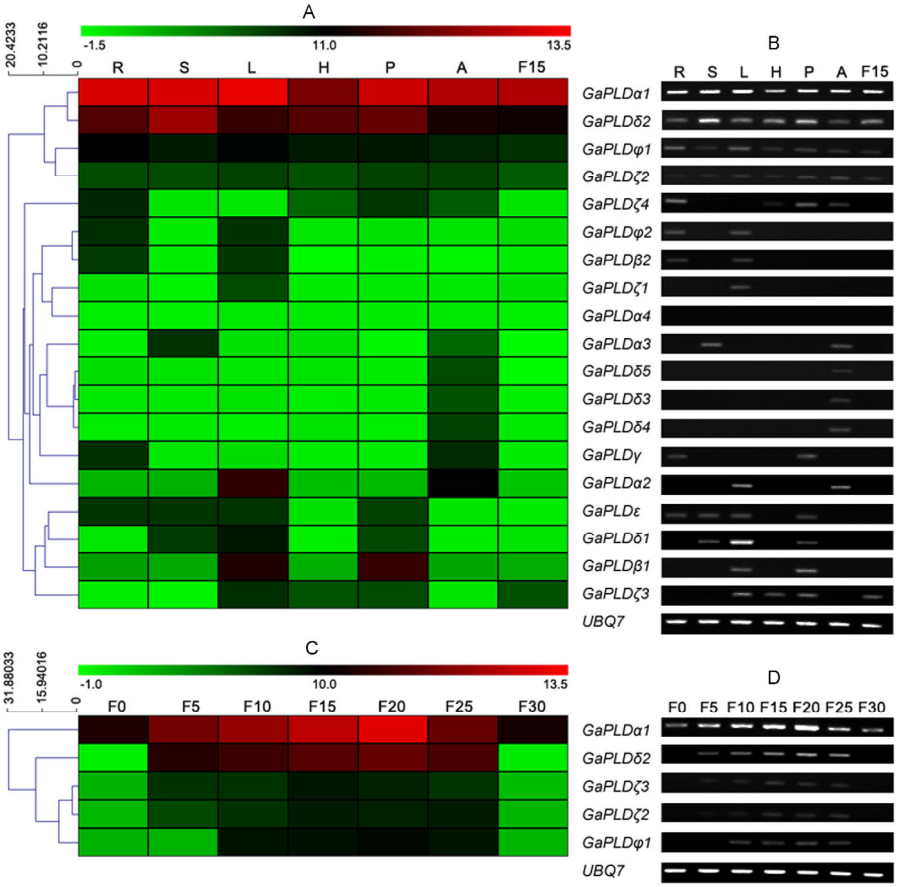

Figure 5 Expression patterns of *GaPLD* genes. A and B, Quantitative (A) and semi-quantitative RT-PCR analysis (B) of *GaPLD* genes in root (R), stem (S), leaf (L), hypocotyl (H), petal (P), anther (A), and fiber at 15 DPA (F15) of cotton plants. C and D, Quantitative (C) and semi-quantitative RT-PCR analysis (D) of the *GaPLD* genes in cotton fibers at different developmental stages. F0, ovules from 0 DPA; and F5 to F30, fibers from 5 to 30 DPA. Results were normalized using cotton *UBQ7* gene expression as the internal control.

high in leaf, petal and anther, but low in root, stem, hypocotyl and fiber, indicating that they might play an important role in the development of leaf and flower. In particular, *GaPLDα1*, *GaPLDδ2*, *GaPLDζ2, GaPLDζ3* and *GaPLDφ1* could be detected in fiber, implying that they might functionally related with the cotton fiber development (Figure 5A). The semi-quantitative RT-PCR analysis also gave the similar results, and also indicated that the amplified segment for each *GaPLD* gene was very specific (Figure 5B).

To further determine the roles of these five *GaPLD* genes in fiber development, qRT-PCR was also performed at seven representative stages of fiber development (0, 5, 10, 15, 20, 25 and 30 DPA). The expression levels of *GaPLDα1* and *GaPLDδ2* in fibers were quite higher than those of others (Figure 5C). In detail, the expression level of *GaPLDα1* increased from 0 to 20 DPA, and reached to the peak at 20 DPA, and after then, the expression reduced. For *GaPLDδ2*, although its expression was low during 10 DPA, the levels rose from 15 to 25 DPA, and then fell from 25 to 30 DPA (Figure 5C). However, the other three genes, *GaPLDζ2, GaPLDζ3* and *GaPLDφ1*, expressed at very low levels during the whole stages of fiber development (Figure 5C). Additionally, the semi-quantitative RT-PCR analysis gave us the similar results as that quantitative RT-PCR analysis did (Figure 5D).

Taken together, *GaPLDα1* and *GaPLDδ2* were highly expressed in fiber, and their expression differentiated at different developing stages, indicating that they might play distinct functions in fiber development. For *GaPLDα1*, its expression pattern was obviously correspond to dynamics of reactive oxygen species (ROS) in differentiating cotton fibers (Yang et al., 2008b). There were some evidences that ROS may function as a developmental signal in the differentiation of secondary walls in cotton (Potikha et al., 1999). Moreover, *GaPLDα1*-ortholog gene in *Arabidopsis*, *AtPLDα1*, hydrolyzed structural phospholipids in the biological membranes to produce PA and a free head group. In *Arabidopsis*, PA could function as a signal to activate the NADPH-oxidase and lead to the ROS production (Zhang et al., 2009). Consistent with the previous proteomic analysis, our results indicated that *GaPLDα1* was highly expressed in fibers at 20 DPA, H_2O_2 level of which was at the peak (Potikha et al., 1999; Yang et al., 2008b), and suggested that *GaPLDα1* might participate in signal transduction for the release of ROS via activation of NADPH-oxidase in elongating fiber cells. The other fiber-related *GaPLD* gene, *GaPLDδ2*, might also be associated with ROS. In *Arabidopsis*, the ortholog of *GaPLDδ2*, *AtPLDδ*, was proven to be activated by H_2O_2. Deficiency of *AtPLDδ* could render *Arabidopsis* cells more sensitive to H_2O_2-promoted programmed cell death than the wild type (Wang and Wang, 2001; Wang et al., 2006; Zhang et al., 2003; Zhang et al., 2005; Zhang et al., 2009). Furthermore, a recent study suggested that AtPLDδ could interact with another cytosolic glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, and this interaction was promoted by ROS. Knockdown of *AtPLDδ* also impeded stomatal response to abscisic acid (ABA) and H_2O_2, also placing AtPLDδ downstream of H_2O_2 (Guo et al., 2012). All of these results indicated that *GaPLDδ2* might also function during ROS-mediated fiber development.

CONCLUSIONS

In this work, we systematically analyzed the PLD gene family in *G. arboreum* genome. 19 *GaPLD* genes were identified and divided into six clades (α, β/γ, δ, ε, ζ and φ) by their phylogenetic relationships. The *GaPLDs* within the same clade were highly similar in terms of sequence characteristics, including gene structure and protein domain. The chromosomal mapping and evolution analysis suggested that the segmental duplication events were the main cause of the expansion of *GaPLD* gene family. These duplication events had happened approximate 19.61~20.44 million years ago, when large-scale recent genome duplication events occurred in cotton. Notably, the expression of *GaPLDα1* and *GaPLDδ2* were highly expressed in developing fibers, and might play roles in fiber development. Taken together, the results and information described in this work provide a solid basis for the further investigation of evolution history and biological functions of *PLD* genes in cotton.

MATERIALS AND METHODS

Plant materials preparation

Cotton (*G. arboreum* cultivar "Shixiya1") was cultivated in a normal agronomic field from May to September under standard conditions in Beijing. The seeds were kindly provided by the Cotton Research Institute, Chinese Academy of Agricultural Sciences. When cotton plants were in full bloom (approximately 90 d after planting), different cotton tissues including roots, stems, leaves, hypocotyls, petals, and anthers were harvested. And cotton fibers were harvested at 0, 5, 10, 15, 20, 25 and 30 d post anthesis (DPA). All of the samples were immediately frozen in liquid nitrogen and then stored at −80°C until RNA extraction.

Identification of PLD gene family in *G. arboreum*

The conserved HKD domain based on Hidden Markov Model (HMM) (PF00614) was obtained from Pfam protein family database (http://pfam.sanger.ac.uk/) (Finn et al.,

2014). In order to identify the *PLD* encoding genes of *G. arboreum*, the HHM profile of HKD domain and corresponding sequences from *Arabidopsis* and rice (*Oryza sativa*) were subsequently employed as query to perform the blastp and tblastn algorithms (Altschul et al., 1997) and HMMER search (http://hmmer.janelia.org/) (Eddy, 2009) against the *G. arboreum* genome databases (http://cgp.genomics.org.cn/) (Li et al., 2014). All the redundant sequences were discarded from further analysis based on Clustal W alignment (Thompson et al., 1994). Furthermore, to verify the reliability of the initial results, all non-redundant candidate *PLD* sequences were analyzed to confirm the presence of the two conserved HKD domains using the InterproScan program (Quevillon et al., 2005). The *PLD* sequences in *Arabidopsis* and rice were retrieved from the published references (Li et al., 2007; Qin and Wang, 2002). The theoretical molecular weight (Mw) and isoelectric point (pI) were calculated by ExPASy (http://cn.expasy.org/tools). Subcellular localization was analyzed using the CELLO v2.5 server (http://cello.life.nctu.edu.tw/).

Chromosomal location analysis and gene duplication

Paralogous *GaPLD* gene pairs were identified on the basis of multiple sequence alignments. The following criteria were adopted: the shorter sequences covers over 70% of the longer sequence after alignment and the minimum identity of aligned regions is 70%, as described previously (Gu et al., 2002; Yang et al., 2008a). Pairwise alignment of nucleotide sequences of *GaPLD* paralogs was performed using Clustal W. Gaps in the alignments were removed manually by Bioedit. The *Ka* (nonsynonymous substitution rates) and *Ks* (synonymous substitution rates) values of the paralogous genes were estimated by the program *KaKs*_Calculator (Zhang et al., 2006). As the theory of molecular clock, *Ks* could be used as the proxy for time and the conserved flanking protein-coding genes was used to estimate the dates of the segmental duplication events (Shiu et al., 2004). The mean *Ks* value was calculated for each of duplicated gene pairs and then used to date the duplication events. *Ks* values greater than 2.0 were discarded because higher *Ks* values are associated with a large degree of uncertainty because of saturation of substitutions. The *Ks* values were then used to calculate the approximate date of the duplication event ($T=Ks/2\lambda$), assuming clock-like rates (λ) of synonymous substitution of 1.5×10^{-8} substitutions/synonymous site/year for diploid cotton species (Blanc and Wolfe, 2004).

Phylogenetic analysis

Multiple sequence alignment was conducted on the amino acid sequences of PLD proteins in *G. arboreum*, *Arabidopsis*, rice and cacao (*Theobroma cacao*) genomes using Clustal W with the default settings (Thompson et al., 1994). Subsequently, MEGA 6.0 was employed to construct an unrooted phylogenetic tree based on alignments using the Neighbor Joining (NJ) method with *P*-distance and pairwise gap deletion parameters engaged (Tamura et al., 2013). The reliability of the trees obtained was tested using bootstrapping with 1,000 replicates. Furthermore, Maximum likelihood and Minimal Evolution methods were also applied to validate the results from the NJ tree. The PLD protein sequences in cacao were from *T. cacao* genome sequence databases (http://www.phytozome. net/cacao).

Gene structure and domain analysis and conserved motif identification

Exon-intron organization of *PLD* genes was generated by the GSDS (Gene Structure Display Server) algorithm (Guo et al., 2007), by comparing the coding sequences (CDS) of *GaPLD* genes with their corresponding genomic sequences.

Protein sequences derived from the *GaPLD* genes were examined using the domain analysis program Pfam with the default cut off parameters.

GaPLD protein sequences were submitted to online Multiple Expectation maximization for Motif Elicitation (MEME) program for identification of conserved motifs (Bailey et al., 2006). The following parameters were employed in the analysis: the maximum number of motifs 20; minimum motif width 6; and maximum motif width 50. The identified protein motifs were further annotated with ScanProsite (de Castro et al., 2006).

RNA isolation and real-time quantitative RT-PCR

Total RNA of cotton tissues was extracted from frozen tissue using RNAprep pure plant kit (TIANGEN, Beijing) according to the manufacturer's protocol. 2 μg of RNA was used as the template for the first-strand cDNA synthesis using an RNA PCR kit (AMV, version 3.0, TaKaRa, Japan). PCR was performed using the gene-specific primers (Table S6). Real-time quantitative RT-PCR was performed with a Mini opticon Real-Time PCR System (Bio-Rad, USA) according to the supplier's protocol. Each reaction mixture contains 8 μL of DNase/RNase free water, 10 μL Real-time SYBR Green PCR master mix, 1 μL diluted cDNA product from RT-PCR reaction and 1 μL gene-specific primers. Three biological replicates conducted for each tissue and each biological replicate was technically repeated three times. The thermal cycle applied was as follows: 95°C for 10 min followed by 45 cycles of denature at 95°C for 15 s and annealing and elongation at 60°C for 60 s. The expression values of *GaPLD* genes tested were normalized with an internal reference gene *UBQ7* (Gene accession number: DQ116441). The relative expression levels were calculated using the comparative $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Each PCR was run in triplicate in each assay. A heatmap for gene expression patterns were generated with the software MultiExperiment Viewer (MeV).

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by grants from the State Key Basic Research and Development Plan (2010CB126003), and the National Transgenic Animals and Plants Research Project (2011ZX08005-003, 2011ZX08009-003). The authors would like to acknowledge members of the Laboratory of Molecular Biology at Tsinghua University for critical discussions.*

Adams, K.L., Cronn, R., Percifield, R., and Wendel, J.F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100, 4649–4654.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, L., Shi, Z., Berard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon, V., Sabau, X., Axtell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Golser, W., Song, X., Clement, D., Rivallan, R., Tahi, M., Akaza, J.M., Pitollat, B., Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S., and Lanaud, C. (2011). The genome of *Theobroma cacao*. Nat Genet 43, 101–108.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34, W369–W373.

Blanc, G., and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667–1678.

Cao, X. (2015). Whole genome sequencing of cotton--a new chapter in cotton genomics. Sci China Life Sci 58, 515–516.

Chen, G., Snyder, C.L., Greer, M.S., and Weselake, R.J. (2011). Biology and Biochemistry of Plant Phospholipases. Crit Rev Plant Sci 30, 239–258.

de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34, W362–W365.

Desai, A., Chee, P.W., Rong, J., May, O.L., and Paterson, A.H. (2006). Chromosome structural changes in diploid and tetraploid A genomes of *Gossypium*. Genome 49, 336–345.

Du, D., Cheng, T., Pan, H., Yang, W., Wang, J., and Zhang, Q. (2013). Genome-wide identification, molecular evolution and expression analyses of the phospholipase D gene family in three Rosaceae species. Sci Hortic-Amsterdam 153, 13–21.

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform 23, 205–211.

Elias, M., Potocky, M., Cvrckova, F., and Zarsky, V. (2002). Molecular diversity of phospholipase D in angiosperms. BMC genomics 3, 2.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Res 42, D222–D230.

Gu, Z., Cavalcanti, A., Chen, F., Bouman, P., and Li, W. (2002). Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. Mol Biol Evol 19, 256–262.

Guo, A., Zhu, Q., Chen, X., and Luo, J. (2007). GSDS: a gene structure display server. Yi Chuan 29, 1023–1026.

Guo, L., Devaiah, S.P., Narasimhan, R., Pan, X., Zhang, Y., Zhang, W., and Wang, X. (2012). Cytosolic glyceraldehyde-3-phosphate dehydrogenases interact with phospholipase Ddelta to transduce hydrogen peroxide signals in the *Arabidopsis* response to stress. Plant Cell 24, 2200–2212.

Hanahan, D.J., and Chaikoff, I.L. (1947). A new phospholipide-splitting enzyme specific for the ester linkage between the nitrogenous base and the phosphoric acid grouping. J Biol Chem 169, 699–705.

Hong, Y., Devaiah, S.P., Bahn, S.C., Thamasandra, B.N., Li, M., Welti, R., and Wang, X. (2009). Phospholipase D epsilon and phosphatidic acid enhance Arabidopsis nitrogen signaling and growth. Plant J 58, 376–387.

Hong, Y., Pan, X., Welti, R., and Wang, X. (2008). Phospholipase D alpha 3 is involved in the hyperosmotic response in *Arabidopsis*. Plant Cell 20, 803–816.

Hong, Y., Zhang, W., and Wang, X. (2010). Phospholipase D and phosphatidic acid signalling in plant response to drought and salinity. Plant Cell Environ 33, 627–635.

Hu, G., Koh, J., Yoo, M.J., Grupp, K., Chen, S., and Wendel, J.F. (2013). Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. New Phytol 200, 570–582.

Jacobs, A.C., Hood, I., Boyd, K.L., Olson, P.D., Morrison, J.M., Carson, S., Sayood, K., Iwen, P.C., Skaar, E.P., and Dunman, P.M. (2010). Inactivation of phospholipase D diminishes *Acinetobacter baumannii* pathogenesis. Infect Immun 78, 1952–1962.

Kargiotidou, A., Kappas, I., Tsaftaris, A., Galanopoulou, D., and Farmaki, T. (2010). Cold acclimation and low temperature resistance in cotton: *Gossypium* hirsutum phospholipase Dalpha isoforms are differentially regulated by temperature and light. J Exp Botany 61, 2991–3002.

Kim, D.U., Roh, T.Y., Lee, J., Noh, J.Y., Jang, Y.J., Hoe, K.L., Yoo, H.S., and Choi, M.U. (1999). Molecular cloning and functional expression of a phospholipase D from cabbage (*Brassica oleracea* var. *capitata*). Biochim Biophys Acta 1437, 409–414.

Lein, W., and Saalbach, G. (2001). Cloning and direct G-protein regulation of phospholipase D from tobacco. Biochim Biophys Acta 1530, 172–183.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., Liu, W., Shi, C., Xiao, G., Gou, C., Ye, W., Xu, X., Zhang, X., Wei, H., Li, Z., Zhang, G., Wang, J., Liu, K., Kohel, R.J., Percy, R.G., Yu, J., Zhu, Y., Wang, J., and Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. Nature genetics 46, 567–572.

Li, G., Lin, F., and Xue, H. (2007). Genome-wide analysis of the phospholipase D family in *Oryza sativa* and functional characterization of PLD beta 1 in seed germination. Cell Res 17, 881–894.

Liu, Q., Zhang, C., Yang, Y., and Hu, X. (2010). Genome-wide and molecular evolution analyses of the phospholipase D gene family in Poplar and Grape. BMC Plant Biol 10, 117.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408.

McDermott, M., Wakelam, M.J.O., and Morris, A.J. (2004). Phospholipase D. Biochem Cell Biol 82, 225–253.

Mendonsa, R., and Engebrecht, J. (2009). Phospholipase D function in Saccharomyces cerevisiae. Biochim Biophys Acta 1791, 970–974.

Pappan, K., Qin, W., Dyer, J.H., Zheng, L., and Wang, X. (1997). Molecular cloning and functional analysis of polyphosphoinositide-dependent phospholipase D, PLDbeta, from *Arabidopsis*. J Biol Chem 272, 7055–7061.

Paterson, A.H., Wendel, J.F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K.C., Shu, S., Udall, J., Yoo, M.J., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M.V., Gong, L., Grimwood, J., Grover, C., Grupp, K., Hu, G., Lee, T.H., Li, J., Lin, L., Liu, T., Marler, B.S., Page, J.T., Roberts, A.W., Romanel, E., Sanders, W.S., Szadkowski, E., Tan, X., Tang, H., Xu, C., Wang, J., Wang, Z., Zhang, D., Zhang, L., Ashrafi, H., Bedon, F., Bowers, J.E., Brubaker, C.L., Chee, P.W., Das, S., Gingle, A.R., Haigler, C.H., Harker, D., Hoffmann, L.V., Hovav, R., Jones, D.C., Lemke, C., Mansoor, S., ur Rahman, M., Rainville, L.N., Rambani, A., Reddy, U.K., Rong, J.K.,

Saranga, Y., Scheffler, B.E., Scheffler, J.A., Stelly, D.M., Triplett, B.A., Van Deynze, A., Vaslin, M.F., Waghmare, V.N., Walford, S.A., Wright, R.J., Zaki, E.A., Zhang, T., Dennis, E.S., Mayer, K.F., Peterson, D.G., Rokhsar, D.S., Wang, X., and Schmutz, J. (2012). Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature 492, 423–427.

Pei, Y. (2015). The homeodomain-containing transcription factor, GhHOX3, is a key regulator of cotton fiber elongation. Sci China Life Sci 58, 309–310.

Peng, X., and Frohman, M.A. (2012). Mammalian phospholipase D physiological and pathological roles. Acta Physiol 204, 219–226.

Pleskot, R., Li, J., Zarsky, V., Potocky, M., and Staiger, C.J. (2013). Regulation of cytoskeletal dynamics by phospholipase D and phosphatidic acid. Trends Plant Sci 18, 496–504.

Potikha, T.S., Collins, C.C., Johnson, D.I., Delmer, D.P., and Levine, A. (1999). The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. Plant Physiol 119, 849–858.

Qin, C., and Wang, X. (2002). The *Arabidopsis* phospholipase D family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains. Plant Physiol 128, 1057–1068.

Qin, W., Pappan, K., and Wang, X. (1997). Molecular heterogeneity of phospholipase D (PLD). Cloning of PLDgamma and regulation of plant PLDgamma, -beta, and -alpha by polyphosphoinositides and calcium. J Biol Chem 272, 28267–28273.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res 33, W116–W120.

Shiu, S.H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., and Li, W.H. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. Plant Cell 16, 1220–1234.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30, 2725–2729.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673–4680.

Ueki, J., Morioka, S., Komari, T., and Kumashiro, T. (1995). Purification and characterization of phospholipase D (PLD) from rice (*Oryza sativa* L.) and cloning of cDNA for PLD from rice and maize (*Zea mays* L.). Plant Cell Physiol 36, 903–914.

Wang, C., and Wang, X. (2001). A novel phospholipase D of *Arabidopsis* that is activated by oleic acid and associated with the plasma membrane. Plant Physiol 127, 1102–1112.

Wang, C., Zien, C.A., Afitlhile, M., Welti, R., Hildebrand, D.F., and Wang, X. (2000). Involvement of phospholipase D in wound-induced accumulation of jasmonic acid in *Arabidopsis*. Plant Cell 12, 2237–2246.

Wang, G., Ryu, S., and Wang, X. (2012). Plant phospholipases: an overview. Methods Mol Biol 861, 123–137.

Wang, X., Devaiah, S.P., Zhang, W., and Welti, R. (2006). Signaling functions of phosphatidic acid. Prog Lipid Res 45, 250–278.

Wang, X., Xu, L., and Zheng, L. (1994). Cloning and expression of phosphatidylcholine-hydrolyzing phospholipase D from Ricinus communis L. J Biol Chem 269, 20312–20317.

Wang, X. (2005). Regulatory functions of phospholipase D and phosphatidic acid in plant growth, development, and stress responses. Plant Physiol 139, 566–573.

Wanjie, S.W., Welti, R., Moreau, R.A., and Chapman, K.D. (2005). Identification and quantification of glycerolipids in cotton fibers: reconciliation with metabolic pathway predictions from DNA databases. Lipids 40, 773–785.

Whitaker, B.D., Smith, D.L., and Green, K.C. (2001). Cloning, characterization and functional expression of a phospholipase Dalpha cDNA from tomato fruit. Physiol Plantarum 112, 87–94.

Yang, S., Zhang, X., Yue, J., Tian, D., and Chen, J. (2008a). Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Genet Genomics : MGG 280, 187–198.

Yang, Y., Bian, S., Yao, Y., and Liu, J. (2008b). Comparative proteomic analysis provides new insights into the fiber elongating process in cotton. J Proteome Res 7, 4623–4637.

You, X., Zhang, Y., Li, L., Li, Z., Li, M., Li, C., Zhu, J., Peng, H., and Sun, J. (2014). Cloning and molecular characterization of phospholipase D (PLD) delta gene from longan (*Dimocarpus longan* L). Mol Biol Rep 41, 4351–4360.

Zhang, W., Wang, C., Qin, C., Wood, T., Olafsdottir, G., Welti, R., and Wang, X. (2003). The oleate-stimulated phospholipase D, PLDdelta, and phosphatidic acid decrease H_2O_2-induced cell death in *Arabidopsis*. Plant Cell 15, 2285–2295.

Zhang, W., Yu, L., Zhang, Y., and Wang, X. (2005). Phospholipase D in the signaling networks of plant response to abscisic acid and reactive oxygen species. Biochim Biophys Acta 1736, 1–9.

Zhang, Y., Zhu, H., Zhang, Q., Li, M., Yan, M., Wang, R., Wang, L., Welti, R., Zhang, W., and Wang, X. (2009). Phospholipase dalpha1 and phosphatidic acid regulate NADPH oxidase activity and production of reactive oxygen species in ABA-mediated stomatal closure in *Arabidopsis*. Plant Cell 21, 2357–2377.

Zhang, Z., Li, J., Zhao, X., Wang, J., Wong, G., and Yu, J. (2006). KaKs_Calculator: calculating *Ka* and *Ks* through model selection and model averaging. Genomics, Proteomics Bioinformatics 4, 259–263.

Zhao, J., and Wang, X. (2004). *Arabidopsi*s phospholipase D alpha 1 interacts with the heterotrimeric G-protein alpha-subunit through a motif analogous to the DRY motif in G-protein-coupled receptors. J Biol Chem 279, 1794–1800.

Zhao, J., Zhou, D., Zhang, Q., and Zhang, W. (2012). Genomic analysis of phospholipase D family and characterization of GmPLDalphas in soybean (*Glycine max*). J Plant Res 125, 569–578.

SUPPORTING INFORMATION

Table S1 Nucleic acid, deduced amino acid and promoter sequences of *GaPLD* genes

Table S2 Sequence identities among the ORF regions of the *GaPLD* genes

Table S3 Sequence identities among the GaPLD proteins

Table S4 The number of *PLD* genes in *G. arboreum* and other plant species

Table S5 The sequences of the 20 motifs by motif analysis of GaPLDs

Table S6 Primers used for quantitative and semi-quantitative RT-PCR

Figure S1 Alignment of sequences of MEME motif 1 in PLD genes in Gossypium arboreum. The gray front indicates the sites which flanking MEME motif 1 and the other colorful front indicate identical and conserved amino acid residues present in MEME motif 1. The alignment of sequences of MEME motif 1 is ordered by P-value.

Figure S2 Alignment of sequences of MEME motif 5 in *PLD* genes in *Gossypium arboreum*. The gray front indicates the sites which flanking MEME motif 5 and the other colorful front indicate identical and conserved amino acid residues present in MEME motif 5. The alignment of sequences of MEME motif 5 is ordered by *P*-value.

Figure S3 Alignment of sequences of MEME motif 2 in *PLD* genes in *Gossypium arboreum*. The gray front indicates the sites which flanking MEME motif 2 and the other colorful front indicate identical and conserved amino acid residues present in MEME motif 2. The alignment of sequences of MEME motif 2 is ordered by *P*-value.

Figure S4 Alignment of sequences of MEME motif 3 in *PLD* genes in *Gossypium arboreum*. The gray front indicates the sites which flanking MEME motif 3 and the other colorful front indicate identical and conserved amino acid residues present in MEME motif 3. The alignment of sequences of MEME motif 3 is ordered by *P*-value.

Figure S5 Alignment of sequences of MEME motif 4 in *PLD* genes in *Gossypium arboreum*. The gray front indicates the sites which flanking MEME motif 4 and the other colorful front indicate identical and conserved amino acid residues present in MEME motif 4. The alignment of sequences of MEME motif 4 is ordered by *P*-value.

The supporting information is available online at life.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.