

RESEARCH ARTICLE

# Genome-Wide Comparative Analysis of the Phospholipase D Gene Families among Allotetraploid Cotton and Its Diploid Progenitors

Kai Tang<sup>1</sup>, Chun-Juan Dong<sup>2</sup>, Jin-Yuan Liu<sup>1\*</sup>

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

\* [liujy@mail.tsinghua.edu.cn](mailto:liujy@mail.tsinghua.edu.cn)



**OPEN ACCESS**

**Citation:** Tang K, Dong C-J, Liu J-Y (2016) Genome-Wide Comparative Analysis of the Phospholipase D Gene Families among Allotetraploid Cotton and Its Diploid Progenitors. PLoS ONE 11(5): e0156281. doi:10.1371/journal.pone.0156281

**Editor:** David D Fang, USDA-ARS-SRRC, UNITED STATES

**Received:** March 4, 2016

**Accepted:** May 11, 2016

**Published:** May 23, 2016

**Copyright:** © 2016 Tang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** 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 and 2011ZX08009-003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

In this study, 40 phospholipase D (PLD) genes were identified from allotetraploid cotton *Gossypium hirsutum*, and 20 PLD genes were examined in diploid cotton *Gossypium raimondii*. Combining with 19 previously identified *Gossypium arboreum* PLD genes, a comparative analysis was performed among the PLD gene families among allotetraploid and two diploid cottons. Based on the orthologous relationships, we found that almost each *G. hirsutum* PLD had a corresponding homolog in the *G. arboreum* and *G. raimondii* genomes, except for *GhPLDβ3A*, whose homolog *GaPLDβ3* may have been lost during the evolution of *G. arboreum* after the interspecific hybridization. Phylogenetic analysis showed that all of the cotton PLDs were unevenly classified into six numbered subgroups: α, β/γ, δ, ε, ζ and φ. An N-terminal C2 domain was found in the α, β/γ, δ and ε subgroups, while phox homology (PX) and pleckstrin homology (PH) domains were identified in the ζ subgroup. The subgroup φ possessed a single peptide instead of a functional domain. In each phylogenetic subgroup, the PLDs showed high conservation in gene structure and amino acid sequences in functional domains. The expansion of GhPLD and GrPLD gene families were mainly attributed to segmental duplication and partly attributed to tandem duplication. Furthermore, purifying selection played a critical role in the evolution of PLD genes in cotton. Quantitative RT-PCR documented that allotetraploid cotton PLD genes were broadly expressed and each had a unique spatial and developmental expression pattern, indicating their functional diversification in cotton growth and development. Further analysis of *cis*-regulatory elements elucidated transcriptional regulations and potential functions. Our comparative analysis provided valuable information for understanding the putative functions of the PLD genes in cotton fiber.

## Introduction

Upland cotton *Gossypium hirsutum* is the world's most valuable fiber crop [1, 2]. It has been widely grown in over 80 countries and accounts for more than 90% of commercial cotton production worldwide [3]. *G. hirsutum* is also studied as a model polyploid plant. It was a classic natural allotetraploid (AADD,  $2n = 4x = 52$ ) that arose from interspecific hybridization approximately 1 to 2 million years ago (mya) between the A genome diploid species *Gossypium arboreum* (AA,  $2n = 2x = 26$ ) and the D genome diploid species *Gossypium raimondii* (DD,  $2n = 2x = 26$ ) [2, 4]. In general, allopolyploid plants grow more vigorously and adaptively than their parents, and thereby allotetraploid cotton produces a higher yield and superior quality of fibers than their diploid progenitors under similar conditions [5]. Cotton fiber (commonly known as cotton lint) is the most important natural and renewable material for the textile industry and profoundly affects the world economy and human daily life [6].

Phospholipase D (PLD) is a major type of phospholipase in plants and catalyzes the hydrolysis of phospholipids at the terminal phosphodiester bond to produce a free head group and phosphatidic acid (PA) [7]. Two highly conserved HxKxxxxD (HKD) domains are the common feature of PLD proteins. The functional domains near the N-terminus are diverse. Some PLDs contained a  $\text{Ca}^{2+}$ -dependent phospholipid-binding C2 domain, while others contained the phox homology (PX) and pleckstrin homology (PH) domains. Up until now, the PLD gene families have been identified in various plants [8–11]. In *Arabidopsis*, for example, 12 *AtPLD* genes were identified, encoding three *PLD* $\alpha$ s, two *PLD* $\beta$ s, three *PLD* $\gamma$ s, one *PLD* $\delta$ , one *PLD* $\epsilon$  and two *PLD* $\zeta$ s [8]. In the rice genome, 17 PLD genes were identified, encoding 14 C2-PLDs ( $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ - and  $\epsilon$ - types PLDs), two PX/PH-PLDs ( $\zeta$ -type PLDs), and one single peptide-PLD (SP-PLD,  $\phi$ -type PLD), which possesses a signal peptide near the N-terminus instead of C2 or PX/PH domains [9]. Different PLDs are found to have different reaction requirements, lipid selectivity and subcellular location [12]. Increasing studies indicate that PLD and its product PA have been implicated in multiple plant growth and developmental processes, such as seed germination [13], seedling growth [14], pollen tube elongation [15], root hair growth [16], hypocotyl elongation [17] and leaf senescence [18]. PLD and PA are also found to play pivotal roles in signaling plant responses to various abiotic and biotic stresses, such as drought [19], salt [20], cold [21], light [14], wounding [22] and pathogen [23].

On the basis of comparative proteomic analysis in allotetraploid cotton, PLD might be a key enzyme participating in the regulation of secondary cell wall synthesis in cotton fibers [24]. Meanwhile, its product PA obviously accumulated at the initial stage of secondary cell wall thickening [25]. It was reported that PLD is responsible for a key event in signal transduction for the release of reactive oxygen species (ROS), particularly  $\text{H}_2\text{O}_2$  via NADPH-oxidase [26]. It was also shown that a ROS burst occurs during the transition from elongation to secondary cell wall synthesis in developing fibers [27].  $\text{H}_2\text{O}_2$  functions as a developmental signal in the differentiation of secondary cell walls in fiber [28]. Thus, PLD might be directly involved in  $\text{H}_2\text{O}_2$  accumulation at the early stage of secondary cell wall synthesis and may play an important role in the regulation of fiber development. In addition, PLD was also found to take part in the cold stress response in cotton fibers [21].

In our previous study, 19 *GaPLDs* were identified in the diploid cotton *G. arboreum*, and their expression profiles provided us some clues on functional diversity [29]. However, we still only have a limited understanding of expansion pattern, molecular evolution and functional diversification of this gene family in widely cultivated allotetraploid cotton. Recently, the *G. hirsutum* genome has been sequenced, opening a new chapter of cotton genomic studies [30, 31]. In this study, the PLD gene families were identified in *G. hirsutum* (*GhPLDs*) and *G. raimondii* (*GrPLDs*). Then, all of these *PLDs* were compared with the *GaPLDs* to evaluate the

phylogenetic relationship, sequence characteristics, functional divergence and selective pressure analyses. In addition, the expression profiles of *GhPLD* genes in different allotetraploid cotton tissues were examined, and the *cis*-regulatory elements were also analyzed to account for the expression specificity and transcriptional regulation. This study will provide a better understanding of the expansion and diversification of the allotetraploid cotton PLD gene family and advance PLD functional research to enhance our ability to manipulate fiber and agronomic production of cotton.

## Materials and Methods

### Plant materials

Allotetraploid cotton (*G. hirsutum* cultivar 'CRI 35') was grown in the field under standard conditions at the Tsinghua University in China. When cotton plants were in full bloom (approximately 90 days after planting), we collected different cotton tissues, including roots, stems, leaves, petals, and stamens. Cotton fibers were harvested at 0, 5, 10, 15, 20 and 25 days post anthesis (dpa). All of these samples were immediately frozen in liquid nitrogen and then stored at -80°C until RNA extraction. Each sample was collected from 40 individual cotton plants.

### Database search for cotton PLD genes

*G. hirsutum* genome data were downloaded from the CottonGen database (<http://www.cottongen.org>) [30], and *G. raimondii* genome data were obtained from the Phytozome v9.1 database (<http://www.phytozome.net/>) [32]. To identify the *G. hirsutum* and *G. raimondii* PLD genes, the PLD protein sequences from *G. arboreum*, *Arabidopsis* and rice were used as queries in BLASTP search [33]. Moreover, the HMMER search was performed in the annotation database using the HKD domain (PF00614) as a keyword. All of the candidate PLD genes were confirmed for two HKD domains by Pfam (<http://pfam.sanger.ac.uk/search>) and InterproScan databases (<http://www.ebi.ac.uk/interpro/search/sequence-search>).

### Sequence analysis methods

Multiple sequence alignment of PLD proteins was performed using Clustal W with standard settings [34]. The program MUSCLE was also used to perform multiple sequence alignments to confirm the ClustalW results [35].

Sequence identities of cotton PLDs at both the nucleotide and amino acid levels were calculated with the program DNASTAR Lasergene (<http://www.dnastar.com/>). The exon-intron structures of PLD genes were generated online by the Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn/>).

The theoretical molecular weight (Mw) and the isoelectric point (pI) of the deduced cotton PLD proteins 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 *cis*-regulatory elements in the promoter sequences were analyzed using two publicly available databases: Database of Plant *Cis*-acting Regulatory DNA Elements (PLACE) (<http://www.dna.affrc.go.jp/PLACE/>) and the Plant CARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

### Phylogenetic tree construction

The program MEGA 6.0 was used to align the deduced amino acid sequences of cotton PLD genes to construct a phylogenetic tree [36]. A neighbor joining (NJ) consensus tree was

constructed with the following parameters: P-distance, pairwise gap deletion, and bootstrap (1,000 replicates). Then, Maximum likelihood and Minimal Evolution methods of MEGA 6.0 were applied to validate the results from the NJ method. Meanwhile, Maximum Parsimony method of PHYLIP software [37] was also employed to create a new phylogenetic tree to validate the results from the NJ method. The tree topologies from different methods were very similar.

## Chromosomal location and gene duplication

The detailed physical positions of PLD genes were obtained from the CottonGen database and the Phytozome v9.1 database. Chromosomal localization and collinear relationships of genes were visualized using the program Circos [38].

Paralogous PLD genes were chosen from the results of sequence identity calculations and the phylogenetic tree. Gene duplication events were indicated by shared aligned sequence covering >70% of the longer gene and similarity of the aligned regions of >70%. A tandem duplication event has been defined as paralogous genes that are physically close to each other on the chromosomes [39]. A segmental duplication event has been defined as paralogous genes that result from large-scale events such as whole genome duplication or duplications of large chromosomal regions [40].

The  $K_a$  (nonsynonymous substitution rate) and  $K_s$  (synonymous substitution rate) values of the paralogous genes were estimated by the  $KaKs\_Calculator$  [41]. Mean  $K_s$  could be used as the proxy for time and the conserved flanking protein-coding genes were also used to estimate dates of the segmental duplication events [42]. Based on the  $\lambda$  of synonymous substitution  $1.5 \times 10^{-8}$  substitutions/ synonymous site/year for cotton [43], the approximate age of duplicated events of the duplicate PLD gene pairs was estimated ( $Time = K_s/2\lambda$ ). Moreover, the  $K_a/K_s$  ratio was used to show the selection pressure for the duplicate PLD genes. A  $K_a/K_s$  ratio >1, <1 or = 1 indicates positive, negative (purifying selection) and neutral evolution, respectively [44].

## RNA extraction and real-time quantitative RT-PCR

Total RNAs were extracted using the RNAPrep pure plant kit according to the manufacturer's instructions (TIANGEN, Beijing, China) from different cotton tissues. The first strand cDNAs were synthesized using the Takara Reverse Transcription System (TaKaRa, Shuzo, Otsu, Japan).

Quantitative real-time PCR (qRT-PCR) reactions were performed in Mini Opticon Real-Time PCR System (Bio-Rad, Hercules, California, USA) using the SYBR Green Master Mix Reagent (TaKaRa, Shuzo, Otsu, Japan) under standardized thermal cycling conditions according to the manufacturer's protocol. Each PCR reaction (20  $\mu$ L) contained 10  $\mu$ L real-time PCR Mix, 0.5  $\mu$ L of each gene-specific primer (S1 Table), and appropriately diluted cDNA. *GhUBQ7* (Accession number: DQ116441) was used as an internal reference gene. Each PCR reaction was repeated three times independently. The specificity of the reactions was verified by melting curve analysis, and products were further confirmed by agarose gel electrophoresis. The comparative  $2^{-\Delta\Delta CT}$  method was used to calculate the relative expression levels [45]. Heat maps were generated with the MultiExperiment Viewer (MeV) [46].

## RNA-sequencing data analysis

The high-throughput RNA-sequencing (RNA-seq) data of *G. hirsutum* for expression analysis in roots, stems, leaves, petals, stamens and fibers at 0, 5, 10, 20 and 25 dpa were downloaded from the National Center for Biotechnology Information Short Read Archive (<http://www.>

[ncbi.nlm.nih.gov/sra/](http://ncbi.nlm.nih.gov/sra/)) with the accession numbers SRX797899, SRX797900, SRX797901, SRX797903, SRX797904, SRX797909, SRX797917, SRX797918, SRX797919 and SRX797920, respectively [30]. The transcript abundance of each gene was calculated by the fragments per kilobase of exon model per million mapped reads (FPKM) with Cufflinks software (<http://cufflinks.cbc.umd.edu/>). The  $\log_2$ (FPKM) values were utilized for generating the heat maps using the MeV software.

## Results

### PLD genes in *G. hirsutum* and *G. raimondii* genomes

The BLASTP and HMMER searches against *G. hirsutum* and *G. raimondii* genomes were performed to identify PLD genes. Then, the candidate PLD genes were confirmed through similarity searches against Pfam and InterproScan databases. Finally, a total of 40 *G. hirsutum* PLDs (*GhPLDs*) and 20 *G. raimondii* PLDs (*GrPLDs*) were identified (Table 1 and S2 Table). The properties of newly found cotton PLDs were analyzed by ExpASY and CELLO v2.5 servers. The ORF lengths of these genes ranged from 1,545 bp to 3,693 bp, which encoded polypeptides of 514 aa to 1,230 aa with predicted molecular weights ranging from 57.91 kD to 140.56 kD (Table 1). The theoretical pIs ranged from 5.37 to 9.13. Most of the PLDs were predicted to be localized in the cytoplasm (Table 1).

The phylogenetic analysis allowed the classification of these cotton PLDs into six subgroups ( $\alpha$ ,  $\beta/\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$  and  $\varphi$ ) (Fig 1A). According to the phylogenetic relationships with orthologs in *G. arboreum*, 20 *GrPLDs* were named as *GrPLD* $\alpha$ 1-*GrPLD* $\alpha$ 4, *GrPLD* $\beta$ 1-*GrPLD* $\beta$ 3, *GrPLD* $\gamma$ , *GrPLD* $\delta$ 1-*GrPLD* $\delta$ 5, *GrPLD* $\epsilon$ , *GrPLD* $\zeta$ 1-*GrPLD* $\zeta$ 4, *GrPLD* $\varphi$ 1 and *GrPLD* $\varphi$ 2 (Fig 1A and Table 1). Taking into account the gene location and orthologous relationship, we designated 40 *GhPLDs* as *GhPLD* $\alpha$ 1A/ $\alpha$ 1D-*GhPLD* $\alpha$ 4A/ $\alpha$ 4D, *GhPLD* $\beta$ 1A/ $\beta$ 1D-*GhPLD* $\beta$ 3A/ $\beta$ 3D, *GhPLD* $\gamma$ A/ $\gamma$ D, *GhPLD* $\delta$ 1A/ $\delta$ 1D-*GhPLD* $\delta$ 5A/ $\delta$ 5D, *GhPLD* $\epsilon$ A/ $\epsilon$ D, *GhPLD* $\zeta$ 1A/ $\zeta$ 1D-*GhPLD* $\zeta$ 4A/ $\zeta$ 4D, *GhPLD* $\varphi$ 1A/ $\varphi$ 1D and *GhPLD* $\varphi$ 2A/ $\varphi$ 2D (Fig 1 and Table 1). Every *GhPLD* had its own orthology in both diploid relatives, with the exception of *GhPLD* $\beta$ 3A. After attempting to amplify by specific primers and scanning the genome again, we still did not find a putative *GaPLD* $\beta$ 3. Thus, we deduced that it might have been lost during the evolution of *G. arboreum* after interspecific hybridization.

Among these cotton PLD subgroups, the *PLD* $\delta$  constituted the largest, containing 20 members. The second largest subgroups ( $\alpha$  and  $\zeta$ ) both consisted of 16 members. The subgroup  $\beta/\gamma$  was a unique bi-type subgroup which was comprised of 11 *PLD* $\beta$ s and four *PLD* $\gamma$ s. In the subgroups  $\varphi$  and  $\epsilon$ , eight and four members were identified, respectively (Fig 1A). Notably, subgroups  $\zeta$  and  $\varphi$  were far from the other four subgroups in evolutionary distance. Based on the functional domain prediction by searching the Pfam database, subgroups  $\zeta$  and  $\varphi$  belonged to PX/PH-PLD and SP-PLD subfamilies, respectively, whereas the other subgroups fell into the subfamily C2-PLD. Overall, according to the phylogenetic relationships of cotton PLDs, it was speculated that these multiple subgroups might play specialized roles in the adaptive evolution of cotton.

### Sequence characteristics of cotton PLD genes

To understand phylogenetic relationships, we calculated the sequence identities of pairwise cotton PLDs at both the nucleotide and amino acid level and also compared the exon-intron structures of individual PLDs. The genes that belong to the same subgroup had high identity to each other, especially for the ones with an orthologous relationship (S1 Fig). For different subgroups, the nearer evolutionary distance they were in, the higher sequence identities they had, such as the *PLD* $\beta/\gamma$  and *PLD* $\delta$  subgroups (Fig 1A and S1 Fig). Significantly, four gene clusters

**Table 1. The PLD genes in *G. hirsutum* and *G. raimondii* and properties of the deduced proteins.**

Gene name	Gene ID	Chromosomal location <sup>a</sup>	ORF(bp)	Protein <sup>b</sup>			
				Size (aa)	MW (kDa)	pI	Subcellular localization <sup>c</sup>
<i>GhPLDα1A</i>	Gh_A10G0662	A10(-): 11048194–11066260	2,424	807	91.53	5.37	C
<i>GhPLDα1D</i>	Gh_D10G0730	D10(+): 8439309–8443570	2,424	807	91.53	5.48	C
<i>GhPLDα2A</i>	Gh_A06G1624	A06(-): 101165820–101169178	2,418	805	91.6	5.5	C
<i>GhPLDα2D</i>	Gh_D06G1993	D06(-): 61526855–61530236	2,418	805	91.55	5.55	C
<i>GhPLDα3A</i>	Gh_A07G0222	A07(+): 2700343–2703900	2,484	827	93.99	6.65	C
<i>GhPLDα3D</i>	Gh_D07G0279	D07(+): 2893589–2897211	2,484	827	94.07	6.76	C
<i>GhPLDα4A</i>	Gh_A05G0436	A05(-): 4839584–4842566	2,454	817	92.9	6.71	C
<i>GhPLDα4D</i>	Gh_D05G0560	D05(+): 4511885–4514866	2,454	817	93.08	6.62	C
<i>GhPLDβ1A</i>	Gh_A11G0758	A11(+): 7513190–7519138	3,375	1,124	125.68	7.7	N
<i>GhPLDβ1D</i>	Gh_D11G0886	D11(+): 7677837–7683822	3,375	1,124	125.53	7.2	N
<i>GhPLDβ2A</i>	Gh_A12G0942	A12(+): 59715352–59718967	2,355	784	88.99	8.16	C
<i>GhPLDβ2D</i>	Gh_D12G1032	D12(+): 36125870–36130292	3,168	1,055	119.57	8.28	N
<i>GhPLDβ3A</i>	Gh_A11G0761	A11(+): 7538255–7544383	3,390	1,129	125.42	6.7	N
<i>GhPLDβ3D</i>	Gh_D11G0888	D11(+): 7700913–7707032	3,393	1,130	125.44	6.63	N
<i>GhPLDγ A</i>	Gh_A08G1600	A08(+): 94594467–94598013	2,556	851	95.54	8.19	C
<i>GhPLDγ D</i>	Gh_D08G1911	D08(+): 57054075–57057621	2,556	851	95.9	7.6	C
<i>GhPLDδ1A</i>	Gh_A12G0540	A12(-): 13280804–13286284	2,550	914	96.46	6.69	C
<i>GhPLDδ1D</i>	Gh_D12G0556	D12(-): 10278949–10284459	2,550	914	96.37	6.69	C
<i>GhPLDδ2A</i>	Gh_A02G1674	A02(+): 83053172–83057736	2,415	804	90.98	7.18	C
<i>GhPLDδ2D</i>	Gh_D03G0048	D03(+): 312090–316688	2,571	856	96.97	6.63	C
<i>GhPLDδ3A</i>	Gh_A01G0880	A01(-): 20697344–20701722	2,538	845	95.68	6.64	C
<i>GhPLDδ3D</i>	Gh_D01G0915	D01(-): 15285684–15289947	2,538	845	95.78	6.72	C
<i>GhPLDδ4A</i>	Gh_A05G2168	A05(-): 24844619–24854267	3,693	1,230	140.562	8.7	N
<i>GhPLDδ4D</i>	Gh_D05G2424	D05(-): 24298286–24302113	2,595	864	98.669	8.44	N
<i>GhPLDδ5A</i>	Gh_A12G0243	A12(-): 3722040–3726903	2,547	848	96.03	6.79	C
<i>GhPLDδ5D</i>	Gh_D12G0242	D12(-): 3218723–3223600	2,550	849	96.12	7.03	C
<i>GhPLDε A</i>	Gh_A02G1124	A02(-): 60940992–60943829	2,307	768	87.87	6.45	C
<i>GhPLDε D</i>	Gh_D03G0546	D03(-): 10422279–10425127	2,307	768	87.72	6.75	C
<i>GhPLDζ1A</i>	Gh_A01G0213	A01(+): 2083336–2091108	3,321	1,106	125.96	6.1	C
<i>GhPLDζ1D</i>	Gh_Sca005308G01	scaffold5308(+): 923–8640	3,321	1,106	126.01	6.1	C
<i>GhPLDζ2A</i>	Gh_A10G1486	A10(+): 81489677–81498158	3,345	1,114	126.71	6.33	N
<i>GhPLDζ2D</i>	Gh_D10G1729	D10(+): 48605329–48613791	3,345	1,114	126.49	6.31	N
<i>GhPLDζ3A</i>	Gh_A02G0494	A02(+): 7274495–7288655	2,919	972	109.9	6.28	N
<i>GhPLDζ3D</i>	Gh_D02G0554	D02(+): 7493511–7502060	2,946	981	111.07	6.28	N
<i>GhPLDζ4A</i>	Gh_A11G2458	A11(+): 83020505–83026075	3,327	1,108	126.32	5.57	C
<i>GhPLDζ4D</i>	Gh_D11G2775	D11(+): 57479136–57489254	3,192	1,063	121.3	5.56	C
<i>GhPLDφ1A</i>	Gh_A07G0265	A07(-): 3306792–3314618	1,545	514	57.91	7.23	E
<i>GhPLDφ1D</i>	Gh_D07G0321	D07(-): 3400982–3403213	1,545	514	57.94	7.99	E
<i>GhPLDφ2A</i>	Gh_A06G0922	A06(-): 37290132–37292468	1,380	459	52.17	8.98	PM
<i>GhPLDφ2D</i>	Gh_D06G1085	D06(-): 24321096–24323423	1,623	540	61.55	9.13	PM
<i>GrPLDα1</i>	Gorai.011G083000.1	Chr11(+): 8380923–8385184	2,424	807	91.58	5.44	C
<i>GrPLDα2</i>	Gorai.010G225600.1	Chr10(-): 59751136–59754500	2,418	805	91.51	5.63	C
<i>GrPLDα3</i>	Gorai.001G033300.1	Chr1(+): 3069538–3073522	2,484	827	93.88	6.33	C
<i>GrPLDα4</i>	Gorai.009G057800.1	Chr9(+): 4158704–4161686	2,454	817	93.14	6.55	C
<i>GrPLDβ1</i>	Gorai.007G094300.1	Chr7(+): 6927130–6933259	3,393	1130	125.61	6.73	N
<i>GrPLDβ2</i>	Gorai.008G115700.1	Chr8(+): 34857730–34862153	3,162	1053	119.23	7.59	N

(Continued)

Table 1. (Continued)

Gene name	Gene ID	Chromosomal location <sup>a</sup>	ORF(bp)	Protein <sup>b</sup>			
				Size (aa)	MW (kDa)	pI	Subcellular localization <sup>c</sup>
<i>GrPLDβ3</i>	Gorai.007G094100.1	Chr7(+):6908403–6914399	3,375	1124	125.64	7.51	N
<i>GrPLDγ</i>	Gorai.004G206700.1	Chr4(+):53722628–53726174	2,556	851	95.93	7.6	C
<i>GrPLDδ1</i>	Gorai.008G061600.1	Chr8(-):9809029–9814536	2,550	849	96.25	6.69	C
<i>GrPLDδ2</i>	Gorai.003G005500.1	Chr3(-):302357–306946	2,565	854	96.8	6.55	C
<i>GrPLDδ3</i>	Gorai.002G120800.1	Chr2(-):17310003–17314270	2,562	853	96.72	6.59	C
<i>GrPLDδ4</i>	Gorai.009G268500.1	Chr9(-):22371067–22374881	2,523	840	95.74	8.6	PM
<i>GrPLDδ5</i>	Gorai.008G027300.1	Chr8(-):3218926–3223804	2,550	849	96.14	6.91	C
<i>GrPLDε</i>	Gorai.003G059700.1	Chr3(-):10351275–10354153	2,307	768	87.6	6.67	C
<i>GrPLDζ1</i>	Gorai.002G028800.1	Chr2(+):2142619–2150335	3,321	1106	125.91	6.21	C
<i>GrPLDζ2</i>	Gorai.011G193900.1	Chr11(+):46880104–46888580	3,300	1099	124.6	6.15	C
<i>GrPLDζ3</i>	Gorai.005G062100.1	Chr5(+):6558010–6566183	3,273	1090	124.04	6.18	N
<i>GrPLDζ4</i>	Gorai.007G302800.1	Chr7(+):51639889–51645452	3,237	1078	123.16	5.58	C
<i>GrPLDφ1</i>	Gorai.001G038100.1	Chr1(-):3551503–3553734	1,545	514	58	7.19	E
<i>GrPLDφ2</i>	Gorai.010G118800.1	Chr10(-):23679895–23682222	1,545	514	58.28	8.96	E

a. Chromosomal location: [ChrNo.(Orientation): start-end], '+' and '-' indicated the forward and reverse orientation, respectively.

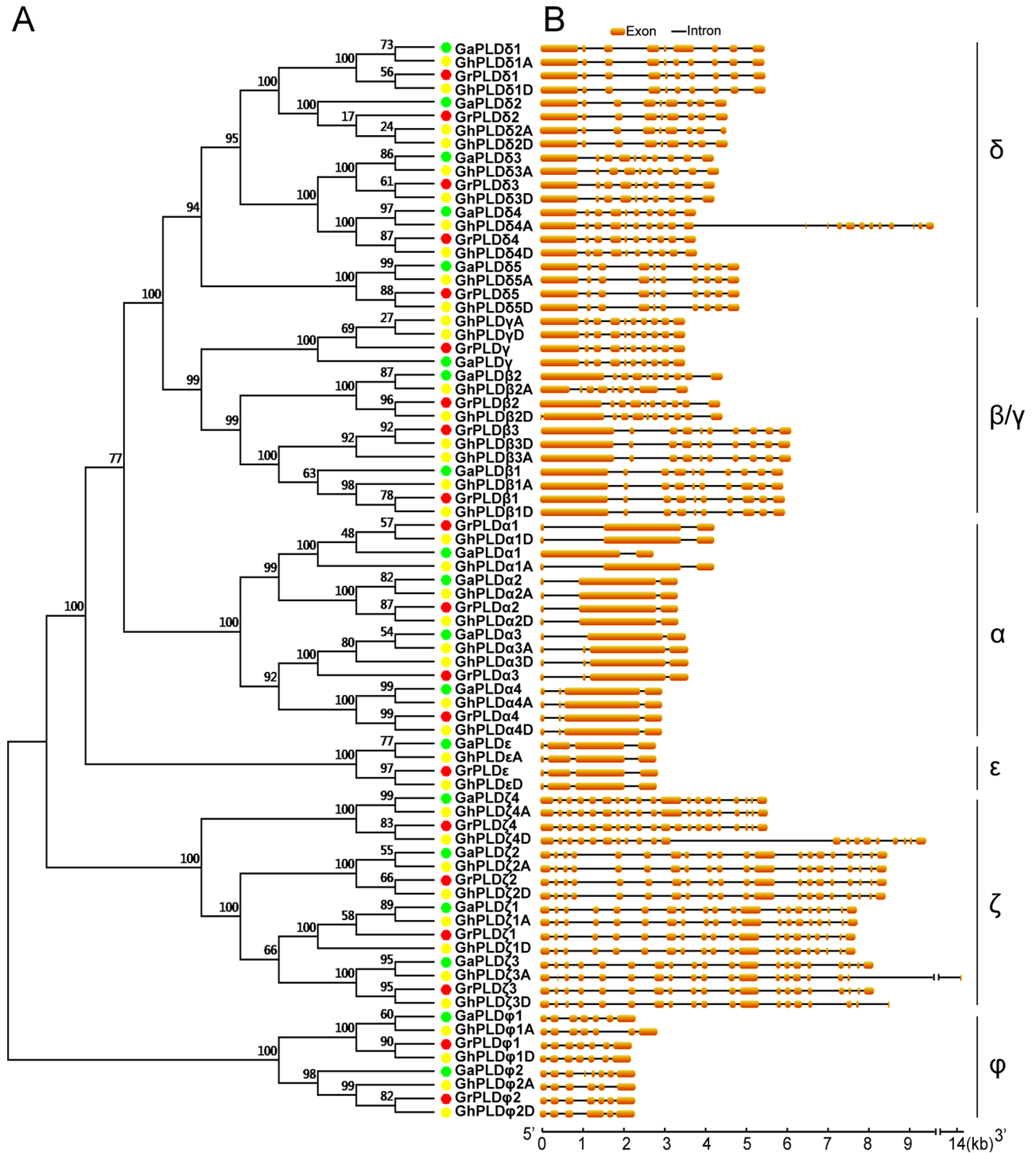
b. The theoretical molecular weight (MW) and isoelectric point (pI) were calculated by ExPASy. Subcellular localization was analyzed using the CELLO v2.5 server. The Nucleic acid and deduced amino acid sequences of each *GhPLD* and *GrPLD* gene are listed in [S2 Table](#).

c. C, cytoplasm; N, nucleus; PM, plasma membrane; E, extracellular space.

doi:10.1371/journal.pone.0156281.t001

(*PLDβ1s-PLDβ2s-PLDβ3s*, *PLDδ1s-PLDδ2s*, *PLDδ3s-PLDδ4s*, and *PLDφ1s-PLDφ2s*) with more than 90% identity existed in three subgroups, indicating that they might originate from gene duplication events ([S1 Fig](#)). The detailed illustration showed the distribution and position of introns within each of the cotton PLD genes ([Fig 1B](#)). A total of 680 introns were found in the cotton PLD genes, with an average intron number of 8.6 per gene and an average intron length of 258.9 bp ([S3 Table](#)). The members of the individual subgroups shared similar intron numbers, consistent with the phylogenetic classification of the cotton PLDs ([Fig 1](#) and [S3 Table](#)). For example, both subgroups *PLDβ/γ* and *PLDδ* included members with approximately nine introns.

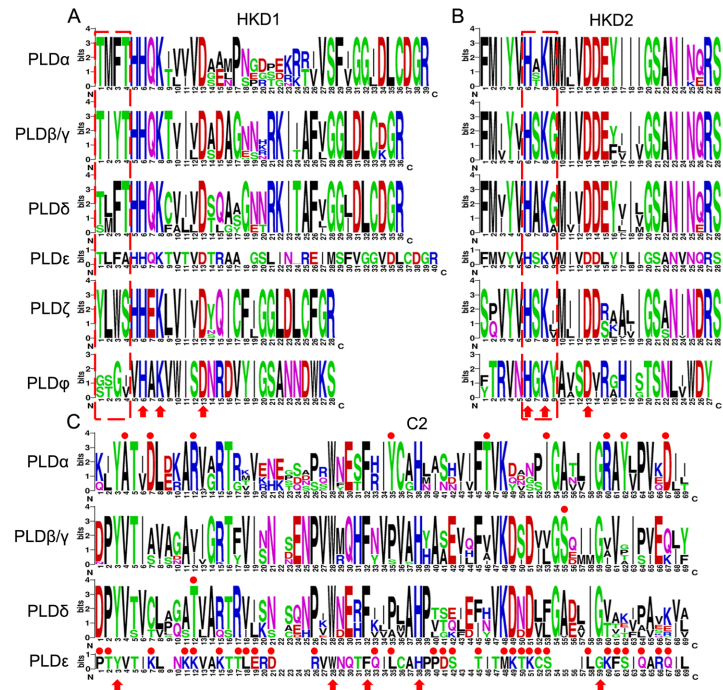
To reveal the typical domain characteristics of cotton PLD subgroups, comparative analyses for the conservation of amino acid residues in functional domains were performed on the basis of the alignments of these PLDs. The HKD1 domains were relatively more diverse than the HKD2 domains ([Fig 2A and 2B](#)). Both of the HKD1 and HKD2 domains contained three highly conserved amino acids (6H, 8K and 13D), which might have a key functional importance within these two domains ([Fig 2A and 2B](#)). The members of the individual subgroups possessed some marker amino acids in the HKD1 and HKD2 domains, for instance, "TMFT" and "H[A/ST]KM" in *PLDαs*, "TIYT" and "HSGK" in *PLDβ/γs*, "[T/S][L/M]FT" and "HAK [G/A]" in *PLDδs*, "TLFA" and "HSKV" in *PLDεs*, "YLWS" and "HSK[I/L/V]" in *PLDζs*, and "[G/S][S/T]G[IV]" and "HGKY" in *PLDφs* ([Fig 2A and 2B](#)). Similarly, the members of individual cotton PLD subfamilies also contained some specific amino acids, for example, "HHQK" in C2-PLDs, "HHEK" in PX/PH-PLDs, and "VHAK" in SP-PLDs ([Fig 2A](#)). For C2 domain sequences, five amino acids (3Y, 28W, 32F, 38H and 59G) are highly conserved, while some other amino acids were extremely specific for their own PLD subgroup, such as 4A, 7D, 12R, 35Y, 46T, 53I, 60R, 62Y and 67D for *PLDα* ([Fig 2C](#)).



**Fig 1. Phylogenetic relationship and gene structure of the cotton PLD genes.** A. The phylogenetic tree was conducted using MEGA 6.0 software with the NJ method with bootstrapping analysis (1,000 replicates). The numbers beside the branches indicate the bootstrap values that support the adjacent nodes. Dots of different colors represented the different cotton species (Green, *G. arboreum*; red, *G. raimondii*; yellow, *G. hirsutum*). B. Schematic diagram for the exon-intron organization of cotton PLD genes. The orange boxes and black lines indicate the exons and introns, respectively.

doi:10.1371/journal.pone.0156281.g001





**Fig 2. The conservation of amino acid residues in two HKDs and C2 domains were presented in different PLD subgroups.** Numbers on the x-axis represent the sequence positions of each domain. The y-axis represents the information content measured in bits. A. Alignment of the HKD1 domains; B. Alignment of the HKD2 domains; C. Alignment of the C2 domains.

doi:10.1371/journal.pone.0156281.g002

### Gene duplication events in GhPLD and GrPLD gene families

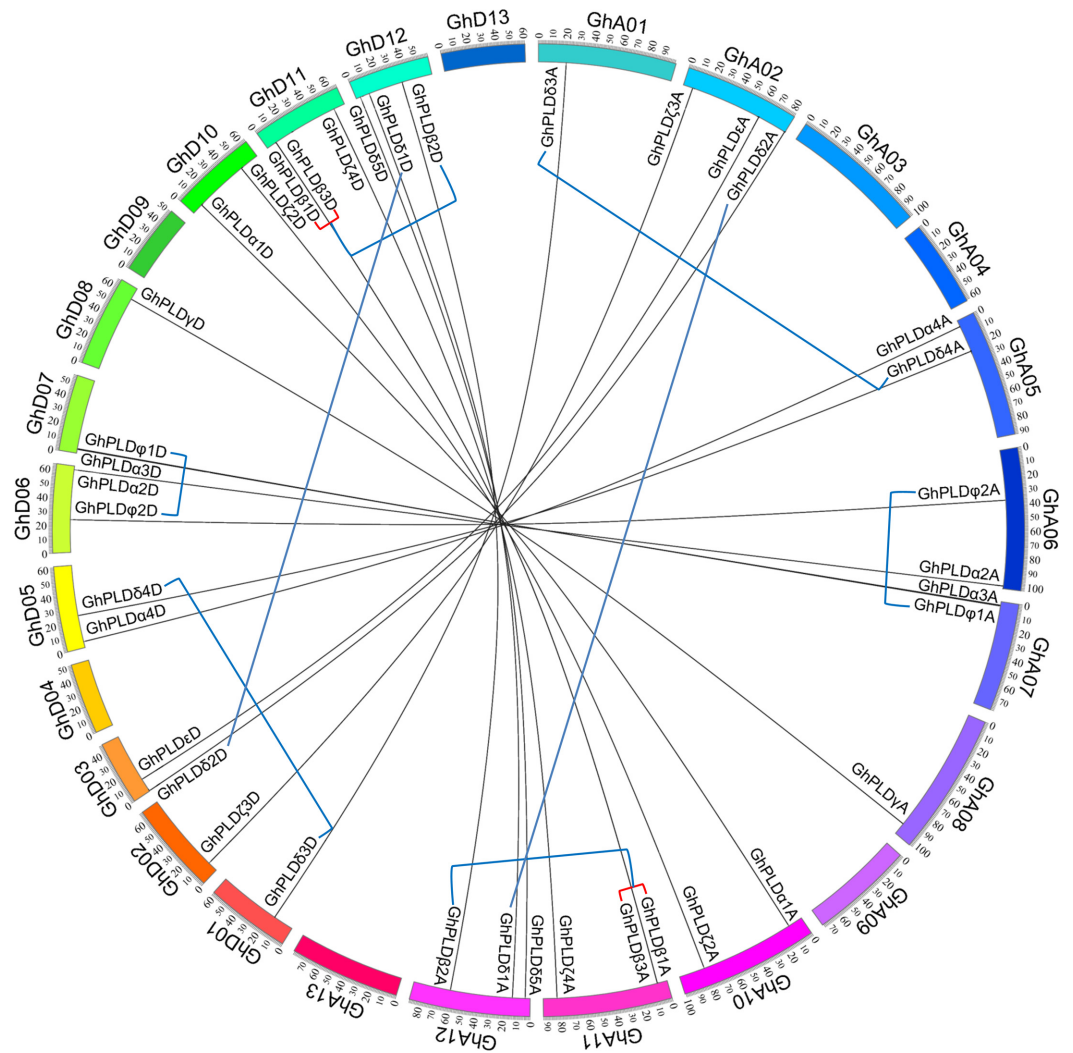
After mapping all cotton PLD genes to their corresponding chromosomes, we found that they were unevenly distributed on the chromosomes of three cotton species (Fig 3 and S2 Fig). For the *GhPLD* gene family, 39 out of 40 PLD genes were assigned to 19 of the 26 *G. hirsutum* chromosomes (none were assigned to A03, A04, A09, A13, D04, D09 or D13), while one gene *GhPLDζ1D* was assigned to an as of yet unmapped scaffold (Fig 3). For the *GrPLD* gene family, all 20 PLD genes were assigned to 10 of the 13 *G. raimondii* chromosomes (none were assigned to Gr6, Gr12 or Gr13) (S2 Fig). We further assessed the contribution of gene duplication to the expansion of these two cotton PLD gene families. In *G. hirsutum*, eight segmental duplications (*GhPLDβ1A/β2A*, *GhPLDδ1A/δ2A*, *GhPLDδ3A/δ4A*, *GhPLDφ1A/φ2A*, *GhPLDβ1D/β2D*, *GhPLDδ1D/δ2D*, *GhPLDδ3D/δ4D* and *GhPLDφ1D/φ2D*) and two tandem duplications (*GhPLDβ1A/β3A* and *GhPLDβ1D/β3D*) occurred from 17.88 to 28.01 mya (Fig 3 and S4 Table). In *G. raimondii*, four segmental duplication (*GrPLDβ1/β2*, *GrPLDδ1/δ2*, *GrPLDδ3/δ4* and *GrPLDφ1/φ2*) and one tandem duplication (*GrPLDβ1/β3*) jointly took place during the time of 18.06~21.93 mya (Fig 3 and S4 Table). All identified gene duplication events in the three cotton PLD families happened before the split of two diploid cottons. Thus, we speculate that an unidentified tandem duplication event also happened previously in *G. arboreum*, followed by a specific gene loss event of *GaPLDβ3*.

To investigate which type of selection pressure had been involved in the divergence after gene duplication events, we calculated the *Ka/Ks* ratios (synonymous substitutions to non-synonymous substitutions) for the duplicated cotton PLD gene pairs on the basis of coding sequences. The resulting pairwise comparison data showed that all of the paralogous genes had *Ka/Ks* ratios of < 1, suggesting that cotton duplicated PLD genes had experienced strong

purifying selection pressure. Strong functional constraints had a bearing on the evolution of these gene families, reflecting the essential roles of PLDs in cotton.

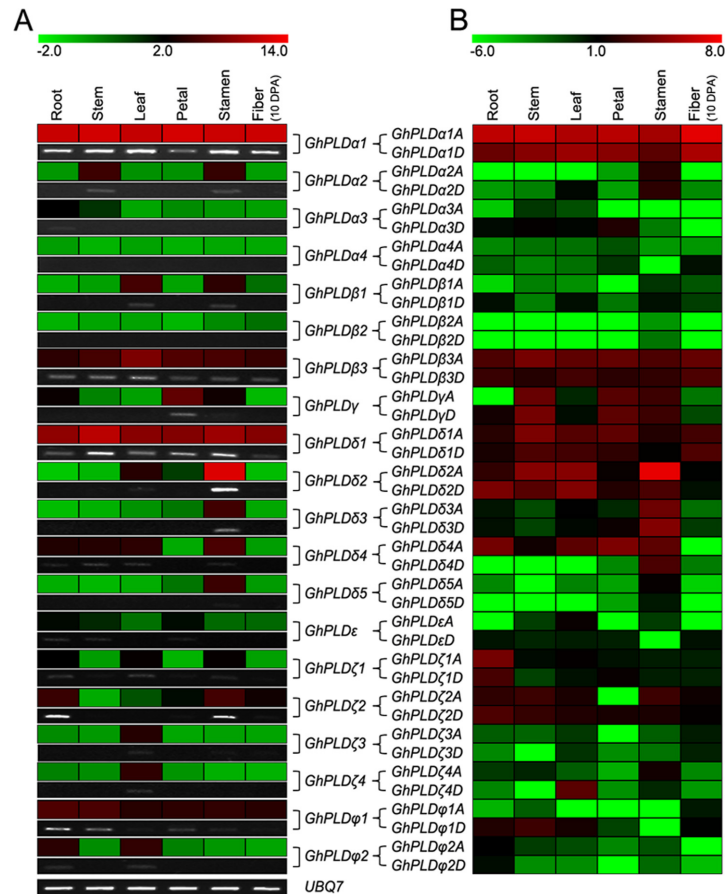
### Spatial expression profiles of GhPLD genes

To better reveal the potential functions of PLDs in allotetraploid cotton, the expression profiles of *GhPLDs* were investigated by both quantitative RT-PCR and RNA-seq data analysis. Because of the extremely high similarity between the mRNAs of the *GhPLDxA-GhPLDxD* gene pairs (such as *GhPLDα1A* and *GhPLDα1D*) and their nearly identical transcript sizes, we could not distinguish them using quantitative RT-PCR. Therefore, we regarded *GhPLDxA-GhPLDxD* as one combination named *GhPLDx* and investigated its expression level by quantitative RT-PCR, and then distinguished the portion of *GhPLDα1A* and that of *GhPLDα1D* by analyzing RNA-seq data, which included both gene expression levels and the information of gene location. Alternatively, the expression profiles of *GhPLDs* obtained from qRT-PCR and semi-quantitative PCR, were confirmed by RNA-seq data analysis.



**Fig 3. Syntenic relationship of the PLD genes in *G. hirsutum*.** The syntenic relationships between *GhPLDxA*s and *GhPLDxD*s according to the CottonGen database are illustrated using the program Circos. Tandem and segmental duplicated PLD genes are connected by red and blue lines. The chromosomes of *G. hirsutum* are designated as GhA01—GhA13 and GhD01—GhD13.

doi:10.1371/journal.pone.0156281.g003



**Fig 4. Expression patterns of *GhPLDs* in different cotton tissues.** Expression patterns of *GhPLDs* in roots, stems, leaves, petals, stamens, and 10 DPA fibers. A. Real-time quantitative RT-PCR and semi-quantitative PCR results. B. RNA-seq data analysis results. The Illumina reads for expression analysis in cotton roots, stems, leaves, petals, stamens and 10 DPA fibers are retrieved from the NCBI SRA database. The color scale at the top of the left column heat map indicates the relative expression levels where light green indicates low and red indicates high. The color scale at the top of the right column heat map represents the FPKM values normalized log<sub>2</sub> transformed counts where light green indicates low and red indicates high.

doi:10.1371/journal.pone.0156281.g004

We found that *GhPLD* genes have rather broad expression patterns across a variety of cotton tissues under normal growth conditions including roots, stems, leaves, petals, stamens and fibers. The gene combination from the PLD $\alpha$  subgroup, *GhPLD $\alpha$ 1*, had the most preferential expression levels in roots, stems, leaves and fibers and maintained high levels of expression in petals and stamens (Fig 4A). *GhPLD $\delta$ 2*, with the highest expression abundance only in stamens, belongs to the largest subgroup (PLD $\delta$ ). Another gene combination from this subgroup, *GhPLD $\delta$ 1*, showed the second highest expression level in almost all cotton tissues, and *GhPLD $\beta$ 3*, which belongs to the PLD $\beta$ / $\gamma$  subgroup, expressed the third highest levels (Fig 4A). Similarly, *GhPLD $\zeta$ 2* and *GhPLD $\phi$ 1*, which were from the PLD $\zeta$  and PLD $\phi$  subgroups, respectively, were expressed at moderate levels, while the only gene combination of the remaining subgroup, *GhPLD $\epsilon$* , showed low expression values in all tissues (Fig 4A). Other low expression genes included *GhPLD $\alpha$ 2*, *GhPLD $\beta$ 1*, *GhPLD $\gamma$* , *GhPLD $\delta$ 3*, *GhPLD $\delta$ 4*, *GhPLD $\zeta$ 1*, *GhPLD $\zeta$ 3*, *GhPLD $\zeta$ 4* and *GhPLD $\phi$ 2*. However, *GhPLD $\alpha$ 3*, *GhPLD $\alpha$ 4* and *GhPLD $\beta$ 2* had only weak expression or even undetected expression (Fig 4A). Additionally, the duplicated *GhPLD* gene pairs displayed variant expression patterns. For instance, *GhPLD $\beta$ 2* had almost undetected

expression in all tissues, while *GhPLDβ1* maintained obvious expression levels in leaves and stamens. Examining when and where a gene is expressed in the plant tissues/organs can often lead to clues about gene function. Therefore, these diverse expression patterns of *GhPLDs* allude to functional diversification of this gene family in allotetraploid cotton.

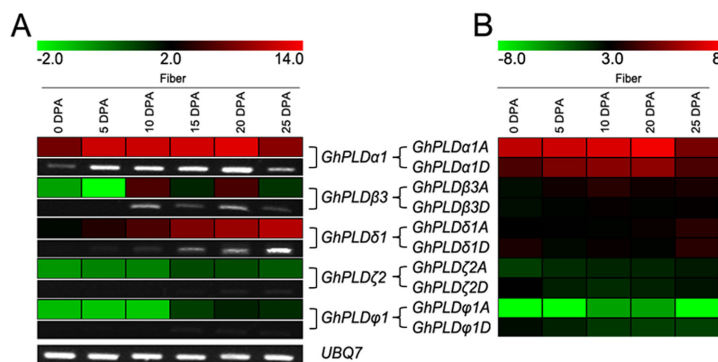
To gain deeper insights into the expression patterns, we analyzed a RNA-seq dataset that encompassed results from six studied cotton tissues. In these surveyed gene combinations, we successfully distinguished the contribution of *GhPLDxAs* and that of *GhPLDxDs*. Most *GhPLDxAs* were more preferentially expressed than *GhPLDxDs* (Fig 4B). For instance, *GhPLDα1A* was expressed more preferentially than *GhPLDα1D* in all studied cotton tissues. However, the opposite correlation was observed for *GhPLDφ1D*, which maintained higher expression levels than *GhPLDφ1A*. Overall, the results of RNA-seq expression data were in close agreement with that of quantitative RT-PCR (Fig 4B).

### Expression patterns of GhPLD genes in developing fibers

It is worth noting that the *GhPLDs* that were expressed in fibers, which is the most valuable economic tissue in cotton, were *GhPLDα1A/D*, *GhPLDβ3A/D*, *GhPLDδ1A/D*, *GhPLDζ2A/D* and *GhPLDφ1A/D*. To survey the expression profiles of these fiber expressed PLD genes, quantitative RT-PCR and RNA-seq data analysis were performed at the representative stages of fiber development (0~25 DPA). Strikingly, among these selected genes, *GhPLDα1* had high expression in elongating fibers. Specifically, expression increased from 0 to 20 DPA, peaking at 20 DPA, and then decreased substantially (Fig 5A). Furthermore, *GhPLDα1A* made more of a contribution than *GhPLDα1D* did at every studied stages of fiber elongation (Fig 5B). These results implied that *GhPLDα1A* and *GhPLDα1D* might jointly play an important role in fiber development, most likely in the initiation stage of secondary cell wall thickening in fiber. In addition, *GhPLDδ1A* and *GhPLDδ1D* may also have an essential role in fiber development around the time of 25 DPA (Fig 5). However, *GhPLDβ3A*, *GhPLDβ3D*, *GhPLDζ2A*, *GhPLDζ2D*, *GhPLDφ1A* and *GhPLDφ1D* had relatively low or even undetectable expression levels in elongating fibers (Fig 5).

### Cis-regulatory elements in the promoters of GhPLD genes

To obtain more insights into the expression patterns and putative functions, the *cis*-regulatory elements were scanned in the promoter regions of *GhPLDs*. According to the expression levels,



**Fig 5. Expression patterns of *GhPLDs* in developing fibers.** Expression patterns of *GhPLDs* were analyzed in cotton fibers at different developmental stages (0, 5, 10, 15, 20, and 25 DPA). A. Real-time quantitative RT-PCR and semi-quantitative PCR results. B. RNA-seq data analysis results. The Illumina reads for expression analysis in developing fibers (0, 5, 10, 20 and 25 dpa) are retrieved from the NCBI SRA database. The color scale at the top of the left column heat map indicates the relative expression levels where light green indicates low and red indicates high. The color scale at the top of the right column heat map represents the FPKM values normalized log<sub>2</sub> transformed counts where light green indicates low and red indicates high.

doi:10.1371/journal.pone.0156281.g005



they might originate from the common ancestor I (Fig 1 and S3 Fig). PLD $\alpha$ s and PLD $\epsilon$ s possessed 2 or 3 introns, suggesting that they originated from the common ancestor II (Fig 1 and S3 Fig). In addition, of these PLDs, PLD $\zeta$ s and PLD $\phi$ s belonged to the PX/PH-PLD and SP-PLD subfamilies, respectively, and had dissimilar intron numbers, suggesting the convergent evolution via two independent evolutionary paths (Fig 1 and S3 Fig). Despite specific features of the functional domains near the N-terminus, the sequence identities of the C-terminus were relatively conserved, especially near the two characteristic HKD domains [47] (Fig 2). Thus, we reasonably speculated that all plant PLDs might have evolved from one original ancestor followed by some unknown changes that took place near the N-terminus, dividing the plant PLD gene family into three distinct PLD subfamilies: C2-PLD, PX/PH-PLD and SP-PLD (S3 Fig).

In the long history of evolution, selection pressure after gene duplication events shaped gene families, resulting in distinct evolutionary patterns among different gene families and even different subgroups in one gene family [48]. The members of the subgroup PLD $\phi$  were believed to be more conserved than those of the subgroups PLD $\beta/\gamma$  and PLD $\delta$ , as they had higher degrees of sequence identity and more similar exon-intron structures than PLD $\beta/\gamma$ s and PLD $\delta$ s (Fig 1 and S1 Fig). This was consistent with our analysis on the rate of molecular evolution, in which the mean values of *Ka/Ks* ratios in the subgroup PLD $\phi$  were smaller than those in the PLD $\beta/\gamma$  and PLD $\delta$  subgroups (S5 Table).

## Functional diversification of allotetraploid cotton PLD genes

Increasing evidence suggests the PLDs are involved in many plant processes including plant development, responses to biotic and abiotic stresses. *Cis*-regulatory elements in gene promoter regions have essential roles in determining tissue-specific and stress-responsive expression patterns of genes [49], which might help to elucidate transcriptional regulation and potential functions of *GhPLDs*. Phytohormones are important regulators of plant growth and development. Cotton fiber development is known to be regulated by gibberellic acid (GA) [50], jasmonic acid (JA) [51], abscisic acid (ABA) [52], ethylene [53] and auxin [54]. Among the 10 *GhPLD* genes expressed in elongating allotetraploid cotton fibers, 5, 4, 4, 2 and 1 genes were found to have *cis*-regulatory elements (P-box, GARE, TGACG-motif, ABRE, ERE and TGA-element) specific for GA, JA, ABA, ethylene and auxin responses in their respective promoters (S6 Table and Fig 6). Three preferentially expressed genes in fiber, including *GhPLD $\alpha$ 1A*, *GhPLD $\delta$ 1A* and *GhPLD $\delta$ 1D*, all had more than three types of hormone response elements, strongly suggesting that the functions of these *GhPLDs* in elongating cotton fibers were synergistically regulated by different phytohormones. Other tissue-specific *GhPLDs*, such as *GhPLD $\gamma$ A*, *GhPLD $\gamma$ D*, *GhPLD $\delta$ 4A* and *GhPLD $\delta$ 4D*, possessed 2~4 types of hormone response elements, indicating that multiple phytohormones also regulated the transcription of *GhPLDs* in other cotton tissues.

Stress response elements related to diverse environmental stimuli were found in the promoter regions of allotetraploid cotton PLD genes (S6 Table and Fig 6). For the vast majority of *GhPLDs*, except for the preferentially expressed genes, the distribution of elements for stress responsiveness was more extensive, explaining why most of the *GhPLDs* exhibited the relatively low or weak expression levels under normal growth conditions (Figs 4 and 6). Adverse environmental stimulus, such as drought, defense, and heat shock, might trigger high levels of transcription of most low or weak expressed *GhPLDs*, as a result of that a number of the specific elements (MBS, TC-rich and HSE) were located in the promoters of these genes (Fig 6C and 6D). Therefore, we hypothesize that *GhPLDs* might also have important functions in the adaptation to adverse environmental stimuli. Overall, complex and diverse transcriptional regulation significantly broadened our view and understanding of the functional diversification of allotetraploid cotton PLDs.

To date, the biological and cellular functions of GhPLD genes remain largely unknown. The current investigation demonstrates some of GhPLD genes that might be involved in cotton development and stress response, and provides important clues for the selection of candidate genes, especially *GhPLD $\alpha$ 1A/D* and *GhPLD $\delta$ 1A/D* for the further studies.

## Conclusions

In conclusion, a total of 40 and 20 PLD genes were identified in *G. hirsutum* and *G. raimondii*, respectively. Our comparative analyses provided valuable insight into the understanding of phylogenetic relationships, sequence characteristics, molecular evolution of PLD genes in allotetraploid cotton and its two diploid progenitors. Moreover, we also characterized the broad spatial and fiber developmental expression profiles of allotetraploid cotton PLDs and expanded the view of transcriptional regulation of these genes. Unveiling the roles of PLD genes in cotton growth, development and stress adaptation processes may facilitate advances in crop variety development and utilization.

## Supporting Information

**S1 Fig. Sequence identity of the cotton PLD genes.** The sequence identities of cotton PLDs at both the nucleotide and amino acid level are calculated with the program DNASTAR. The color scale at the top of the heat map indicates the levels of the sequence identities where light green indicates low and red indicates high. The data at the diagonal lines are equal to 100%. (TIF)

**S2 Fig. Syntenic relationship of the PLD genes in *G. arboreum* and *G. raimondii*.** The syntenic relationships between *GaPLDs* and *GrPLDs* according to the CottonGen and the Phytosome v9.1 database are illustrated using the program Circos. Tandem and segmental duplicated PLD genes are connected by red and blue lines. The chromosomes of *G. arboreum* and *G. raimondii* are designated as Ga1—Ga13 and Gr1—Gr13, respectively. (TIF)

**S3 Fig. The expansion of the allotetraploid cotton PLD family.** The left column represents the phylogenetic relationships of allotetraploid cotton PLDs constructed using the NJ method. Letters in the circles indicate the identified gene duplication event (TD, tandem duplication; SD, segmental duplication). The middle column is the representative exon-intron organization of the PLD subgroup. PLDxA and PLDxD represent *G. hirsutum* PLDxA and *G. hirsutum* PLDxD genes, respectively. The right column indicates the putative common ancestors of allotetraploid cotton PLD subgroups. (TIF)

**S1 Table. Primers used for qRT-PCR and semi-quantitative PCR.** (XLSX)

**S2 Table. Nucleic acid, deduced amino acid, promoter and genomic sequences of *GhPLD* and *GrPLD* genes.** (XLSX)

**S3 Table. Intron number and average length of cotton PLD genes in each subgroup.** (XLSX)

**S4 Table. Duplicated *GhPLDs* and *GrPLDs* and the numbers of conserved protein-coding genes flanking them.** Abbreviation: *Ks*-synonymous substitution rates; SD *Ks*-Standard

deviation *Ks*; Mini *Ks*-Minimum *Ks*; Max *Ks*-Maximum *Ks*; mya-million years ago.  
(XLSX)

**S5 Table. The *Ka/Ks* ratios for duplicate PLD genes in *G. arboreum*, *G. raimondii*, and *G. hirsutum*.**  
(XLSX)

**S6 Table. Putative *cis*-element sequences in promoter regions of cotton PLD genes.** \*N—A, C, G or T; K—G or T; M—A or C; R—A or G; W—A or T; Y—C or T  
(XLSX)

## Acknowledgments

The authors would like to acknowledge members of the Laboratory of Molecular Biology at Tsinghua University for critical discussions. 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 and 2011ZX08009-003).

## Author Contributions

Conceived and designed the experiments: JYL. Performed the experiments: KT. Analyzed the data: KT. Contributed reagents/materials/analysis tools: KT. Wrote the paper: KT CJD JYL.

## References

1. Zhu YX, Li FG. The *Gossypium raimondii* genome, a huge leap forward in cotton genomics. *Journal of integrative plant biology*. 2013; 55(7):570–1. Epub 2013/05/31. doi: [10.1111/jipb.12076](https://doi.org/10.1111/jipb.12076) PMID: [23718577](https://pubmed.ncbi.nlm.nih.gov/23718577/).
2. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant physiology*. 2007; 145(4):1303–10. Epub 2007/12/07. doi: [10.1104/pp.107.107672](https://doi.org/10.1104/pp.107.107672) PMID: [18056866](https://pubmed.ncbi.nlm.nih.gov/18056866/); PubMed Central PMCID: PMC2151711.
3. Wendel JF. New World tetraploid cottons contain Old World cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*. 1989; 86(11):4132–6. Epub 1989/06/01. PMID: [16594050](https://pubmed.ncbi.nlm.nih.gov/16594050/); PubMed Central PMCID: PMC287403.
4. Wendel JF, Albert VA. Phylogenetics of the Cotton Genus (*Gossypium*)—Character-State Weighted Parsimony Analysis of Chloroplast-DNA Restriction Site Data and Its Systematic and Biogeographic Implications. *Syst Bot*. 1992; 17(1):115–43. doi: [10.2307/2419069](https://doi.org/10.2307/2419069) PMID: [151A1992HB45800011](https://pubmed.ncbi.nlm.nih.gov/151A1992HB45800011/).
5. Jiang C, Wright RJ, El-Zik KM, Paterson AH. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95(8):4419–24. Epub 1998/04/29. PMID: [9539752](https://pubmed.ncbi.nlm.nih.gov/9539752/); PubMed Central PMCID: PMC22504.
6. Cao X. Whole genome sequencing of cotton—a new chapter in cotton genomics. *Science China Life sciences*. 2015; 58(5):515–6. Epub 2015/05/02. doi: [10.1007/s11427-015-4862-z](https://doi.org/10.1007/s11427-015-4862-z) PMID: [25929972](https://pubmed.ncbi.nlm.nih.gov/25929972/).
7. Wang G, Ryu S, Wang X. Plant phospholipases: an overview. *Methods Mol Biol*. 2012; 861:123–37. Epub 2012/03/20. doi: [10.1007/978-1-61779-600-5\\_8](https://doi.org/10.1007/978-1-61779-600-5_8) PMID: [22426716](https://pubmed.ncbi.nlm.nih.gov/22426716/).
8. Qin C, Wang X. The Arabidopsis phospholipase D family. Characterization of a calcium-independent and phosphatidylcholine-selective PLD zeta 1 with distinct regulatory domains. *Plant physiology*. 2002; 128(3):1057–68. Epub 2002/03/14. doi: [10.1104/pp.010928](https://doi.org/10.1104/pp.010928) PMID: [11891260](https://pubmed.ncbi.nlm.nih.gov/11891260/); PubMed Central PMCID: PMC152217.
9. Li G, Lin F, Xue HW. Genome-wide analysis of the phospholipase D family in *Oryza sativa* and functional characterization of PLD beta 1 in seed germination. *Cell research*. 2007; 17(10):881–94. Epub 2007/09/19. doi: [10.1038/cr.2007.77](https://doi.org/10.1038/cr.2007.77) PMID: [17876344](https://pubmed.ncbi.nlm.nih.gov/17876344/).
10. Liu Q, Zhang C, Yang Y, Hu X. Genome-wide and molecular evolution analyses of the phospholipase D gene family in Poplar and Grape. *BMC plant biology*. 2010; 10:117. Epub 2010/06/23. doi: [10.1186/1471-2229-10-117](https://doi.org/10.1186/1471-2229-10-117) PMID: [20565843](https://pubmed.ncbi.nlm.nih.gov/20565843/); PubMed Central PMCID: PMC3095279.



11. Zhao J, Zhou D, Zhang Q, Zhang W. Genomic analysis of phospholipase D family and characterization of GmPLDalphas in soybean (*Glycine max*). *Journal of plant research*. 2012; 125(4):569–78. Epub 2011/12/14. doi: [10.1007/s10265-011-0468-0](https://doi.org/10.1007/s10265-011-0468-0) PMID: [22161123](https://pubmed.ncbi.nlm.nih.gov/22161123/).
12. Li MY, Hong YY, Wang XM. Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Bba-Mol Cell Biol L*. 2009; 1791(9):927–35. doi: [10.1016/j.bbali.2009.02.017](https://doi.org/10.1016/j.bbali.2009.02.017) PMID: [18900014](https://pubmed.ncbi.nlm.nih.gov/18900014/).
13. Devaiah SP, Pan X, Hong Y, Roth M, Welti R, Wang X. Enhancing seed quality and viability by suppressing phospholipase D in *Arabidopsis*. *The Plant journal: for cell and molecular biology*. 2007; 50(6):950–7. Epub 2007/06/15. doi: [10.1111/j.1365-3113X.2007.03103.x](https://doi.org/10.1111/j.1365-3113X.2007.03103.x) PMID: [17565616](https://pubmed.ncbi.nlm.nih.gov/17565616/).
14. Kabachevskaya AM, Liakhnovich GV, Kisel MA, Volotovskii ID. Red/far-red light modulates phospholipase D activity in oat seedlings: relation of enzyme photosensitivity to photosynthesis. *Journal of plant physiology*. 2007; 164(1):108–10. Epub 2006/04/20. doi: [10.1016/j.jplph.2006.02.008](https://doi.org/10.1016/j.jplph.2006.02.008) PMID: [16621133](https://pubmed.ncbi.nlm.nih.gov/16621133/).
15. Potocky M, Elias M, Profotova B, Novotna Z, Valentova O, Zarsky V. Phosphatidic acid produced by phospholipase D is required for tobacco pollen tube growth. *Planta*. 2003; 217(1):122–30. doi: [10.1007/s00425-002-0965-4](https://doi.org/10.1007/s00425-002-0965-4) PMID: [123386800014](https://pubmed.ncbi.nlm.nih.gov/123386800014/).
16. Ohashi Y, Oka A, Rodrigues-Pousada R, Possenti M, Ruberti I, Morelli G, et al. Modulation of phospholipid signaling by GLABRA2 in root-hair pattern formation. *Science*. 2003; 300(5624):1427–30. Epub 2003/05/31. doi: [10.1126/science.1083695](https://doi.org/10.1126/science.1083695) PMID: [12775839](https://pubmed.ncbi.nlm.nih.gov/12775839/).
17. Li G, Xue HW. *Arabidopsis* PLD zeta 2 regulates vesicle trafficking and is required for auxin response. *Plant Cell*. 2007; 19(1):281–95. doi: [10.1105/tpc.106.041426](https://doi.org/10.1105/tpc.106.041426) PMID: [1500244757400024](https://pubmed.ncbi.nlm.nih.gov/1500244757400024/).
18. Fan L, Zheng S, Wang X. Antisense suppression of phospholipase D alpha retards abscisic acid- and ethylene-promoted senescence of postharvest *Arabidopsis* leaves. *Plant Cell*. 1997; 9(12):2183–96. Epub 1998/01/23. PMID: [9437863](https://pubmed.ncbi.nlm.nih.gov/9437863/); PubMed Central PMCID: [PMC157067](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC157067/).
19. Mane SP, Vasquez-Robinet C, Sioson AA, Heath LS, Grene R. Early PLDalpha-mediated events in response to progressive drought stress in *Arabidopsis*: a transcriptome analysis. *Journal of experimental botany*. 2007; 58(2):241–52. Epub 2007/01/31. doi: [10.1093/jxb/erl262](https://doi.org/10.1093/jxb/erl262) PMID: [17261695](https://pubmed.ncbi.nlm.nih.gov/17261695/).
20. Bargmann BO, Laxalt AM, ter Riet B, van Schooten B, Merquiol E, Testerink C, et al. Multiple PLDs required for high salinity and water deficit tolerance in plants. *Plant & cell physiology*. 2009; 50(1):78–89. Epub 2008/11/20. doi: [10.1093/pcp/pcn173](https://doi.org/10.1093/pcp/pcn173) PMID: [19017627](https://pubmed.ncbi.nlm.nih.gov/19017627/); PubMed Central PMCID: [PMC2638713](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC2638713/).
21. Kargiotidou A, Kappas I, Tsaftaris A, Galanopoulou D, Farmaki T. Cold acclimation and low temperature resistance in cotton: *Gossypium hirsutum* phospholipase Dalpha isoforms are differentially regulated by temperature and light. *Journal of experimental botany*. 2010; 61(11):2991–3002. Epub 2010/05/19. doi: [10.1093/jxb/erq124](https://doi.org/10.1093/jxb/erq124) PMID: [20478966](https://pubmed.ncbi.nlm.nih.gov/20478966/).
22. Wang C, Zien CA, Afithile M, Welti R, Hildebrand DF, Wang X. Involvement of phospholipase D in wound-induced accumulation of jasmonic acid in *Arabidopsis*. *Plant Cell*. 2000; 12(11):2237–46. Epub 2000/11/23. PMID: [11090221](https://pubmed.ncbi.nlm.nih.gov/11090221/); PubMed Central PMCID: [PMC150170](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC150170/).
23. Bargmann BO, Laxalt AM, Riet BT, Schouten E, van Leeuwen W, Dekker HL, et al. LePLDbeta1 activation and relocalization in suspension-cultured tomato cells treated with xylanase. *The Plant journal: for cell and molecular biology*. 2006; 45(3):358–68. Epub 2006/01/18. doi: [10.1111/j.1365-3113X.2005.02631.x](https://doi.org/10.1111/j.1365-3113X.2005.02631.x) PMID: [16412083](https://pubmed.ncbi.nlm.nih.gov/16412083/).
24. Yang YW, Bian SM, Yao Y, Liu JY. Comparative Proteomic Analysis Provides New Insights into the Fiber Elongating Process in Cotton. *J Proteome Res*. 2008; 7(11):4623–37. doi: [10.1021/Pr800550q](https://doi.org/10.1021/Pr800550q) PMID: [189000003](https://pubmed.ncbi.nlm.nih.gov/189000003/).
25. Wanjie SW, Welti R, Moreau RA, Chapman KD. Identification and quantification of glycerolipids in cotton fibers: reconciliation with metabolic pathway predictions from DNA databases. *Lipids*. 2005; 40(8):773–85. Epub 2005/11/22. PMID: [16296396](https://pubmed.ncbi.nlm.nih.gov/16296396/).
26. Zhang Y, Zhu H, Zhang Q, Li M, Yan M, Wang R, et al. Phospholipase dalpha1 and phosphatidic acid regulate NADPH oxidase activity and production of reactive oxygen species in ABA-mediated stomatal closure in *Arabidopsis*. *Plant Cell*. 2009; 21(8):2357–77. Epub 2009/08/20. doi: [10.1105/tpc.108.062992](https://doi.org/10.1105/tpc.108.062992) PMID: [19690149](https://pubmed.ncbi.nlm.nih.gov/19690149/); PubMed Central PMCID: [PMC2751945](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC2751945/).
27. Potikha TS, Collins CC, Johnson DI, Delmer DP, Levine A. The involvement of hydrogen peroxide in the differentiation of secondary walls in cotton fibers. *Plant physiology*. 1999; 119(3):849–58. Epub 1999/03/09. PMID: [10069824](https://pubmed.ncbi.nlm.nih.gov/10069824/); PubMed Central PMCID: [PMC32100](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC32100/).
28. Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(17):11109–14. Epub 2002/08/03. doi: [10.1073/pnas.162077099](https://doi.org/10.1073/pnas.162077099) PMID: [12154226](https://pubmed.ncbi.nlm.nih.gov/12154226/); PubMed Central PMCID: [PMC123218](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC123218/).

29. Tang K, Dong C, Liu J. Genome-wide analysis and expression profiling of the phospholipase D gene family in *Gossypium arboreum*. *Science China Life sciences*. 2015. Epub 2016/01/01. doi: [10.1007/s11427-015-4916-2](https://doi.org/10.1007/s11427-015-4916-2) PMID: [26718354](https://pubmed.ncbi.nlm.nih.gov/26718354/).
30. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature biotechnology*. 2015; 33(5):531–7. Epub 2015/04/22. doi: [10.1038/nbt.3207](https://doi.org/10.1038/nbt.3207) PMID: [25893781](https://pubmed.ncbi.nlm.nih.gov/25893781/).
31. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature biotechnology*. 2015; 33(5):524–30. Epub 2015/04/22. doi: [10.1038/nbt.3208](https://doi.org/10.1038/nbt.3208) PMID: [25893780](https://pubmed.ncbi.nlm.nih.gov/25893780/).
32. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012; 492(7429):423–7. Epub 2012/12/22. doi: [10.1038/nature11798](https://doi.org/10.1038/nature11798) PMID: [23257886](https://pubmed.ncbi.nlm.nih.gov/23257886/).
33. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997; 25(17):3389–402. Epub 1997/09/01. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/); PubMed Central PMCID: PMC146917.
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*. 1994; 22(22):4673–80. Epub 1994/11/11. PMID: [7984417](https://pubmed.ncbi.nlm.nih.gov/7984417/); PubMed Central PMCID: PMC308517.
35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004; 32(5):1792–7. Epub 2004/03/23. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/); PubMed Central PMCID: PMC390337.
36. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*. 2013; 30(12):2725–9. Epub 2013/10/18. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/); PubMed Central PMCID: PMC3840312.
37. Plotree D, Plotgram D. PHYLIP-phylogeny inference package (version 3.2). *cladistics*. 1989; 5:163–6.
38. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. *Genome research*. 2009; 19(9):1639–45. doi: [10.1101/gr.092759.109](https://doi.org/10.1101/gr.092759.109) PMID: [18099597](https://pubmed.ncbi.nlm.nih.gov/18099597/).
39. Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual review of plant biology*. 2009; 60:433–53. Epub 2009/07/07. doi: [10.1146/annurev.arplant.043008.092122](https://doi.org/10.1146/annurev.arplant.043008.092122) PMID: [19575588](https://pubmed.ncbi.nlm.nih.gov/19575588/).
40. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC plant biology*. 2004; 4:10. Epub 2004/06/03. doi: [10.1186/1471-2229-4-10](https://doi.org/10.1186/1471-2229-4-10) PMID: [15171794](https://pubmed.ncbi.nlm.nih.gov/15171794/); PubMed Central PMCID: PMC446195.
41. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics, proteomics & bioinformatics*. 2006; 4(4):259–63. Epub 2007/05/29. doi: [10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2) PMID: [17531802](https://pubmed.ncbi.nlm.nih.gov/17531802/).
42. Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH. Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell*. 2004; 16(5):1220–34. Epub 2004/04/24. doi: [10.1105/tpc.020834](https://doi.org/10.1105/tpc.020834) PMID: [15105442](https://pubmed.ncbi.nlm.nih.gov/15105442/); PubMed Central PMCID: PMC423211.
43. Blanc G, Wolfe KH. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*. 2004; 16(7):1667–78. Epub 2004/06/23. doi: [10.1105/tpc.021345](https://doi.org/10.1105/tpc.021345) PMID: [15208399](https://pubmed.ncbi.nlm.nih.gov/15208399/); PubMed Central PMCID: PMC514152.
44. Li WH, Gojorbori T. Rapid evolution of goat and sheep globin genes following gene duplication. *Molecular biology and evolution*. 1983; 1(1):94–108. Epub 1983/12/01. PMID: [6599963](https://pubmed.ncbi.nlm.nih.gov/6599963/).
45. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>( $\Delta\Delta C_T$ ) Method. *Methods*. 2001; 25(4):402–8. Epub 2002/02/16. doi: [10.1006/meth.2001.1262](https://doi.org/10.1006/meth.2001.1262) PMID: [11846609](https://pubmed.ncbi.nlm.nih.gov/11846609/).
46. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques*. 2003; 34(2):374–8. Epub 2003/03/05. PMID: [12613259](https://pubmed.ncbi.nlm.nih.gov/12613259/).
47. Kolesnikov YS, Nokhrina KP, Kretynin SV, Volotovskii ID, Martinec J, Romanov GA, et al. Molecular structure of phospholipase D and regulatory mechanisms of its activity in plant and animal cells. *Biochemistry-Moscow+*. 2012; 77(1):1–14. doi: [10.1134/S0006297912010014](https://doi.org/10.1134/S0006297912010014) PMID: [22000001](https://pubmed.ncbi.nlm.nih.gov/22000001/).

48. Wu P, Shao ZQ, Wu XZ, Wang Q, Wang B, Chen JQ, et al. Loss/retention and evolution of NBS-encoding genes upon whole genome triplication of *Brassica rapa*. *Gene*. 2014; 540(1):54–61. doi: [10.1016/j.gene.2014.01.082](https://doi.org/10.1016/j.gene.2014.01.082) PMID: [ISI:000334141000009](https://pubmed.ncbi.nlm.nih.gov/24774859/).
49. Todeschini AL, Georges A, Veitia RA. Transcription factors: specific DNA binding and specific gene regulation. *Trends in genetics: TIG*. 2014; 30(6):211–9. Epub 2014/04/30. doi: [10.1016/j.tig.2014.04.002](https://doi.org/10.1016/j.tig.2014.04.002) PMID: [24774859](https://pubmed.ncbi.nlm.nih.gov/24774859/).
50. Xiao YH, Li DM, Yin MH, Li XB, Zhang M, Wang YJ, et al. Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. *Journal of plant physiology*. 2010; 167(10):829–37. Epub 2010/02/13. doi: [10.1016/j.jplph.2010.01.003](https://doi.org/10.1016/j.jplph.2010.01.003) PMID: [20149476](https://pubmed.ncbi.nlm.nih.gov/20149476/).
51. Wang L, Zhu Y, Hu W, Zhang X, Cai C, Guo W. Comparative Transcriptomics Reveals Jasmonic Acid-Associated Metabolism Related to Cotton Fiber Initiation. *PloS one*. 2015; 10(6):e0129854. Epub 2015/06/17. doi: [10.1371/journal.pone.0129854](https://doi.org/10.1371/journal.pone.0129854) PMID: [26079621](https://pubmed.ncbi.nlm.nih.gov/26079621/); PubMed Central PMCID: PMC4469610.
52. Yang SS, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, et al. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant Journal*. 2006; 47(5):761–75. doi: [10.1111/j.1365-313X.2006.02829.x](https://doi.org/10.1111/j.1365-313X.2006.02829.x) PMID: [ISI:000239700000009](https://pubmed.ncbi.nlm.nih.gov/16461577/).
53. Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, et al. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell*. 2006; 18(3):651–64. Epub 2006/02/08. doi: [10.1105/tpc.105.040303](https://doi.org/10.1105/tpc.105.040303) PMID: [16461577](https://pubmed.ncbi.nlm.nih.gov/16461577/); PubMed Central PMCID: PMC1383640.
54. Zhang M, Zheng X, Song S, Zeng Q, Hou L, Li D, et al. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nature biotechnology*. 2011; 29(5):453–8. Epub 2011/04/12. doi: [10.1038/nbt.1843](https://doi.org/10.1038/nbt.1843) PMID: [21478877](https://pubmed.ncbi.nlm.nih.gov/21478877/).