Mol Genet Genomics

ORIGINAL PAPER

Molecular characterization of two homoeologous elicitor-responsive lipin genes in cotton

Sonia M Phillips . Ian A Dubery . Henriette van Heerden

S.M. Phillips . I.A. Dubery . H. van Heerden
Department of Biochemistry, University of Johannesburg, Kingsway Campus, P.O. Box 524, Auckland Park, 2006, South Africa

Present Address:

S.M. Phillips
Department of Plant Science, University of Pretoria, Hatfield, Pretoria, 0001, South Africa

H. van Heerden
Department of Veterinary Tropical Diseases, University of Pretoria, Private Bag X04, Onderstepoort, 0110, South Africa

I.A. Dubery (✉)
Department of Biochemistry, University of Johannesburg, Kingsway Campus, P.O. Box 524, Auckland Park, 2006, South Africa

Note: Nucleotide sequence data reported are available in the GenBank database under the accession numbers HQ630674 (*GhLIPN-1*) and HQ630675 (*GhLIPN-2*)

Abstract The identification and molecular characterization of two lipin-like gene copies (*GhLIPN*) in cotton, *Gossypium hirsutum*, an allotetraploid derived from two progenitor diploid *Gossypium* species, is described. Sequence analyses of the *GhLIPN* copies, designated *GhLIPN-1* and *-2*, revealed that they contain 11 exons, separated by 10 introns. They each have a 2643 bp open reading frame that encodes 880 aa proteins, and share a 97.7% and 95.5% sequence similarity at the translated nucleotide and amino acid level, respectively. The *GhLIPN* genes have a distinct domain architecture consisting of an archetypical N-terminal lipin domain, followed by a haloacid dehalogenase (HAD) domain towards the C-terminus. A Southern blot did not distinguish between the two gene copies, which suggest that they may be homoeologs rather than paralogs. *GhLIPN-2* is more similar to a homologous sequence from *G. raimondii*, representing the ancestral D-genome, compared to *GhLIPN-1* that matches *G. herbaceum* and that represents the A-genome. Our data indicates that *GhLIPN-1* and *GhLIPN-2* are homoeologs that derive from the A- and the D-diploid genomes respectively. The promoter sequences of *GhLIPN-1* and *-2* differ by 56%, as a result of multiple indels. *In silico* analysis of the promoter regions revealed that both genes both contain numerous putative defence-related and elicitor-responsive *cis*-elements that supports a role for *GhLIPN* in defence responses. Relative quantification real-time PCR confirmed the up-regulation in response to a cell wall-derived *V. dahliae* elicitor, which supported the association of *GhLIPN* with defence signalling. The results add a new dimension to the proposed roles of lipins in plants by suggesting that lipins may have a role in defence signalling.

Keywords Allotetraploid . *Gossypium hirsutum* . haloacid dehalogenase . homoeologs . lipin . phosphatidate phosphatase . *Verticillium dahliae*

Introduction

The fungal phytopathogen *Verticillium dahliae* is widespread, occurring on a variety of economically important crops, including cotton (Klosterman et al. 2009), where infection is characterized by wilting (Göre et al. 2009) and results in significant cotton crop losses (Bolek et al. 2005; Aguado et al. 2010). Differential display transcriptome analysis was previously used to identify changes in gene expression of *V. dahliae*-resistant *G. hirsutum* cells challenged with a cell-wall derived (CWD) *V. dahliae* elicitor, to identify defence-related genes. An up-regulated lipin-like expressed sequence tag (EST) (designated C1B10) was identified (Zwiegelaar 2003). This study is the first report of lipin genes in cotton and provides a foundation for the further investigation of the role(s) of plant lipins.

Lipins are recently identified phosphatidate phosphatases (PAPs) which have a crucial function in the lipid metabolism of eukaryotes (Reue and Zhang 2008). Owing to their PAP activity (Type 1, Mg^{2+}-dependent), lipins convert phospholipase D (PLD)-derived phosphatidic acid (PA) to diacylglycerol (DAG) and phosphate, in the presence of Mg^{2+}. DAG is further metabolized to produce integral membrane components such as phospholipids and galactolipids, in addition to triacylglycerol (TAG), for lipid storage (Reue and Zhang 2008; Carman and Han 2009). Lipin orthologues have been found in distant eukaryotic species and lipin mutants of diverse species have illustrated the vast implications of lipin activity, which ranges from fat metabolism in mammals to dynamic membrane remodelling in yeast (Han et al. 2007), nematodes (Gorjánácz and Mattaj 2009), plants (Nakamura et al. 2009; Eastmond et al. 2010) and mice (Péterfy et al. 2001; Donkor et al. 2009). The mediation of lipid metabolism by lipins is clearly a dynamic process, which allows the organism to adapt to its ever-changing environment (Csaki and Reue, 2010).

Two lipin orthologs (AtPAH1 and AtPAH2) were identified in the model plant *Arabidopsis thaliana* (Nakamura et al. 2009). *AtPAH1* and *AtPAH2* encode functional PAP1 enzymes, which are responsible for the eukaryotic pathway of DAG synthesis (Mietkiewska et al. 2011). These lipins mediate membrane lipid remodelling, which occurs exclusively through the eukaryotic endoplasmic reticulum (ER)-mediated supply of DAG. Nakamura et al. (2009) proposed that plants employ lipin-mediated lipid remodelling to overcome phosphate shortage stress and *pah1*/2 *A. thaliana* double mutant studies by Eastmond et al. (2010), suggests that AtPAH1/2 function indirectly to repress phospholipid biosynthesis and that this regulation likely occurs at the transcript level for many enzymes.

Commercially cultivated cotton, *Gossypium hirsutum*, is an allotetraploid that derived from a genomic hybridization event between diploid cotton species from Africa (A-genome) and America (D-genome). The

genomic merger occurred approximately 1-2 million years ago (Mya), and prior to this event, the A- and D-diploid species had evolved in isolation for approximately 5-10 million years (Senchina et al. 2003; Wendel and Cronn 2003). Consequently, *G. hirsutum* has two fully differentiated co-resident genomes, described as $A_TD_{T,}$ (Udall et al. 2006), and at each given locus there are two copies each from the maternal ancestral parent (A-genome) and two from the paternal parent (D-genome) (Wendel and Cronn 2003; Adams et al. 2009). The progenitor A- and D-genomes are represented by the extant species *G. herbaceum* (A_1-genome), *G. arboreum* (A_2-genome) and *G. raimondii* (D-genome) (Cronn et al. 1999; Udall et al. 2006; Adams et al. 2009). Many gene copies arising from the A_T and D_T subgenomes have recently been mapped to the A- and D-genomes from these diploid *Gossypium* species (van Deynze et al. 2009; Flagel and Wendel 2010). In polyploids, the allele pairs are called homoeologs, since they derived from the genomic merger of two different species (Wendel 2000; Flagel et al. 2009). Evidence has shown that homoeologs evolved independently following polyploid formation in cotton (Cronn et al. 1999; Small and Wendel 2002). However, Salmon et al. (2010) demonstrated that the co-resident genomes have undergone between 1.8-1.9% nonreciprocal homoeologous exchanges, or gene conversions, since the genomic merger 1–2 Mya.

Although polyploidy is prevalent among angiosperms (Flagel et al. 2009; Flagel and Wendel 2010), with > 70% having undergone one or more episodes of polyploidy at some point (Moore and Purugganan 2005), and > 75% of these resulted in allopolyploids (Jackson and Chen 2009), the vast genetic implications of polyploidy on plant diversification and adaptive evolution have only recently become apparent due to large-scale investigations into comparative genomics, transcriptomics and epigenomics (Flagel and Wendel 2009; Ainouche and Jenczewski 2010). Many molecular changes arise immediately following allopolyploid formation, of which the most notable involve changes in gene expression (Adams and Wendel 2005; Dong and Adams 2011).

Materials and methods

Plant material, elicitor preparation and induction of cell suspension cultures

Cotton (*Gossypium hirsutum*) cv OR_{19} (*V. dahliae* resistant cultivar) cell suspension cultures were established from callus tissue and grown in the dark at 25 $\pm$ 3°C on a continuous rotary shaker at 120 rpm (Phillips et al. 2012). All experiments were performed on cells in the logarithmic growth phase, 2-3 days after sub-cultivation. The *V. dahliae* elicitor was prepared from the heat-released fraction of the mycelial cell walls (Dubery and Slater

1997). Cotton cell suspensions (25 ml) were treated with 5 µg/ml CWD *V. dahliae* elicitor, or only culture media for the controls. The suspensions were incubated for 24 h at 25 °C (Phillips et al. 2013).

Genome walking

Genomic DNA was isolated using a CTAB method (Murray and Thompson 1980). DNA (6 µg) was digested with 2.5 U/µg DNA of the following restriction enzymes: *Stu*I, *Dra*I, *Pvu*II, *Eco*RV, *Sna*B, *Ssp*I, *Pdi*I, *Eco*47III, *Sac*I, *Xba*I, *Eco*RI and *Hind*III (Fermentas, St. Leon-Rot, Germany). The reaction was incubated for 16 h at 37 °C. Adaptors 1 (5'-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT-3') and 2 (5'-ACCAGCCC-3') from a GenomeWalker™ Universal kit (Clontech, Madison, WI, USA) were annealed together to yield a 25 µM genome walker adaptor using an initial incubation at 94°C for 5 min, after which the temperature was reduced every min by 1 °C, until the temperature reached 4 °C. The ends of each of the digested genomic library fragments (10 µl) were ligated to 25 µM genome walker adaptor with 3 U T4 DNA ligase (Bioline, London, UK) at 16°C for 16 h. The reaction was diluted 5 times with ddH_2O and primary and secondary/nested PCR were performed on the adaptor-ligated, restriction-digested cotton DNA. Gene-specific primers (GSPs) were designed to amplify genomic sequences upstream and downstream of the 213 bp EST, named C1B10. The primary PCRs (25 µl) contained 1x Ex Taq™ buffer, 0.2 mM dNTPs (Takara, Madison, WI, USA), 0.2 µM adaptor primer 1 (AP1) (5'-GTAATACGACTCACTATAGGGC-3'), 0.2 µM GSP (Additional file 1: Table S1), 1.25 U TaKaRa Ex Taq™ (Takara Bio Inc.), and 0.1-1 µg adaptor-ligated, digested DNA template. The PCR conditions consisted of an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 20 s, annealing at 55°C for 30 s, and elongation at 72°C for 4 min, followed by a final elongation step at 72°C for 15 min. The primary PCRs were diluted 1:49 in ddH_2O and 0.5 µl of the diluted primary PCR products were used as a template for the secondary (nested) PCR with 0.2 µM adaptor primer 2 (AP2) (5'-ACTATAGGGCACGCGTGGT-3') and an internal GSP (Additional file 1: Table S1). The same cycling conditions were used for the secondary PCRs, except the annealing temperature was increased to 60-65°C to improve sensitivity. The PCR products were purified, cloned into a pGEM®-T Easy vector (Promega, Fitchburg, WI, USA) and sequenced.

Rapid amplification of cDNA ends (RACE)

The 5' and 3' RACE reactions were performed with a 5'/3' RACE kit, 2nd Generation (Roche Diagnostics, Mannheim, Germany) in order to obtain the full-length cDNA sequence of each gene copy. The protocol was carried out according to the manufacturer's instructions with minor modifications. Briefly, total RNA was isolated from cotton cell suspensions with an RNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) and mRNA was isolated from the total RNA with an Oligotex mRNA Mini Kit (Qiagen). mRNA (250 – 500 ng) was reverse transcribed to cDNA with 0.5 µM GSP (Additional file 1: Table S1), and 25 U Transcriptor Reverse Transcriptase (Roche Diagnostics) in 10 µl reaction. The 5' single-stranded cDNA molecule was purified with a Nucleospin® Extract kit (Macherey-Nagel, Düren, Germany) or DNA Clean and Concentrator™ kit (Zymo Research, Freiburg, Germany). A homopolymeric A-tail was then added to the 3' end of the cDNA strand with a recombinant Terminal Transferase (Roche Diagnostics) and dATP. Primary PCRs were performed with 0.2 µM oligo dT-anchor primer (5'-GACCACGCGTATCGATGTCGACT$_{16}$V-3' V = A, C or G), and 0.2 µM of a second GSP (Additional file 1: Table S1). The PCR cycling conditions consisted of denaturation at 95°C for 2 min, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 55°C for 30 s, extension at 72°C for 2 min, followed by a final extension at 72°C for 7 min. The primary 5' RACE reaction products were diluted 1:49 with ddH_2O and used as the templates in secondary 5' RACE reactions. The secondary 5' RACE reactions were performed with 0.2 µM oligo dT-anchor primer and 0.2 µM of a third internal GSP (Additional file 1: Table S1). The same cycling parameters were used for the secondary 5' RACE PCR, except the annealing temperature was raised to 60-65°C to improve sensitivity. The 5' RACE products were purified, cloned into a pGEM®-T Easy vector and sequenced.

For 3' RACE, 1 µl (40 – 80 ng) mRNA was reverse transcribed to cDNA with 0.5 µM 3' RACE adapter (5'-GCGAGCACAGAATTAATACGACTCACTATAGGT$_{12}$VN-3' V = A, C or G, N = any base) from an Ambion FirstChoice® RLM-RACE kit (Ambion, Naugatuck, CT, USA) and 25 U Transcriptor Reverse Transcriptase (Roche Diagnostics) in a 10 µl reaction according to the manufacturer's instructions. A primary PCR was performed with 0.2 µM 3' RACE outer primer (5'-GCGAGCACAGAATTAATACGACT-3') (Ambion) and 0.2 µM GSP (Additional file 1: Table S1). The PCR cycling conditions consisted of denaturation at 95°C for 2 min, followed by 35 cycles of denaturation at 94°C for 20 s, annealing at 55°C for 30 s, extension at 72°C for 30 s, final extension at 72°C for 7 min. The primary 3' RACE reaction products were diluted 1:49 with ddH_2O and used as templates in the secondary 3' RACE reaction, performed with 0.2 µM 3' RACE inner

primer (5'-CGCGGATCCGAATTAATACGACTCACTATAGG-3' (Ambion) and 0.2 µM third internal GSP (Additional file 1: Table S1). The cycling conditions were the same as the primary 3' RACE PCR, except the annealing temperature was raised to 60-65°C to increase specificity. The 3' RACE products were purified, cloned into a pGEM®-T Easy vector and sequenced.

Obtaining the full-length *GhLIPN* sequence

Three PCR reactions were performed to obtain genomic sequences not covered by the genome walking reactions. Each PCR contained 0.04 U ExSel high fidelity DNA polymerase (JMR Holdings, Kent, UK), 1 X reaction buffer that provided 2 mM $MgSO_4$ final concentration (f.c.), 0.2 mM dNTPs (Bioline), 0.2 µM of each GSP (Additional file 1: Table S1) and 50-135 ng DNA template. The PCR cycling parameters included an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 20 s, annealing at 60°C for 30 s and elongation at 70°C for 2.5 min, followed by a final elongation step at 70°C for 7 min. The PCR products were purified, cloned into a pGEM®-T Easy vector and sequenced.

A RT-PCR was performed to obtain cDNA sequence not covered by the RACE reactions. PCR was performed with 0.5 µl cDNA template, 0.08 U ExSel high fidelity DNA polymerase (JMR Holdings), 1 X reaction buffer (at a 2 mM $MgSO_4$ f.c.), 0.2 mM dNTPs (Bioline) and 0.2 µM of each primer (LiD 1F (5'-CCTTAGTTGGAAGGGATTGGA-3') and RR3 (5'-TTCCTTACAAGATGAACCCCAAC-3'). The PCR cycling parameters consisted of an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 20 s, annealing at 60°C for 30 s and elongation at 70°C for 1 min, followed by a final elongation step at 70°C for 7 min. The PCR products were purified, cloned into a pGEM®-T Easy vector and sequenced.

Obtaining the full-length genomic sequences of each *GhLIPN* gene copy

Primers were designed to distinguish between two gene copies based upon the *GhLIPN* sequence data. The putative homoeologs were designated *GhLIPN-1* and *GhLIPN-2* based on a 13 bp insertion/deletion site in intron 6, which was initially the main distinguishing feature between the genomic sequences obtained (***1*** = No insert , ***2*** = Insert present). The insert sequence was present in the primer used to amplify *GhLIPN-2* and absent in the primers used to amplify *GhLIPN- 1* (Additional file 1: Table S1). Each PCR contained 1 X reaction buffer (at 2 mM $MgSO_4$ f.c.), 0.2 mM dNTPs (Bioline), 0.2 µM of each homoeolog-specific primer (HSP), 0.04 U ExSel

DNA polymerase (JMR Holdings) and 133 ng DNA template. The PCR cycling parameters were an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 20 s, annealing at 60°C for 30 s and elongation at 70°C for 4 min, followed by a final elongation step at 70°C for 7 min. The PCR products were purified, cloned into a pGEM®-T Easy vector and sequenced.

Amplification of each *GhLIPN* promoter

Primers were designed to distinguish between the promoter regions of both gene copies. A 17 bp indel was absent in the reverse primer used to amplify the *GhLIPN-2* promoter and present in the reverse primer used to amplify the *GhLIPN-1* promoter (Additional file 1: Table S1). The PCR reaction was identical to above mentioned, except 0.08 U ExSel DNA polymerase (JMR Holdings) was added. The cycling parameters consisted of an initial denaturation at 94°C for 2 min, then 35 cycles of denaturation at 94°C for 20 s, annealing at 50-60°C for 30 s and elongation at 70°C for 2 min, followed by a final elongation step at 70°C for 7 min. The PCR products were purified, cloned into a pGEM®-T Easy vector and sequenced.

Sequence analyses, assembly and alignments

The possibility of introduced errors in PCR amplification and sequencing was kept to a minimum by using a high-fidelity Taq polymerase in all the PCRs, and sequencing three or more clones from each single single PCR. DNA sequencing was performed according to the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA). Sequences were analysed and assembled with ChromasPro (Technelysium, Brisbane Australia), whereas DNAssist (Version 5.1) and CLC Main Workbench was utilized for routine genomic sequence alignments. Homologous protein sequences were identified and compared to characterized gene products with the blastp (basic local alignment search tool - BLAST) search algorithm provided by the NCBI/NIH (http://www.ncbi.nlm.nih.gov.BLAST). Alignments between homologous protein sequences were performed with ClustalW in ChromasPro.

A phylogenetic tree showing the relationship between the lipin homologs and their ancestral representatives from *G. rainmondii* and *G. herbaceum* was constructed with neighbourhood joining analysis using MEGA5 (Tamura et al. 2011). The 811 bp sequence from *G. herbaceum* was amplified with Herb_lip_F1

(5'-GCTTTGTTTATGTAGGATGG-3') and Herb_lip_R1 (5'-GTTCAATGGATGAGGTTG-3') (Additional file 1, Table S1) and sequenced (Inqaba Biotech, Pretoria, South Africa).

PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) and PLACE (http://www.dna.affrc.go.jp/PLACE), two web-based software programs, were used to identify putative *cis*-elements in the promoter regions. Several bioinformatics web-based software programs were used for complementary prediction and/or analysis of the domain architecture of the gene products and identification of signature motif regions. These included: InterPro (www.ebi.ac.uk/interpro), Pfam (http://pfam.sanger.ac.uk), ExPASy proteomics server (expert protein analysis system) from the Swiss Institute of Bioinformatics (SIB) (http://au.expasy.org), PROSITE (http://ca.expasy.org/prosite), ProtParam (protein identification and analysis tool) (http://au.expasy.org/tools/protparam.html) PANTHER (protein analysis through evolutionary relationships) (www.pantherdb.org), SMART (simple modular architecture research tool) (http://smart.embl-heidelberg.de), PRODOM (protein domain) (http://prodom.prabi.fr/prodom/current/html), My Hits - Motif Scan using hidden Markov models (HMMs) (http://myhits.isb-sib.ch/cgi-bin/motif_scan) and GO (gene ontology) (www.geneontology.org).

Southern blot

Genomic DNA was extracted from cotton cell suspensions using a CTAB protocol (Sambrook et al. 1989). DNA (30 µg) was restriction-digested (3 U restriction enzyme/µg DNA) overnight at 37°C with *Xba*I, *Eco*RI, and *Hind*III (Fermentas, St. Leon-Rot, Germany) in a total volume of 200 µl. The restriction digests were purified with a sodium acetate precipitation to remove contaminants and reduce the volume of the digested DNA. The purified DNA (20 µg) was electrophoresed at 4°C on a 0.8% (w/v) TAE (40 mM Tris-acetate; 1 mM EDTA) agarose gel at 20 V and DNA was transferred to a Hybond-N^+ nylon membrane (Amersham, Little Chalfont, UK) using an upward transfer system (Sambrook et al. 1989). The membrane was baked at 80°C for 2 h, to immobilize the DNA. The 123 bp probe targeting the third exon was prepared with a PCR digoxygenin (DIG) Probe Synthesis Kit (Roche Diagnostics), according to the manufacturer's instructions using forward (RTF2) and reverse (RTR2) primers (Additional file 1: Table S1). The membrane was hybridized with 20 pM (final concentration) of the heat-denatured DIG-labelled probe in pre-warmed ULTRAhyb™ Ultrasensitive Hybridization Buffer (Ambion) for 20 h at 42 °C with constant agitation. Detection of the membrane was carried out with a DIG Luminescence Detection Kit (Roche Diagnostics), according to the manufacturer's instructions.

Time study of gene expression in cotton cell suspensions induced with the CWD *V. dahliae* elicitor

Cotton cell suspensions (cv OR_{19}) were treated with 5 µg/ml CWD *V. dahliae* elicitor. RNA was isolated from the suspensions at 0 (calibrator), 2, 4, 6, 8, and 10 h with an RNeasy® Plant Mini Kit (Qiagen). Total RNA was digested with an RNase-free DNase (Promega) to remove any carry-over DNA contamination. The RNA concentration and quality was assessed by spectrophotometry and 1.2% agarose gel electrophoresis, respectively. A Transcriptor Reverse Transcriptase (25 U) (Roche Diagnostics) was used to reverse transcribe 1 µg RNA to cDNA in a 10 µl RT-PCR reaction according to the manufacturer's instructions. Relative qRT-PCR was performed with *18S* as a reference gene using a LightCycler® FastStart DNA Master^{PLUS} SYBR Green I kit (Roche Diagnostics) on a RotorGene 3000 (Corbett Research, Sydney, Australia), using 2 µl cDNA as the template and 0.2 µM of forward (RTF2) and reverse (RTR2) primers (Additional file 1: Table S1). The cycling conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 10 s, 60°C for 8 s, 72°C for 10 s. The experiment was performed in duplicate with two technical PCR repeats. The qRT-PCR data was processed according to the standard curve method (Larionov et al. 2005). A non-treated, 0 h calibrator was used to normalize the expression of *GhLIPN* (designated as 1-fold), with all experimentally-derived quantities reported as an n-fold difference relative to the calibrator, as described in Wong and Medrano (2005).

Results

Analysis of the full-length *GhLIPN* sequences

The full genomic sequence containing the *GhLIPN* gene was identified and characterized from the C1B10 EST using genome walking, PCR, 5'/3' RACE and RT-PCR. The genomic sequence contained untranscribed regions at both the 5' (upstream) and 3' (downstream) ends (Additional file 2: Figures S1 and S2). The transcribed *GhLIPN* sequences were identified from 5'/3'-RACE and RT-PCR, and translated to reveal that *GhLIPN* encodes a lipin phosphatidate phosphatase (PAP). During *GhLIPN* sequence assembly, it became apparent that two copies of the gene (designated *GhLIPN-1* and -2) were amplified. Homoeolog-specific PCR amplification of *GhLIPN*, using homoeolog-specific gene primers (HSPs) (Additional file 1: Table S1), was subsequently performed to obtain the full genomic sequence of each gene copy. The sizes of the genomic sequences containing the *GhLIPN* gene copies were 8423 and 6704 bp, for *GhLIPN-1* (Genbank: HQ630674) and *GhLIPN-*

2 (Genbank: HQ630675), respectively and their genomic sequences had an overall sequence divergence of 2.5% (ungapped).

The putative *GhLIPN* transcription start site (TSS) (Additional file 3: Figure S3) was determined from a 5' RACE reaction that produced a 1846 bp fragment (results not shown). The transcription end site was determined from the longest of two 3' RACE products (a 448 bp fragment, results not shown), which matched *GhLIPN-1*. The size of the *GhLIPN-1* 3'-untranslated region (UTR) was 351bp and poly(A) signal sequences were identified downstream of the 3'-end (Additional file 2: Figure S1). Due to the preferential amplification of *GhLIPN-1* in the 3' RACE and downstream genome walking reactions, the experimentally obtained *GhLIPN-1* genomic sequence included the 3'-UTR, as well as genomic sequence following the transcription end site (obtained from a 1402 bp fragment) (Additional file 2: Figure S1). In contrast, the genomic sequence obtained for *GhLIPN-2* did not contain the full 3'-UTR, but ended at the same site as the C1B10 EST, after the translation stop site (Additional file 2: Figure S2).

Sequences obtained from the RACE and RT-PCRs were assembled and aligned with the genomic DNA sequences to obtain the full cDNA sequence of each *GhLIPN* gene copy and to identify intron splice junctions. The homoeologs each contain 11 exons, separated by 10 introns at the same intron-exon splice junctions (Figure 1). Unusually, the first exon is in the 5'-UTR and is therefore non-coding. There are therefore 10 coding exons in the *GhLIPN-1* and -2 open reading frames (ORFs), separated by 9 introns. The intron sequences differ by 2.97% and the exons differ by 2.27% (ungapped). The sizes of the full-length cDNA transcripts are 4 467 and 4 479 bp, for *GhLIPN-1* and -2, respectively, and the ORF sequence is 2643 bp for both *GhLIPN* gene copies. The translation start and end sites were determined by translation of the spliced *GhLIPN* transcripts from the first ATG codon to the first in-frame stop codon (TAA) (Additional file 2: Figures S1 and S2). The ATG translation start site is located 715 and 695 bp downstream from the predicted TSSs in the genomic sequences of *GhLIPN-1* and -2, respectively, due to the intron in the 5'-UTR. The ATG start site is, however, only 204 and 186 bp downstream from the predicted TSSs of the *GhLIPN-1* and -2 spliced transcripts, respectively (Additional file 2: Figures S1 and S2).

GhLIPN amino acid sequence comparison and physico-chemical properties

The deduced GhLIPN-1 and -2 amino acid sequences have a length of 880 aa, an estimated molecular mass of 96.72 kDa and 96.94 kDa respectively and a calculated pI of 4.70 (deduced from DNAssist and ExPASy). A

ClustalW alignment between the GhLIPN-1 and -2 amino acid sequences revealed 40 mismatches (4.50% of the total length), of which 16 and 5 are conserved and semi-conserved, respectively (Additional file 3: Figure S3). A 5 aa DxDxT/V HAD domain catalytic motif and 17 aa bipartite nuclear localization signal (NLS) are conserved in both proteins (Figure 1A and Additional file 3: Figure S3). Four amino acid mismatches are evident in the N-terminal lipin domains of the GhLIPN proteins, of which two are conserved (S/T and Q/E) and two are semi-conserved (A/P and A/V). Only one (unconserved) mismatch occurs in the HAD domains of the GhLIPN proteins (where GhLIPN-2 has an Arg in place of an Ile at position 824).

Domain and post-translational modification analysis

The *GhLIPN-1* and *-2* genes were confirmed to code for lipin proteins, based on the results of several motif scans. The lipin domain occurs at the N-terminus of the proteins and a HAD-like domain is found towards the C-terminus within a region loosely referred to as the C-terminal lipin (C-LIP) domain. Figure 1A displays the various motifs graphically and indicates where on the protein sequences they are located. GhLIPN-1 and -2 were analysed further to examine potential post-translational modification sites (Additional file 4: Table S2).

GhLIPN homology

The GhLIPN amino acid sequences were used in a Position-Specific Iterated (PSI)-BLAST of the non-redundant (nr) *Viridiplantae* NCBI database to compare them with currently available homologous lipin domain-containing protein sequences in plants (hereon referred to as lipin-like). Of the eleven identified *Viridiplantae* species with homologous lipin-like proteins seven were from seed plants, three from green algae and one from a moss (Table 1). The lipin-like proteins with the greatest homology to the GhLIPN homoeologs had E-scores of 0. These were from *Vitis vinifera* (XP_002274246), with a percentage identity of 61-62%, and *A. thaliana* AtPAH1 (At3g09560, NP_187567), with percentage identity of 53%. AtPAH2 (At5g42870, NP_199101), an ortholog of AtPAH1 had an E-score of $2e^{-153}$ and $4e^{-155}$ and a percentage identity of 57% to GhLIPN-1 and -2, respectively.

An alignment between the GhLIPN homoeologs and the lipin-like proteins from *V. vinifera* (XP_002274246) and *A. thaliana* (NP_187567, representing AtPAH1) is shown in Additional file 5, Figure S4. The HAD domains are found within C-LIP domains, which were identified from an alignment of lipins from different taxonomic groups (Péterfy et al. 2001). Overall, a higher degree of similarity is evident within the

conserved N- and C-terminal lipin domains compared to the sequences between the two domains. The DxDxT/V catalytic signature motifs of the HAD domain and the NLS are highly conserved among the plant lipin-like proteins. The same high degree of conservation was evident when the HAD domains of the *GhLIPN* homoeologs were aligned to those of lipin-like proteins in other plant species.

Ancestral genome prediction of GhLIPN homoeologs

In an attempt to determine which ancestral genome (A or D) each of the *GhLIPN* homoeolog copies are derived from, a BLAST search of the official cotton EST database (http://www.agcol.arizona.edu/cgi-bin/pave/Cotton/index.cgi) was performed to find sequences representing *GhLIPN* orthologs from *G. arboreum* /*G. herbaceum* and *G. raimondii* (extant species representing the ancestral A- and D-genomes, respectively). An orthologous EST (Genbank: CO087195), comprised of 5' and 3' mRNA, from *G. raimondii* (D-genome) had an E-score of 0 to the *GhLIPN* homoeologous cDNA sequences. No sequences representing orthologous lipin genes were identified from *G. arboreum* / *G. herbaceum* (A-genome) on NCBI at the time of the study so a portion of the gene was amplified from *G. herbaceum* genomic DNA for sequence comparisons. The spliced nucleotide *GhLIPN* sequences (transcribed sequence from 1-210 nt and genomic sequence from 211-1021 nt in the alignment) were aligned with the *G. raimondii* CO087195 EST and an 811 bp genomic region amplified from *G. herbaceum* (Additional file 6: Figure S5A). The translated amino acid sequences were also compared (Figure S5B). A phylogenetic tree was constructed based on a 500 bp region where all four sequences were aligned in the nucleotide alignment (from the beginning of the *G. herbaceum* sequence to the end of the *G. raimondii* sequence) to predict the genomic ancestry of the homoeologs (Figure 2). It is clear from the multiple sequence alignments and the phylogenetic analysis that the *GhLIPN-2* homoeolog derived from the ancestral D-genome (represented by the CO087195 EST from *G. raimondii*) and that the *GhLIPN-1* homoeolog derived from the ancestral A-genome (represented by the 811 bp amplified fragment from *G. herbaceum*).

Copy number determination

A Southern blot with genomic DNA from *G. hisutum* digested with *Xba*I, *Eco*RI, and *Hind*III was performed to determine the copy number of *GhLIPN*. High stringency hybridization of the DIG-labelled probe, targeting the

third exon (Figure 1B), each produced a single band (Figure 3). A single band is consistent with the gene sequences as there are identical resitiction sites for the enzymes used in both the two gene copies.

Promoter analyses and *GhLIPN* expression in response to the cell wall-derived *Verticillium dahliae* elicitor

Promoter sequences corresponding to each gene copy were obtained by PCR, performed upstream of an identified 17 bp indel in the 5'-untranslated region (5'-UTR) of *GhLIPN* with HSPs. The obtained *GhLIPN-1* promoter sequence was 1464 bp compared to the 1289 bp obtained for *GhLIPN* -2 (Additional file 7, Figure S6). The alignment of the promoter regions shows a large number of SNPs between the *GhLIPN* promoters, as well as many indels. The largest indel is a 437 bp fragment, present only in the *GhLIPN-1* promoter. The *GhLIPN-1* and -2 promoters were examined for *cis*-acting elements, associated with the binding of specific regulatory transcription factors (TFs). The length of the analysed promoter sequence was restricted to 1 kb from the TSSs (+1) for *GhLIPN-2*, however, to facilitate the alignment between the promoters, 1441 bp of the *GhLIPN-1* promoter was included in the analysis. A canonical TATA-box was not identified in the analysed promoter sequences of either gene. Despite the substantial difference in length between the analysed portions of the promoters, most of the identified putative *cis*-elements related to defense -, stress - and hormone responses were present in both the *GhLIPN-1* and *-2* promoters (Table 2 and Additional file 7, Tables S3A-C).

To confirm the up-regulation of the *GhLIPN* transcripts by the CWD *V. dahliae* elicitor, and to investigate the time period of the elicitation, cotton cell suspensions from the resistant OR_{19} cv were treated with the elicitor over a 10 h period and transcription of *GhLIPN* was monitored with qRT-PCR. Transcription of *GhLIPN* increased up to 6 h, whereby the maximum average fold-change (FC) was more than at four times higher than the control at 0 min (Figure 4). Transcription of *GhLIPN* increased only slightly after 2 h induction with the elicitor, but had more than doubled between 2 and 4 h post-elicitation. The induced response was transient as after a maximal FC was observed at 6 h, *GhLIPN* transcription levels decreased to a level comparable to 4 h at 8 h post-elicitation. Although a slight increase in transcription was observed between 8 and 10 h post-treatment, it was not significant enough to verify a second marked up-regulation of *GhLIPN* .

Discussion

This is the first study that describes the genomic architecture of lipin homoeologs in a polyploid plant, allotetraploid cotton. We describe the structure of the promoter regions, complete cDNA structure, exon-intron

organization, copy number and the up-regulation of *GhLIPN* in response to a CWD *V. dahliae* elicitor. Based on our findings, we propose that in addition to their purported roles in lipid metabolism, lipins may also be involved in a defence role in cotton against *V. dahliae*.

GhLIPN-1 and -2 **are lipin genes**

During the sequence acquisition and alignments of the genomic *GhLIPN* sequences, it became clear that two highly homologous sequences (designated *GhLIPN-1* and -2) were being differentially amplified. The possibility of the gene copies representing paralogs (i.e. members of a gene family) is unlikely since the Southern blot, produced only a single band (Figure 3) for each of three digests. A single band is consistent with the gene sequences as there are identical resitiction sites for the enzymes used in both the two gene copies. The gene copies in *G. hirsutum* can thus be represented as AADD, since both copies from each parent are generally homozygous (Wendel and Cronn 2003; van Deynze et al. 2009). This suggests that the *GhLIPN* gene copies are likely homoeologs that derive from the two separate ancestral genomes within *G. hirsutum*. In contrast, most non-polyploid plant species have two paralogous copies of lipin genes (Nakamura et al. 2009).

The *GhLIPN* homoeologs contain complex intron-exon structures, with a high level of conservation in both the size and position of the introns, and conformed to GT-AG splice junctions (Breathnach and Chambon 1981). The first *GhLIPN* intron interrupts the 5'-UTR of the homoeologs, and consequently, the first exon is in the 5'-UTR and is therefore noncoding. An examination of the gene structure of sequenced lipins in the NCBI database revealed that a noncoding exon in the 5'-UTR appears to be a conserved feature in the lipins of more evolved eukaryotes. Peterfy et al., (2005) reported alternatively spliced lipin isoforms in mouse adipocytes that exhibit distinct expression patterns, subcellular localization, and complementary roles (Csaki and Reue, 2010). In this study, with undifferentiated cultured cotton cells, no indication of alternative splicing of the *GhLIPN* homoeologs were observed.

An investigation of the degree of sequence divergence between homoeologous genes provides insight into the evolutionary dynamics at play between duplicated loci since reciprocal recombination, gene conversion, and other forms of non-independence among homoeologues remain evolutionary possibilities (Small and Wendel 2002; Salmon et al. 2010). An alignment of the *GhLIPN-1* and -2 genomic sequences indicated multiple putative homoeo-SNPs (Salmon et al. 2010), and several indels, varying in size from between 1-3, to 13 and 17 bp. None of the indels occur in the coding exons, and the translated nucleotide sequences of both homoeologs

have an equal number of amino acids. Overall, the sequence divergence between the ungapped genomic DNA of the homoeologs was 2.5%. Most of the putative homoeo-SNPs in the translated *GhLIPN* gene sequences resulted in non-synonymous mutations (40), while comparatively fewer non-synonymous mutations were observed (18) (Additional file 3: Figure S3), which suggests that the homoeologs may have undergone differential selection during evolution.

The results of this study reveal a high sequence conservation in the ungapped *GhLIPN* genomic sequences (97.50%). Additionally, the putative homoeo-SNPs differentiating the *GhLIPN* gene copies were distributed fairly evenly across the A_T and D_T *GhLIPN* genomic sequences, indicating that sequence conservation is high in both introns and exons alike. However, a comparison of the sequence divergence of the ungapped regions spanning the introns and exons within the *GhLIPN* ORFs indicates that the sequence diversity of the introns (2.97%) is slightly more than that of the coding exons (2.27%) as expected. Further, genomic sequences averaged across six amplicons between *G. raimondii* (D-genome) and *G. arboreum* (A-genome) have previously been reported to show ~ 3.7-fold more polymorphisms in introns than exons (Chee et al. 2004).

Genomic origin analysis

The analysis and genome designation of homoeologous gene pairs provides information on the differential evolutionary pressures imposed upon each subgenome (Small et al. 1999). However, due to the shortage of sequence information for both diploid and allotetraploid *Gossypium* species, relatively few homoeologous genes from allotetraploid *Gossypium* have been classified as belonging to either A_T or D_T genomes. The extant species, *G. arboreum* (or *G. herbaceum*) and *G. raimondii*, best represent the ancestral A- and D-genome progenitors, respectively (van Deynze et al. 2009), and sequences from these diploid species are therefore used to determine the parentage of homoeologous sequences from allotetraploids. Given that the A- and D-genomes have undergone little evolution since the polyploidization event 1-2 Mya (Wendel 2000; Salmon et al. 2010) and because they are divergent enough in both diploids and allotetraploids (van Deynze et al. 2009), individual homoeo-SNPs are sufficient to determine genome ancestry (e.g. Small et al. 1999; Senchina et al. 2003; Udall et al. 2006; Udall, 2008; van Deynze et al. 2009; Salmon et al. 2010).

The genomic ancestry of the homoeologs was determined by comparing their sequences to a homologous EST from *G. raimondii* (CO087195), representing the D-genome diploid ancestor and an 811 bp product amplified from *G. herbaceum* genomic DNA, representing the A-genome diploid ancestor. The homoeo-

SNPs from the*GhLIPN-2* homoeolog generally agree with the *G. raimondii* sequence, while those from *GhLIPN-1* consistently (except in one instance) agree with the *G. herbaceum* sequence (Additional file 8: Figure S5A and B). The phylogenetic tree supports these results (Figure 3) and suggests that the *GhLIPN-2* homoeolog belongs to the D_T-genome and *GhLIPN-1* to the A_T-genome of *G. hirsutum*. The D_T genome is consistently more divergent to the D-genome from *G. raimondii* compared to the A_T genome and the A-genome from *G. arboreum* or *G. herbaceum* (Senchina et al. 2003, van Deynze et al. 2009). *G. arboreum* and *G. herbaceum* are therefore better representatives of the A-genome ancestor than *G. raimondii* is of the D-genome ancestor (Senchina et al. 2003). Moreover, the evolutionary rate of divergence in the D_T genome is often significantly higher than that of the A_T genome, although the specific divergence rate does appear to be gene-specific (Small and Wendel 2002; Senchina et al. 2003, van Deynze et al. 2009).

In silico protein analysis

The GhLIPN homoeologs contain an N-terminal lipin domain (N-LIP) and a HAD domain within their C-LIP, which are distinctive features of lipins (Figure 1A). These domains are highly conserved among plant lipin orthologs (Additional file 5, Figure S4) and this conservation is maintained across diverse eukaryotic species (Péterfy et al. 2001; Reue and Zhang 2008). Mutations within these domains are linked to various metabolic diseases in mammals (Péterfy et al. 2001; Donkor et al. 2009), signifying their functional importance (Reue and Zhang 2008). The HAD domain of the GhLIPN homoeologs contains a catalytic active site with a distinctive signature sequence, DxDxT/V (Carman and Han, 2009), indicating that they are functional PAPs. This highly conserved site is usually present as DVDGT in plants and DIDGT in mammals, and classifies lipins as members of the HAD-like superfamily (Burroughs et al. 2006; Reue and Zhang, 2008). Mutating either or both of the Asp (D) residues renders the PAP activity of lipins inactive (Han et al. 2007; Mietkiewska et al. 2011). Other conserved amino acid residues within the N-LIP and HAD domains have also been identified as crucial for PAP activity in mammals. These residues are a Gly in the N-LIP domain (Gly^{83} in the GhLIPN aa sequences) and a Ser in the HAD domain (Ser^{729} in the GhLIPN aa sequences) (Donkor et al. 2009).

Lipins are primarily soluble proteins occurring in the cytoplasm, although they localize in the membrane fractions (ER and nucleus) under specific circumstances (Péterfy et al. 2005; Liu and Gerace 2009). When expressed in yeast, AtPAH1 is found predominantly in the cytoplasm. However, upon elicitation with oleic acid, AtPAH1 migrates to the nucleus, suggesting that it has a role in regulating gene expression

(Mietkiewska et al. 2011). The bipartite NLS reportedly plays a role in the nuclear localization of lipins (Péterfy et al. 2010). The position of the NLS varies between the eukaryotic taxonomic groups. In mammals it is typically found towards the N-terminus, following the lipin domain, whereas in plants it is located towards the C-terminus, within the HAD domain (Figure 1A and Figure S4). In addition to the HAD active site motif, some mammalian C-LIP domains also contain an leucine-rich LxxIL motif reported to be involved in transcriptional coactivator functions (Reue and Zhang, 2008). This was not present in the *GhLIPN* homoeologs and the *V. vinifera* and *A. thaliana* proteins, although leucine- rich sequences (LLFLS, LLNL, LFPSL and LHTL) do occur downstream of the DVDGT motif (Additional file 5, Figure S4).

Promoter analyses

The determined promoter sequences of the *GhLIPN* homoeologs differ significantly, mostly due to indels (Additional file 7, Figure S6). Most notably, the *GhLIPN-2* promoter (predicted to be from the D_T genome) lacks a 437 bp insert that is present in the *GhLIPN-1* promoter (predicted to be from the A_T genome).

The promoters of homoeologous genes frequently exhibit a certain degree of *cis*-regulatory divergence (Chaudhary et al. 2009). However, despite the sequence discrepancies between the *GhLIPN* promoters, the majority of the putative *cis*-elements identified in the analyzed promoter regions occur in both promoters, although their relative positions to the TSS and frequency of occurrence generally differ (Additional file 7, Figure S6). The *GhLIPN* homoeologs contain numerous *cis*-elements, suggesting that they respond to a large and diverse repertoire of TFs (Additional file 7, Tables S3A-C). There is a positive correlation between genes that respond to diverse stimuli and the amount of *cis*-elements in their promoters, with genes functioning in transcriptional regulation, stress responses and signalling processes containing the most *cis*-elements (Walther et al. 2007). This is in line with the recently suggested role of plant lipins functioning in the regulation of gene expression (Mietkiewska et al. 2011). A distinctive feature of the *GhLIPN* promoters was the number of potential stress-responsive *cis*-elements, particularly those involved in plant defence and elicitation (Table 2 and Additional file 7: Table S3), which suggests that they have a defence-associated gene expression.

Two archetypical defence-related *cis*-elements found in both *GhLIPN* promoters are the W-boxes (TTGAC[C/T]), which bind WRKY TFs, and GCC boxes (AGCCGCC), which bind ethylene-response factors (ERFs). WRKY TFs have become iconic defence-associated TFs since they are over-represented in defence-related gene promoters (van den Burg and Takken 2009). Furthermore, several studies have shown that systemic

acquired resistance (SAR)-induced genes, co-regulated with *PR-1*, have an over-representation (~ 4.30) of W-box or W-box-like motifs in their promoters (Maleck et al. 2000; Eulgem 2005). The *GhLIPN-1* and *-2* homoeologs fulfil this criterion with six and three non-overlapping W-boxes in their promoters. The promoters also each contain two conserved ERF-binding GCC boxes which play a key role in elicitor-induced activation of defence genes during plant-pathogen interactions (Rushton et al. 2002). ERF TFs are regulated by ethylene signalling, typically as part of the plant-defence response (Ohme-Takagi et al. 2000). However, other defence-signalling hormones, mostly jasmonic acid, but also salicylic acid, are also reportedly involved in ERF regulation (Lorenzo and Solano 2005; Leon-Reyes et al. 2009). In addition, both *GhLIPN* promoters contain *cis*-element sites for several other defence-associated TF families, namely, MYB, MYC, DOF and GT1 TFs. Several members of these TF families regulate defence-related genes (Jalali et al. 2006).

Several abiotic stress-responsive *cis*-elements were also detected in the *GhLIPN* promoters (Additional file 7: Table S3C). These *cis*-elements are mostly dehydration- and cold-responsive, and hormonally regulated by ABA. There is ample evidence that demonstrates the complicated networks of TFs involved in regulating and integrating multiple stress stimuli of both a biotic and abiotic nature (Cheong et al. 2002; Wyrick and Young 2002). This elaborate cross-talk between diverse regulatory pathways (Koorneef and Pieterse 2008; Huang et al. 2012) results in considerable overlap with regards to the different *cis*-elements, and their corresponding TFs, regulating the different stress signalling pathways.

Expression dynamics of *GhLIPN* in response to elicitation with the cell wall-derived *Verticillium dahliae* elicitor and the potential role of *GhLIPN* in plant defence signalling

Transcriptional bias is frequently observed among the homoeologs of polyploids due to subfunctionalization (Chaudhary et al. 2009; Flagel and Wendel 2010.) Howcever, stress conditions can exert differential effects on homoeolog gene expression, which varies according to the gene, stress and organ type (Dong and Adams 2011).

The up-regulation of *GhLIPN* transcription in response to the *V. dahliae* elicitor was confirmed using qRT-PCR (Figure 4). The transcription was significantly up-regulated after 4 h, and peaked at 6 h post-elicitation with an approximate 4.4 fold-induction. This transient expression pattern is similar to that of an early-response defence gene and supports the observations made on the presence of *cis* elements involved in elicitor- and defense responses. Due to the highly homologous nucleotide sequences, the primer pair used was not able to distinguish between the transcripts from the two homoeologs. Although recent results indicate that genome-

wide expression level dominance between homeologs in AADD allotetrapoid cotton is biased towards the A-genome (Yoo et al. 2013), and thus *GhLIPN-1*, the co-amplification of transcripts originating from the *GhLIPN-2* homoeolog cannot be excluded.

Plant lipins perform key roles in the reorganisation of glycerolipids by hydrolysing phospholipase D (PLD) generated phosphatidic acid (Nakamura et al. 2009; Eastmond et al. 2010). PA is a phospholipid signalling molecule and its levels constantly fluctuate in response to various stimuli (Testerink and Munnik 2005). Both PA and PLD have been implicated in various biotic and abiotic plant stress responses which include pathogen infection and oxidative stress (de Torres Zabella et al. 2002; Testerink and Munnik 2005). In addition, PA and certain classes of PLDs are up-regulated in response to various elicitors (den Hartog et al. 2003; Yamaguchi et al. 2005; Li et al. 2009). PA and PLD activity often mediate hormonal signalling and have been reportedly associated with ABA and ET-triggered stress responses (Fan et al. 1997; Jacob et al. 1999). The specific timing of PLDs and subsequent PA production does, however, differ depending on the class of PLD, the specific elicitor, the type of plant material assayed and the conditions of elicitation (de Torres Zabella et al. 2002; den Hartog et al. 2003; Testerink and Munnik, 2005; Li et al. 2009).

It is vital for signalling molecules to be rapidly down-regulated to pre-stimulation levels following signal transduction, in order to tightly couple their response to the stress stimuli (Munnik 2001). Therefore, the concomitant transcriptional up-regulation of a PA attenuating gene is expected following PA signal transduction. The timing of the transcriptional up-regulation of *GhLIPN* in response to elicitation suggests that the PAP activity may be responsible for the attenuation of PLD-derived PA in defence signalling. This possibility is consistent with the observed up-regulation of *AtPAH1* and *AtPAH2* in early defence responses to *Pseudomonas syringae*, *Blumeria graminis* and salicylic acid treatment (https://www.genevestigator.com).

Conclusions

Previous studies have shown that lipins from *A. thaliana* are involved in ER-mediated lipid remodelling and it is likely that the *GhLIPN* homoeologs perform similar functions in cotton. However, the results obtained in the present study add a new dimension to the proposed roles of lipins in plants by suggesting that lipins may have a role in defence signalling. Specifically, we suggest that, based on prevalence of defence-related *cis*-elements in its promoter, and its expression profile in response to elicitation, *GhLIPN* may play a role in the defence signalling responses of cotton against *V. dahliae*, possibly through attenuation of the PLD-derived PA stress

signal. This intriguing aspect warrants further investigation and together with further studies involving the genome-specific expression of *GhLIPN* will add much needed information to the changes in the cotton transcriptome in response to pathogens.

Acknowledgements

This work was supported by the South African National Research Foundation and the University of Johannesburg.

References

Adams K, Wendel J (2005) Novel patterns of gene expression in polyploid plants. *Trends Genet* 21:539-543

Adams K, Flagel L, Wendel J (2009) Responses of the cotton genome to polyploidy. In: Paterson A (ed) *Genetics and Genomics of Cotton*. Springer, New York, pp 419-429

Aguado A, de Los Santos B, Gamane D, García del Moral L, Romero F (2010) Gene effects for cotton-fiber traits in cotton plant (*Gossypium hirsutum* L.) under *Verticillium* conditions. *Field Crop Res* 116:209-217

Ainouche M, Jenczewski E (2010) Focus on polyploidy. *New Phytol* 186:1-4

Bolek Y, El-Zik K, Pepper A, Bell A, Magill C, Thaxton P, Reddy O (2005) Mapping of Verticillium wilt resistance genes in cotton. *Plant Sci* 168:1581-1590

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu Rev Biochem* 50:349-383

Burroughs A, Allen K, Dunaway-Mariano D, Aravind L (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol* 361:1003-1034

Carman G, Han G-S (2009) Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. *J Biol Chem* 284:2593-2597

Chaudhary B, Flagel L, Stupar R, Udall J, Verma N, Springer N, Wendel J (2009) Reciprocal silencing, transcriptional bias and functional divergence of homoeologs in polyploid cotton (*Gossypium*). *Genetics* 182:503-517

Chee P, Rong J, Williams-Coplin D, Schulze S, Paterson A (2004) EST derived PCR-based markers for functional gene homologues in cotton. *Genome* 47:449-462

Cheong Y, Chang H-S, Gupta R, Wang X, Zhu T, Luan S (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *Plant Physiol* 129:661-677

Cronn R, Small R, Wendel J (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc Natl Acad Sci USA* 96:14406-14411

Csaki LS, Reue K (2010) Lipins: multifunctional lipid metabolism proteins. *Annu Rev Nutrition* 30: 257-272

den Hartog M, Verhoef N, Munnik T (2003) Nod factor and elicitors activate different phospholipid signaling pathways in suspension-cultured alfalfa cells. *Plant Physiol* 132:311-317

de Torres Zabella M, Fernandez-Delmond I, Niittyla T, Sanchez P, Grant M (2002) Differential expression of genes encoding *Arabidopsis* phospholipases after challenge with virulent or avirulent *Pseudomonas* isolates. *Mol Plant Microbe Interact* 15:808-816

Donkor J, Zhang P, Wong S, O'Loughlin L, Dewald J, Kok B, Brindley D, Reue K (2009) A conserved serine residue is required for the phosphatidate phosphatase activity but not the transcriptional coactivator functions of lipin-1 and lipin-2. *J Biol Chem* 284:29968-29978

Dong S, Adams KL (2011) Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol* 190:1045-1057

Dubery I, Slater V (1997) Induced defence responses in cotton leaf discs by elicitors from *Verticillium dahliae*. *Phytochemistry* 44:1429-1434

Eastmond PJ, Quettier A-L, Kroon JTM, Craddock C, Adams N, Slabas AR (2010) Phosphatidic acid phosphoryhydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in *Arabidposis*. *Plant Cell* 22:2796-2811

Eulgem T (2005) Regulation of the *Arabidopsis* defence transcriptome. *Trends Plant Sci* 10: 71-78.

Fan L, Zheng S, Wang X (1997) Antisense suppression of phospholipase Dα retards abscisic acid- and ethylene-promoted senescence of postharvest *Arabidopsis* leaves. *Plant Cell* 9:2183-2196

Flagel L, Chen L, Chaudhary B, Wendel J (2009) Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. *J Hered* 100:487-490

Flagel L, Wendel J (2009) Gene duplication and evolutionary novelty in plants. *New Phytol* 183:557-564

Flagel L, Wendel J (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol* 186:184-193

Göre M, Caner O, Altin N, Aydin M, Erdogan O, Filizer F, Büyükdögerlioglu A (2009) Evaluation of cotton cultivars for resistance to pathotypes of *Verticillium dahliae*. *Crop Prot* 28:215-219

Gorjánácz M, Mattaj IW (2009) Lipin is required for efficient breakdown of the nuclear envelope in *Caenorhabditis elegans*. *J Cell Sci* 122:1963-1969

Han G-S, Siniossoglou S, Carman G (2007) The cellular functions of the yeast lipin homologue Pah1p are dependent on its phosphatidate phosphatase activity. *J Biol Chem* 51:37026-37035

Huang J-C, Piater LA, Dubery IA (2012) Identification and characterization of a differentially expressed NAC transcription factor gene in MAMP-treated *Arabidopsis thaliana*. *Physiol Molec Plant Pathol* 80:19-27

Jackson S, Chen Z (2009) Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol* 13:1-7

Jacob T, Ritchie S, Assmann S, Gilroy S (1999) Abscisic acid signal transduction in guard cells is mediated by phospholipase D activity. *Proc Natl Acad Sci USA* 96:12192-12197

Jalali B, Bhargava S, Kamble A (2006) Signal transduction and transcriptional regulation of plant defence responses. *Phytopathology* 154:65-74

Klosterman S, Atallah Z, Vallad G, Subbarao K (2009) Diversity, pathogenicity, and management of *Verticillium* species. *Annu Rev Phytopathol* 47:39-62

Koorneef A, Pieterse C (2008) Cross talk in defence signaling. *Plant Physiol* 146:839-844

Larionov A, Krause A, Miller W (2005) A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics* 6:62-77

Leon-Reyes A, Spoel S, de Lange E, Abe H, Kobayashi M, Tsuda S, Millenaar F, Welschen R, Ritsema T, Pieterse C (2009) Ethylene modulates the role of nonexpressor of pathogenesis-related genes1 in cross talk between salicylate and jasmonate signaling. *Plant Physiol* 149:1797-1809

Li M, Hong Y, Wang X (2009) Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochim Biophys Acta* 1791:927-935

Liu G-H, Gerace L (2009) Sumoylation regulates nuclear localization of lipin-1α in neuronal cells. *PLoS ONE* 4:e7031 doi:10.1371/journal.pone.0007031

Lorenzo O, Solano R (2005) Molecular players regulating the jasmonate signalling network. *Curr Opin Plant Biol* 8:532-540

Maleck K, Levine A, Eulgem T, Morgan A, Schmidt J, Lawton K, Dangl J, Dietrich R (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature* 26:403-410

Mietkiewska E, Siloto RMP, Dewald J, Shah S, Brindley DN, Weselake RJ (2011) Lipins from plants are phosphatidate phosphatases that restore lipid synthesis in a *pah1Δ* mutant strain of *Saccharomyces cerevisiae*. *FEBS J* 278:764-775

Moore R, Purugganan M (2005) The evolutionary dynamics of plant duplicate genes *Curr Opin Plant Biol* 8:122-128

Munnik T (2001) Phosphatidic acid an emerging plant lipid second messenger. *Trends Plant Sci* 6:227-233

Murray M, Thompson W (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321-4325

Nakamura Y, Koizumi R, Shui G, Shimojima M, Wenk M, Ito T, Ohta H (2009) *Arabidopsis* lipins mediate eukaryotic pathway of lipid metabolism and cope critically with phosphate starvation. *Proc Natl Acad Sci USA* 106:20978-20983

Ohme-Takagi M, Suzuki K, Shinshi H (2000) Regulation of ethylene-induced transcription of defence genes. *Plant Cell Physiol* 41:1187-1192

Péterfy M, Phan J, Xu P, Reue K (2001) Lipodystrophy in the *fld* mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nature Genet* 27:121-124

Péterfy M, Phan J, Reue K (2005) Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis. *J Biol Chem* 280:32883-32889

Péterfy M, Harris T, Fujita N, Reue K (2010) Insulin-stimulated interaction with 14-3-3 promotes cytoplasmic localization of lipin-1 in adipocytes. *J Biol Chem* 285:3857-3864

Phillips SM, Dubery IA, van Heerden H (2012) Molecular characterization of an elicitor- responsive Armadillo repeat gene from cotton. *Mol Biol Rep* 39:8513-8523

Phillips SM, Dubery IA, van Heerden H (2013) Molecular characterization of an elicitor-responsive lectin receptor-like kinase gene (GhLec-RK2) from cotton, *Gossypium hirsutum*. *Plant Mol Biol Rep* 31: 9-20

Reue K, Zhang P (2008) The lipin family: dual roles in lipid biosynthesis and gene expression. *FEBS Lett* 582:90-96

Rushton P, Reinstädler A, Lipka V, Lippok B, Somssich I (2002) Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound -induced signalling. *Plant Cell* 14:749-762

Salmon A, Flagel L, Ying B, Udall J, Wendel J (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol* 186:123-134

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory Press, Cold Spring Harbour

Senchina D, Alvarez I, Cronn R, Liu B, Rong J, Noyes R, Paterson A, Wing R, Wilkins T, Wendel J (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol Biol Evol* 20:633-643

Small R, Ryburn J, Wendel J (1999) Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol Biol Evol* 16:491-501

Small R, Wendel J (2002) Differential evolutionary dynamics of duplicated paralogous *Adh* loci in allotetraploid cotton (*Gossypium*). *Mol Biol Evol* 19:597-607

Tamura K, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739

Testerink C, Munnik T (2005) Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends Plant Sci* 10:368-375

Udall J, Swanson J, Nettleton D, Percifield R, Wendel J (2006) A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics* 173:1823-1827

Udall J (2008) The *Gossypium* transcriptome. In: *Genetics and Genomics of Cotton*. Paterson A (ed) Springer, New York, pp 157-186

van den Burg H, Takken F (2009) Does chromatin remodeling mark systemic acquired resistance? *Trends Plant Sci* 14:286-294

van Deynze A, Stoffel K, Lee M, Wilkins T, Kozik A, Cantrell R, Yu J, Kohel R, Stelly D (2009) Sampling nucleotide diversity in cotton. *BMC Plant Biol* 9:125

Walther D, Brunnemann R, Selbig J (2007) The regulatory code for transcriptional response diversity and its relation to genome structural properties in *A. thaliana. PLoS Genet* 3:e11

Wendel J (2000) Genome evolution in polyploids. *Plant Mol Biol* 42:225-249

Wendel J, Cronn R (2003) Polyploidy and the evolutionary history of cotton. *Adv Agron* 78:139-186

Wong M, Medrano, J (2005) Real-time PCR for mRNA quantification. *Biotech* 39:75-85

Wyrick J, Young R (2002) Deciphering gene expression regulatory networks. *Curr Opin Genet Dev* 12:130-136

Yamaguchi T, Minami E, Ueki J, Shibuya N (2005) Elicitor-induced activation of phospholipases plays an important role for the induction of defence responses in suspension-cultured rice cells. *Plant Cell Physiol* 46:579-587

Yoo M-J, Szadkowski E, Wendel JH (2013) Homeoelog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110:171-180

Zwiegelaar M (2003) DDRT-PCR analysis of defence-related gene induction in cotton. *Dissertation*. *RAU University*

Figure legends

Figure 1 A: Exon-intron structure of GhLIPN. Both gene copies contain 11 exons (labelled 1-11 with sizes) and same the same exon-intron junctions. The C1B10 expressed sequence tag (EST) is represented by ⟼. The location of the probe used for Southern blot analysis is indicated with an arrow. **B:** Predicted domain structure of the GhLIPN homoeologs. Analysis of the GhLIPN-1 and -2 domain structures reveals that they contain a conserved lipin domain at the N-terminus and a haloacid dehalogenase (HAD)-like domain in the C-LIP, towards the end of the protein. A bipartite nuclear localization signal (BNLS) occurs within the HAD-like domain.

Figure 2: Phylogenetic tree showing the relationship between the *GhLIPN-1* and *GhLIPN-2* gene copies and their ancestral genomic representatives from *Gossypium raimondii* and *Gossypium herbaceum*. The tree was constructed with neighbourhood joining analysis from MEGA5 of a 500 bp multiple nucleotide sequence alignment (Additional file 6: Figure S5A).

Figure 3: Southern blot analysis to determine the copy number of *GhLIPN*. Cotton genomic DNA (30 µg) was digested with: (1) *Xba*I, (2) *Eco*RI and (3) *Hind*III. The approximate sizes of the bands are shown.

Figure 4: Real-time relative quantitation PCR analysis to confirm induction of *GhLIPN* following elicitation with the *V. dahliae*-derived elicitor. Cotton cell suspensions were treated with 5 µg/ml *V. dahliae* elicitor and RNA was isolated from treated suspensions at the given time points. Error bars represent the SEM of two biological repeats and two technical repeats (n = 4). Significant differences at $P < 0.05$ between the treated samples and the 0 h control (calibrator, designated as 1-fold) are indicated with asterisks.

Description of additional data files

Additional file 1: Supplementary Table S1: *GhLIPN* gene-specific primers used in the genome walking reactions, 5' and 3' RACE reactions, and internal PCRs to characterize the *GhLIPN* gene copies and their promoter regions. A description of all the gene-specific primers used the study is provided.

Additional file 2: Supplementary Figure S1: Genomic sequence containing the *GhLIPN-1* transcript and promoter. This file contains the full genomic sequence containing the *GhLIPN-1* gene and promoter sequence obtained from genome walking and homoeolog-specific PCR amplification. The putative transcription and

translation start and end sites, and exon-intron structure is indicated. The 17 bp insert in exon 1 (not present in the *GhLIPN-2* transcript) that was used for homoeolog-specific amplification of the *GhLIPN-1* promoter is also shown.

Supplementary Figure S2: Genomic sequence containing the *GhLIPN -2* transcript and promoter. This file contains the full genomic sequence containing the *GhLIPN-2* gene and promoter sequence obtained from genome walking and homoeolog-specific PCR amplification. The putative transcription and translation start and end sites, and exon-intron structure is indicated. The 13 bp insert located in intron 6 (not present in the *GhLIPN-1* transcript) that was used for homoeolog-specific amplification of *GhLIPN-2* is also shown.

Additional file 3: Supplementary Figure S3: Comparison of the full-length cDNAs and deduced amino acid sequences of *GhLIPN-2* and *GhLIPN-1* encoded lipin proteins in *Gossypium hirsutum*.

Additional file 4: Supplementary Table S2: Predicted putative post-translational modification sites in the translated GhLIPN gene copies. Potential post-translational modification sites were predicted from bioinformatic analyses, as described in the Materials and Methods.

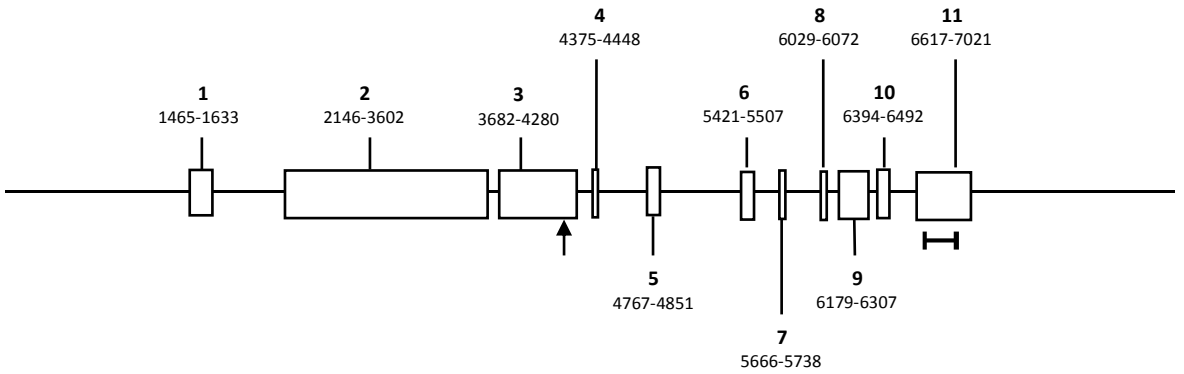
Additional file 5: Supplementary Figure S4: ClustalW alignment of the GhLIPN gene copies with proteins from *Vitis vinifera* and *Arabidopsis thaliana*, the top scoring lipin domain-containing proteins from a NCBI PHI-BLAST.

Additional file 6. Supplementary Figure S5: Alignment of the *GhLIPN* gene copies with a homologous partial mRNA sequence (Genbank: CO087195) from *Gossypium raimondii* and an amplified region from *Gossypium herbaceum*, to determine their genomic origin. **A:** nucleotide alignment. **B:** amino acid alignment.

Additional file 7: Supplementary Table S3: *In silico* analysis of putative *cis*-elements in the *GhLIPN-1* and *-2* promoter sequences.

Figure 1

A


1
1465-1633
2
2146-3602
3
3682-4280
4
4375-4448
5
4767-4851
6
5421-5507
7
5666-5738
8
6029-6072
9
6179-6307
10
6394-6492
11
6617-7021

B


BNLS
775-791
1
105
630
672
829
858
N-terminal lipin
C-LIP
HAD-like domain

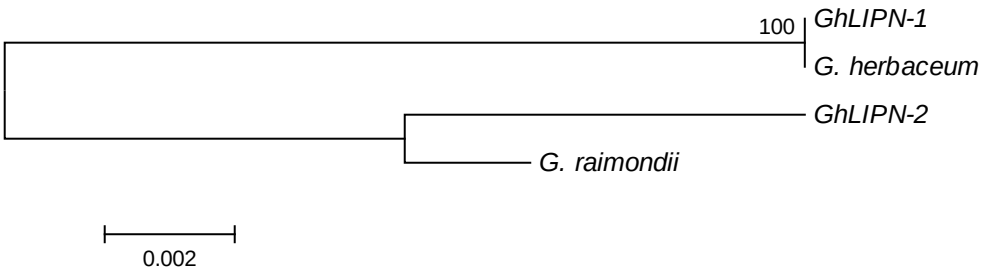


Figure 2: Phylogenetic tree of the *GhLIPN* gene copies and their ancestral representatives from *Gossypium raimondii* and Gossypium *herbaceum*, to determine their putative genomic origin. The tree was constructed with neighbourhood joining analysis in MEGA5 of a 500 bp multiple nucleotide sequence alignment (Additional file 6: Figure S5A).

C

Figure 3

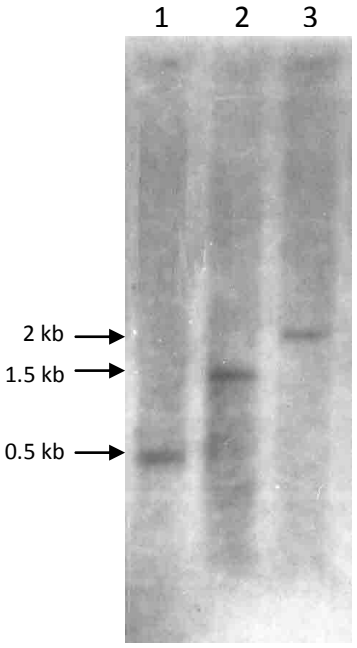


Figure 4

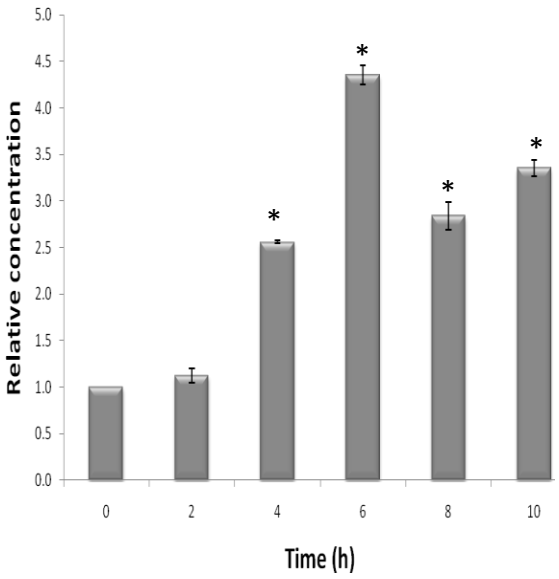

Relative concentration
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
*
*
*
*
0
2
4
6
8
10
Time (h)

Table 1:

Lipin domain-containing proteins from various plant species that exhibit the most homology to the GhLIPN homoeologs -1 and -2, identified from a NCBI PSI-BLAST. Only one lipin-like protein (with the highest E-score in an alignment with the GhLIPN homoeologs) is shown for each of the 11 plant species that had lipin proteins with the highest homology to the GhLIPN homoeologs.

Plant species	**Accession number**	**E-value**		**Identity**	
		-1	***-2***	***-1***	***-2***
Vitis vinifera	XP_002274246	0	0	568/919 (61 %)	574/919 (62 %)
Arabidopsis thaliana	NP_187567 (AtPAH1)	0	0	480/922 (52 %)	483/932 (51 %)
Populus trichocarpa	XP_002323436	$4e^{-179}$	$4e^{-180}$	301/400 (75 %)	302/400 (75 %)
Sorghum bicolor	XP_002441227	$6e^{-166}$	$1e^{-166}$	276/415 (66 %)	277/415 (66 %)
Zea mays	NP_001146282	$7e^{-166}$	$1e^{-166}$	275/415 (66 %)	276/415 (66 %)
Oryza sativa	EAY98334	$1e^{-164}$	$3e^{-165}$	276/415 (66 %)	276/419 (65 %)
Ricinus communis	XP_002510239	$1e^{-139}$	$8e^{-141}$	241/423 (56 %)	243/434 (55 %)
Physcomitrella patens	XP_001751742	$5e^{-116}$	$4e^{-116}$	212/373 (56 %)	212/376 (56 %)
Micromonas pusilla	EEH58259	$5e^{-77}$	$5e^{-76}$	132/267 (49 %)	130/267 (48 %)
Ostreococcus tauri	CAL54337	$4e^{-72}$	$5e^{-72}$	151/402 (37 %)	151/406 (37 %)
Chlamydomonas reinhardtii	XP_001691011	$2e^{-52}$	$6e^{-52}$	91/155 (58 %)	91/155 (58 %)

Table 2: *In silico* analysis of W-box *cis*-elements in the promoters of the *GhLIPN* homoeologs. The promoter regions were obtained by genome walking upstream from the *GhLIPN-1* and-*2* transcription start sites. Homoeolog-specific promoter amplification was used to distinguish between the two promoters of *GhLIPN-1* and *–2*. The analysed promoter sequences were limited to 1 kb from the predicted TSSs. Identical *cis*-elements occurring in the same positions in the promoters of the two homoeologs are indicated by the position number in bold font.

GhLIPN-1* and *GhLIPN-2						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
WBOXATNPR1	+	-882[2]	TTGAC	Binding site for pathogen-induced WRKY transcription factors, in promoter of SAR-regulated NPR1 gene	*Arabidopsis thaliana*	PLACE, PlantCARE
	-	-306[2], -291[1]				
WBOXHVISO1	+	**-29[B]**	TGACT	Sugar signalling, binds to a novel sugar-responsive WRKY transcription factor, SUSIBA2	*Hordeum vulgare*	PLACE
	-	-307[2], -292[1]				
WBOXNTCHN48	+	**-30[B]**	CTGACY	Specifically binds elicitor-induced WRKY transcription factors, involved in elicitor-responsive transcription of defence genes	*Nicotiana tabacum*	PLACE
WBOXNTERF3	+	-881[2], **-29[B]**	TGACY	Involved in activation of ethylene response factor 3 (ERF3), a transcriptional repressor, by wounding	*N. tabacum*	PLACE
	-	-307[2], -292[1]				
WRKY710S	+	**-29[B]**, -881[2]	TGAC	Binds WRKY71, a transcriptional repressor of the gibberellin signalling pathway, specifically binds WRKY proteins within the Pathogenesis-Related Class 10 genes in parsley	*Oryza sativa*, *Petroselinum crispum*	PLACE
	-	-788[1], -454[1], -306[2], -291[1], -124[1]				
WBBOXPCWRKY1/ Box-W1	+	-883[2]	TTTGACY	"WB box" specifically binds elicitor-induced WRKY transcription factors, involved in fungal elicitor-responsive transcription of defence genes	*A. thaliana*, *Avena fatua*, *H. vulgare*, *P. crispum*, *Triticum aestivum*	PLACE, PlantCARE

[B] Both homoeologs, [1] *GhLIPN-1*, [2] *GhLIPN-2*

Additional file 1

Supplementary Table S1: *GhLIPN* gene-specific primers used in the genome walking reactions, 5' and 3' RACE reactions, and internal PCRs to characterize the *GhLIPN* gene, its homoeologs and their promoter regions.

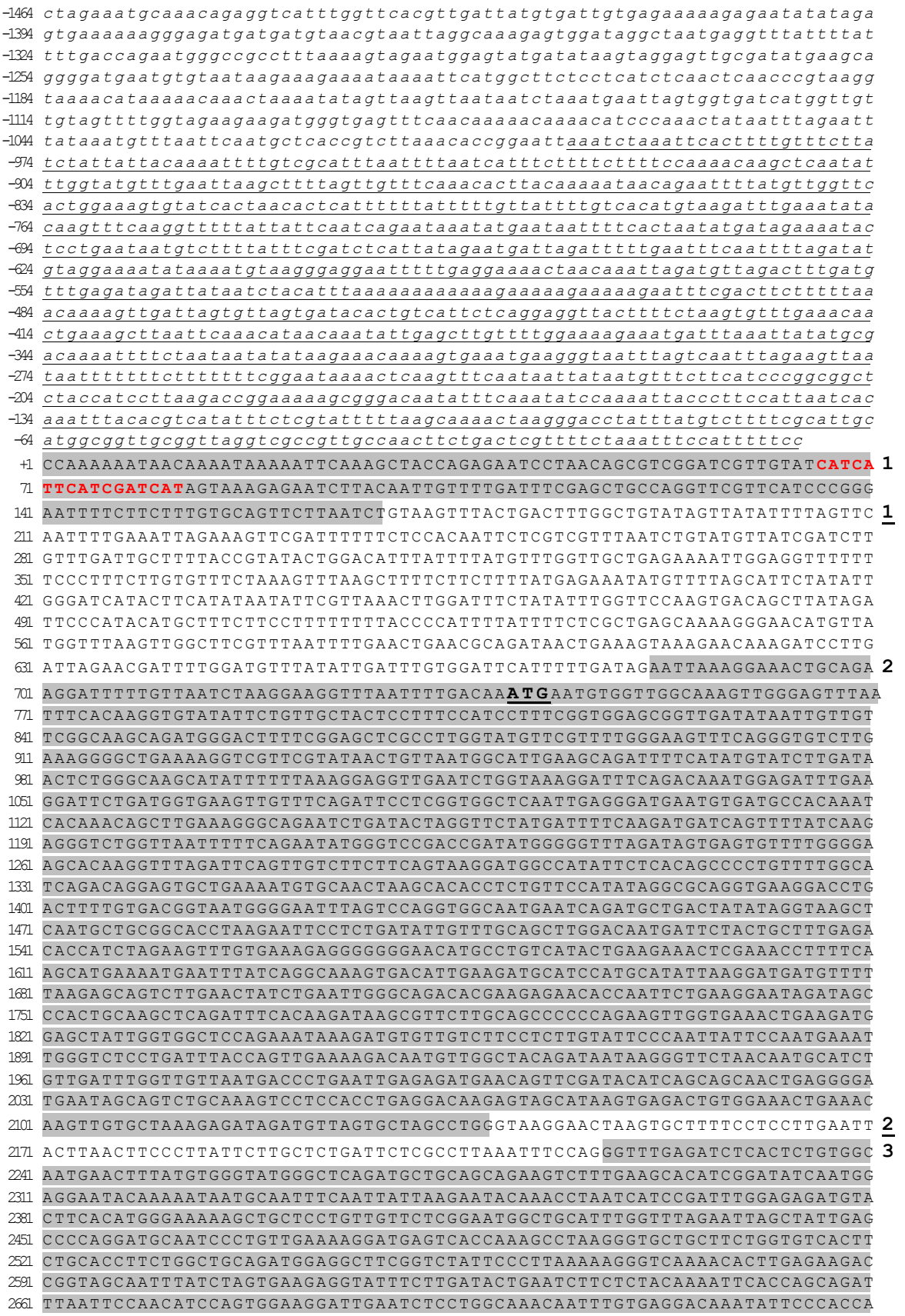
Name	*Sequence* (5' – 3')	*Details*	*Tm*	*GC %*	*Length (nt)*
Genome Walking primers					
First upstream Genome Walk					
R1	CCCCAACAGACCATTTCAAGGTC	primary PCR	65	52	23
R2	CCGTATTCTTTAGGCAATCAAGC	secondary PCR	61	44	23
Second upstream Genome Walk					
GWR1	ACGACGAGAATTGTGGGAAA	primary PCR	58	45	20
GWR2	TTCCCGCGATGAACGAATCT	secondary PCR	60	50	20
Downstream Genome Walk					
F1	ACCAAGGAGATGATCAGTAGC	primary PCR	61	48	21
F2	GGTCTGTTGGGGTTCATCTTG	secondary PCR	63	52	21
5' and 3' RACE primers					
First 5' RACE					
RR3	TTCCTTACAAGATGAACCCCAAC	cDNA synthesis	61	44	23
RR2	CCCCAACAGACCATTTCAAGGTC	primary PCR	65	52	23
RR1	GCTACTGATCATCTCCTTGGT	secondary PCR	61	48	21
Second 5' RACE					
LiDR1	GGAAACAAACCGTCAGGAGA	cDNA synthesis	60	50	20
LiDR2	GAAACTTCTGGTTAGATATGCCTGA	primary PCR	61	40	25
LiDR3	GAAGCTGATACCCGTTCCCCTTA	secondary PCR	65	52	23
Third 5' RACE					
LiDR1	GGAAACAAACCGTCAGGAGA	cDNA synthesis	60	50	20
LiDR2	GAAACTTCTGGTTAGATATGCCTGA	primary PCR	61	40	25
RTR2	CCAGAACCCTGGAGAAGAAA	secondary PCR	60	50	20
Fourth 5' RACE					
LiDR1	GGAAACAAACCGTCAGGAGA	cDNA synthesis	60	50	20
LiDR2	GAAACTTCTGGTTAGATATGCCTGA	primary PCR	61	40	25
5'cDNAR	TCCGAGAACAACAGGAGCAG	secondary PCR	63	55	20
3' RACE					
RF1	GAATGGAATTTTTGGAAAGTGC	primary PCR	57	36	22
RF2	GGAAAGTGCCATTGCAAGATATTGA	secondary PCR	61	40	25
Internal PCR primers					
First PCR					
RTF2	GGATTGAATCTCCTGGCAA	forward	58	45	19
LipinUP2	AATGGATGAGGTTGGTGGAA	reverse	59	43	20
Second PCR					
DNAF	AGCTGCTCCTGTTGTTCTCG	forward	63	55	20
Exon4R	TCATGAGGTGTTCTTCTTATCACTG	reverse	61	40	24
Third PCR					
5'DNAF2	TCAAAGCTACCAGAGAATCCTAACA	forward	61	40	25
5'cDNAR	TCCGAGAACAACAGGAGCAG	reverse	63	55	20
Internal RT-PCR primers					
LiD 1F	CCTTAGTTGGAAGGGATTGGA	Forward	61	48	21
RR3	TTCCTTACAAGATGAACCCCAAC	Reverse	61	44	23

Name	*Sequence* (5' – 3')	*Details*	*Tm*	*GC %*	*Length (nt)*
Homoeolog-specific gene primers [a]					
GhLIPN homoeolog -2					
LipinGW promF1	CACCCCCTATTTTCTCTTTT	Forward	56	40	20
Lipin-2 promR	GATTCTCTT**TACTATAC**AACGATCC	Reverse	60	36	25
GhLIPN homoeolog -1					
LipinGW promF2	TTCACTGTCTTTTCCCTCAC	Forward	58	45	20
Lipin-1 promR	ATGAATGATG**ATAC**AACGATCC	Reverse	57	36	22
Homoeolog-specific promoter primers [b]					
GhLIPN homoeolog -2 promoter					
5'DNAF2	TCAAAGCTACCAGAGAATCCTAACA	forward 5' PCR	61	40	25
AlleleIR	TCACTAATGCCTTTTTGCTTCC	reverse5' PCR	59	41	22
AlleleIF	GGAAGCAAA**AAGGC**ATTAGTGA	forward 3' PCR	59	41	22
RR3	TTCCTTACAAGATGAACCCCAAC	reverse3' PCR	61	44	23
GhLIPN homoeolog -1 promoter					
5'DNAF2	TCAAAGCTACCAGAGAATCCTAACA	forward 5' PCR	61	40	25
AlleleNR	CTCTTTGG**TTAA**A**CCTT**TTTGCTTC	reverse 5' PCR	60	36	25
Allele NF	GAAGCAAA**AAGG**T**TTAA**CCAAAGAG	forward 3' PCR	60	36	25
RR3	TTCCTTACAAGATGAACCCCAAC	reverse 3' PCR	61	44	23
RT-qPCR and Southern blot probe primers					
RT F2	GGATTGAATCTCCTGGCAA	forward	58	45	19
RT R2	CCAGAACCCTGGAGAAGAAA	reverse	60	50	20
***G. herbaceum* primers**					
F1	GCTTTGTTTATGTAGGATGG	forward	56	40	20
R1	GTTCAATGGATGAGGTTG	reverse	58	44	18

[a] The reverse primers are specific for a 13 bp indel which distinguishes between the homoeologs. The nucleotides in red font include 13 bp insert sequence and the sequences surrounding the 13 bp insertion site are in bold font.

[b] The reverse 5' and forward 3' PCR primers are specific for a 17 bp indel which distinguishes between the promoter regions of the homoeologs. The nucleotides in red font include 17 bp insert sequence and the sequences surrounding the 17 bp insertion site are in bold font.

Figure S1


−1464 ctagaaatgcaaacagaggtcatttggttcacgttgattatgtgattgtgagaaaaagagaatatataga
−1394 gtgaaaaaagggagatgatgatgtaacgtaattaggcaaagagtggataggctaatgaggtttattttat
−1324 tttgaccagaatgggccgcctttaaaagtagaatggagtatgatataagtaggagttgcgatatgaagca
−1254 ggggatgaatgtgtaataagaaagaaaataaaattcatggcttctcctcatctcaactcaacccgtaagg
−1184 taaaacataaaaacaaactaaaatatagttaagttaataatctaaatgaattagtggtgatcatggttgt
−1114 tgtagttttggtagaagaagatgggtgagtttcaacaaaaacaaaacatcccaaactataatttagaatt
−1044 tataaatgtttaattcaatgctcaccgtcttaaacaccggaattaaatctaaattcacttttgtttctta
−974 tctattattacaaaatttttgtcgcatttaattttaatcatttcttttctttttccaaaacaagctcaatat
−904 ttggtatgtttgaattaagcttttagttgtttcaaacacttacaaaaataacagaatttttatgttggttc
−834 actggaaagtgtatcactaacactcatttttttatttttgttatttttgtcacatgtaagatttgaaatata
−764 caagtttcaaggtttttattattcaatcagaataaatatgaataatttttcactaatatgatagaaaatac
−694 tcctgaataatgtctttttatttcgatctcattatagaatgattagatttttgaatttcaattttagatat
−624 gtaggaaaatataaaatgtaagggaggaatttttgaggaaaactaacaaattagatgttagactttgatg
−554 tttgagatagattataatctacatttaaaaaaaaaaaagaaaaagaaaaagaatttcgacttctttttaa
−484 acaaaagttgattagtgttagtgatacactgtcattctcaggaggttacttttctaagtgtttgaaacaa
−414 ctgaaagcttaattcaaacataacaaatattgagcttgttttggaaaagaaatgatttaaattatatgcg
−344 acaaaatttttctaataatatataagaaacaaaagtgaaatgaagggtaatttagtcaatttagaagttaa
−274 taatttttttctttttttcggaataaaactcaagtttcaataattataatgtttcttcatcccggcggct
−204 ctaccatccttaagaccggaaaaagcgggacaatatttcaaatatccaaaattacccttccattaatcac
−134 aaatttacacgtcatatttctcgtatttttaagcaaaactaagggacctatttatgtcttttcgcattgc
−64 atggcggttgcggttaggtcgccgttgccaacttctgactcgttttctaaatttccatttttcc
+1 CCAAAAAATAACAAAATAAAAATTCAAAGCTACCAGAGAATCCTAACAGCGTCGGATCGTTGTATCATCA 1
71 TTCATCGATCATAGTAAAGAGAATCTTACAATTGTTTTGATTTCGAGCTGCCAGGTTCGTTCATCCCGGG
141 AATTTTCTTCTTTGTGCAGTTCTTAATCTGTAAGTTTACTGACTTTGGCTGTATAGTTATATTTTAGTTC 1
211 AATTTTGAAATTAGAAAGTTCGATTTTTTCTCCACAATTCTCGTCGTTTAATCTGTATGTTATCGATCTT
281 GTTTGATTGCTTTTACCGTATACTGGACATTTATTTTATGTTTGGTTGCTGAGAAAATTGGAGGTTTTTT
351 TCCCTTTCTTGTGTTTCTAAAGTTTAAGCTTTTCTTCTTTTATGAGAAATATGTTTTAGCATTCTATATT
421 GGGATCATACTTCATATAATATTCGTTAAACTTGGATTTCTATATTTGGTTCCAAGTGACAGCTTATAGA
491 TTCCCATACATGCTTTCTTCCTTTTTTTTACCCCATTTTATTTTCTCGCTGAGCAAAAGGGAACATGTTA
561 TGGTTTAAGTTGGCTTCGTTTAATTTTGAACTGAACGCAGATAACTGAAAGTAAAGAACAAAGATCCTTG
631 ATTAGAACGATTTTGGATGTTTATATTGATTTGTGGATTCATTTTTGATAGAATTAAAGGAAACTGCAGA 2
701 AGGATTTTTGTTAATCTAAGGAAGGTTTAATTTTGACAAATGAATGTGGTTGGCAAAGTTGGGAGTTTAA
771 TTTCACAAGGTGTATATTCTGTTGCTACTCCTTTCCATCCTTTCGGTGGAGCGGTTGATATAATTGTTGT
841 TCGGCAAGCAGATGGGACTTTTCGGAGCTCGCCTTGGTATGTTCGTTTTGGGAAGTTTCAGGGTGTCTTG
911 AAAGGGGCTGAAAAGGTCGTTCGTATAACTGTTAATGGCATTGAAGCAGATTTTCATATGTATCTTGATA
981 ACTCTGGGCAAGCATATTTTTTAAAGGAGGTTGAATCTGGTAAAGGATTTCAGACAAATGGAGATTTGAA
1051 GGATTCTGATGGTGAAGTTGTTTCAGATTCCTCGGTGGCTCAATTGAGGGATGAATGTGATGCCACAAAT
1121 CACAAACAGCTTGAAAGGGCAGAATCTGATACTAGGTTCTATGATTTTCAAGATGATCAGTTTTATCAAG
1191 AGGGTCTGGTTAATTTTTCAGAATATGGGTCCGACCGATATGGGGGTTTAGATAGTGAGTGTTTTGGGGA
1261 AGCACAAGGTTTAGATTCAGTTGTCTTCTTCAGTAAGGATGGCCATATTCTCACAGCCCCTGTTTTGGCA
1331 TCAGACAGGAGTGCTGAAAATGTGCAACTAAGCACACCTCTGTTCCATATAGGCGCAGGTGAAGGACCTG
1401 ACTTTTGTGACGGTAATGGGGAATTTAGTCCAGGTGGCAATGAATCAGATGCTGACTATATAGGTAAGCT
1471 CAATGCTGCGGCACCTAAGAATTCCTCTGATATTGTTTGCAGCTTGGACAATGATTCTACTGCTTTGAGA
1541 CACCATCTAGAAGTTTGTGAAAGAGGGGGGGAACATGCCTGTCATACTGAAGAAACTCGAAACCTTTTCA
1611 AGCATGAAAATGAATTTATCAGGCAAAGTGACATTGAAGATGCATCCATGCATATTAAGGATGATGTTTT
1681 TAAGAGCAGTCTTGAACTATCTGAATTGGGCAGACACGAAGAGAACACCAATTCTGAAGGAATAGATAGC
1751 CCACTGCAAGCTCAGATTTCACAAGATAAGCGTTCTTGCAGCCCCCCAGAAGTTGGTGAAACTGAAGATG
1821 GAGCTATTGGTGGCTCCAGAAATAAAGATGTGTTGTCTTCCTCTTGTATTCCCAATTATTCCAATGAAAT
1891 TGGGTCTCCTGATTTACCAGTTGAAAAGACAATGTTGGCTACAGATAATAAGGGTTCTAACAATGCATCT
1961 GTTGATTTGGTTGTTAATGACCCTGAATTGAGAGATGAACAGTTCGATACATCAGCAGCAACTGAGGGGA
2031 TGAATAGCAGTCTGCAAAGTCCTCCACCTGAGGACAAGAGTAGCATAAGTGAGACTGTGGAAACTGAAAC
2101 AAGTTGTGCTAAAGAGATAGATGTTAGTGCTAGCCTGGGTAAGGAACTAAGTGCTTTTCCTCCTTGAATT 2
2171 ACTTAACTTCCCTTATTCTTGCTCTGATTCTCGCCTTAAATTTCCAGGGTTTGAGATCTCACTCTGTGGC 3
2241 AATGAACTTTATGTGGGTATGGGCTCAGATGCTGCAGCAGAAGTCTTTGAAGCACATCGGATATCAATGG
2311 AGGAATACAAAAATAATGCAATTTCAATTATTAAGAATACAAACCTAATCATCCGATTTGGAGAGATGTA
2381 CTTCACATGGGAAAAAGCTGCTCCTGTTGTTCTCGGAATGGCTGCATTTGGTTTAGAATTAGCTATTGAG
2451 CCCCAGGATGCAATCCCTGTTGAAAAGGATGAGTCACCAAAGCCTAAGGGTGCTGCTTCTGGTGTCACTT
2521 CTGCACCTTCTGGCTGCAGATGGAGGCTTCGGTCTATTCCCTTAAAAAGGGTCAAAACACTTGAGAAGAC
2591 CGGTAGCAATTTATCTAGTGAAGAGGTATTTCTTGATACTGAATCTTCTCTACAAAATTCACCAGCAGAT
2661 TTAATTCCAACATCCAGTGGAAGGATTGAATCTCCTGGCAAACAATTTGTGAGGACAAATATTCCCACCA

```
2731 ACGAGCAGATTGCTTCCTTGAATCTGAAAAATGGTCAAAACATGATTACTTTCAGTTTCTTCTCCAGGGT
2801 TCTGGGAACACAACAGGTATCAATTTTTTCTTTACACAATTCTTCTGGTAGGTTCGCAAGGTTTGTGTTC 3
2871 TGCTTAAATTTTAGCCTAAAAATCTGTGTATCTTTAACAGGTTGAAGCTCATCTTTACTTGTGGAAATGG 4
2941 AACGCAAAGATTGTAATTTCAGATGTGGATGGAACTATTACCAAGTAAGGATTTATCTCCTGATATTGTT 4
3011 GCTCTGCAGTATTTAAAATTTTCCTGTTTCACCCATTTATGGACCTGTCTAAACTAGCTCTGAATCTTAT
3081 TATGATATTATTATGGAAACAGAGAAATTAAACCATATATTCGCATTTGATGCTGAAATTATAACTTCCA
3151 TTATGGTCTTTAGTATTAGAGCAGCAGTGAGCATTCTAGAAGCAGTCATACCCTAATTGATTTGGTGTGA
3221 AATCTTTCTTTCGAATATGTTTTAGTCATGTATAGTGTAATACAAGATGCATGTTTCAAAAACCTTAACA
3291 TATTTCTTTCAGGTCTGATGTCTTAGGCCAGTTTATGCCTTTAGTTGGAAGGGATTGGACACAATCTGGT 5
3361 GTAGCTAATCTTTTCTCAGCTATTAAGGTATGCTTTTCTCAATTGTTTCTTTGTTCAACTTGACTTGACT 5
3431 GAGAAGTTTTTCTTTTCTGGACTATCATGCAAAACTAGATCTAGTAAATTGCTCCACCATTTTTTAGTTT
3501 CACCACAACAGAAATCAAAACTTGTCTCATGTTTCCCTTGCGCTTTACTTTTACTAGTTCTTTATAAAAT
3571 TCCCCACTGATGTAAGTGATGATAAATTTTTTAATGCAAGTTTATAGAGTAACTTTTGATAACCTTAGTA
3641 TCAAGATTATGTGAGCATTATCCTGTCACCATTATGCCCCTCAACATTTATTTTTCTGTTGTGAAATGTA
3711 TTTTTGCACTTCTTAGTTTATATATGGGTAGTGATGAAATTCTTCTCATAAGTTGCATTTGTGTCATCTT
3781 TATTCATCCCTCCTGTAACATGTGTGCATGCACAAACACAATATTGTTTGTAGAGAACAATATCAAACAG
3851 TGTGGCTCTTCTATTTATTCCATTAACCAACATATTTTGGCATGCATATTAACCAATTATTTTCTTGGGT
3921 GTCCGATGTCCAACACTAACTGTATGTTCACAACAGGAGAACGGATATCAGCTTCTATTTCTCAGTGCAC 6
3991 GTGCAATTGTTCAGGCATATCTAACCAGAAGTTTCTTACTTAACCTGAAACAGGTAATGGAATAGCCTGA 6
4061 ATACACTTTTTTACATTTCCCTAATGGAAGCAAAAAGGTTTAACCAAAGAGAAAAGCTAACATTTGGAAA
4131 CCATCTTGTTAAAAGACTTTAAAATCACAAATTTGTAACCGATTTTAACTGATATAGCTTTGTTTATGTA
4201 GGATGGAAAAGCTTTACCCTCTGGACCTGTTGTGATTTCTCCTGACGGTTTGTTTCCCTCATTGTACCGT 7
4271 GAGGGTGAGATCTTTCATTCTTTATTTTCTCTGACATTGTTTGTCTATTGAAGTAATCTAGCTGTTTTTT 7
4341 TAGTTTGCTTCACAATTTTGAGGCTCCAGCTACATGAGAGGAAAAAGAATTCTTATTGTATTCCATCACA
4411 AAACCTTTAAATGATAAACGTGGCATTATTTGTTGTCTTGGCCTTTTCTTTATATCAGAATCATTAGGCT
4481 TTATCATCAAGTTACATTGCTTTTCTTCTTTTACTTCCTTTTTAATTATTTATGGGTTCCATTCTAATAA
4551 ACTATACTTGGCAGTGATAAGAAGAACACCTCATGAATTCAAAATAGCTTGTTTAAAGGTGAATTCTGAA 8 8
4621 ATATTCAACCTTGTTTCTAAATATACAACGGAACTATTACTGCATCCAGATCAGTCTTTTCACCTATGAT
4691 TTCTTTCTTATACATGAATTGCAGAATATCAGGAAACTTTTCCCTTCCGACTACAATCCATTTTATGCAG 9
4761 GCTTTGGAAACAGAGACACAGATGAACTCAGTTATAAAGAAATTGGCATCCCAAAGGGAAAAATATTTAT
4831 TATTAATCCAAAGGTACATGTTTTTCATGTCCTACTTATGAGAAATCTATATTGCAATTACCATATCAAA 9
4901 TTTATAACAATCGGGCTCAAAACTGGCAGGGTGAGGTGGCTGTAAGTTATTGTATGAACACGAGGTCATA 10
4971 CACATCATTGCATACTCTTGTAGAAGAGATGTTTCCACCAACCTCATCCATTGAACTGGTAAGGCATCAA 10
5041 TTTCTCGCCTGTTTATATATGATGCAAAAGAACAATTGAAAATTTATCTACAACTGTCTCTTATTTGTTT
5111 TGGAGCCTTCTTGATATCCTTGATATTTTAATACGTTTGCAGGAAGATTATAACGAATGGAATTTTTGGA 11
5181 AAGTGCCATTGCAAGATATTGAGTAACATAGTTGATTGCCTAAAGAATACGGTTTAGTTATACCAAGGAG
5251 ATGATCAGTAGCGTAATTTGTCCCATAGGTAAGAAATTGTTATGATATTCTTCAAGTTCTTAAACTCGCT
5321 TCCTCACTGACATTCATATTCTTTGGTGACCTTGAAATGGTCTGTTGGGGTTCATCTTGTAAGGAATAAA
5391 TAATTCAAGTGGATAAAATCAAAGCCATTGCCTACAATAGCTATATCTATCTTTAGAGATTAAAATATTC
5461 TCCCCAATTATAAAGTTCAGTAAAAATGATAGCATTTCATACGTAGTTTTCTTTATTATATTCATCCATC
5531 ATTTATAAATGTGGGGTTCATAGAAAAttttaatagatctcacgtggatctattgtgcttacataattgt
5601 tgtataataattattaagtaatatcgttcacaaaatggttcctctttagtaaccatttttaacaatttcat
5671 tttgcgggataagaaataaatgcaacttatgtgtatatgtttaaaaccaagctgttctcagtgttcttca
5741 ggctactaacagtgacctaagaaaagcttgtataggtctttaatggacttatattagaaaagctctgaaa
5811 acccaataaaatgattgtaataattcaaactactgtagactgtaatggtaccaatcttatctggaatcct
5881 gaaatattacattcaaacctgcattcaaaccaatagtactaacaggtgtgtatctatgacacaaaacaaa
5951 aacaacttttgaaggttagtaccataaatccacttgtaaaaggtgtttctgagggcgactcccctgaga
6021 cattggccgcaactattctggggttgggtcgtgtcggactcttgatttatttttttcattcacatgactt
6091 gaaattgagatattgcttaaaagacattaaattgtttatcactcagctcaatagatagggcttggtcttg
6161 aacttgtagatatcaggttggagttgtaattgcaactttgagtcattgtaacagtggctcggttggaaag
6231 cctgagaaggtacatgatgcagccatgccccatgcagacgcgggctggcatgagagttaaaaagtgaaa
6301 aagctaattgccccttcagaaagtgagaactgtggtgtacatgaacaaagcttagggtgctcagaccagca
6371 tcacgtgatcacctctgcttcatttcttgcagtaaattcgactttacccgtatttcagagcagtgatttg
6441 ggttttggatatgggatccttctttcgcatctttatcaatattgtttcatttttcatcattttacccctt
6511 ggttaagatttacatctcaaaaactcttatacactactagatattattaccaatcagcaaaagattgacc
6581 ccgatcacaaggtgacaatataataaaatatcagtctatgttaaatccaaactgttacttaagacagcta
6651 attattatatttttccttgttcattaatttgttgctcttaggttaggtcttagctgatgtgaaacaaaaac
6721 acttggaatctaggtgcaatgttgttattgaatattgcggctggtcttggatcagtgttgtagaattgga
6791 caaagctcaaaagtaaacagttactgcaagaggagacaagctaaaaaagaaaaaccctcgagccttggga
6861 taaactgcatcccaacaccaaaatattggagcttatacttgttttctatttttgggggggatgactctag
6931 tttatgtttgtttttagtgaggtcaacag
```

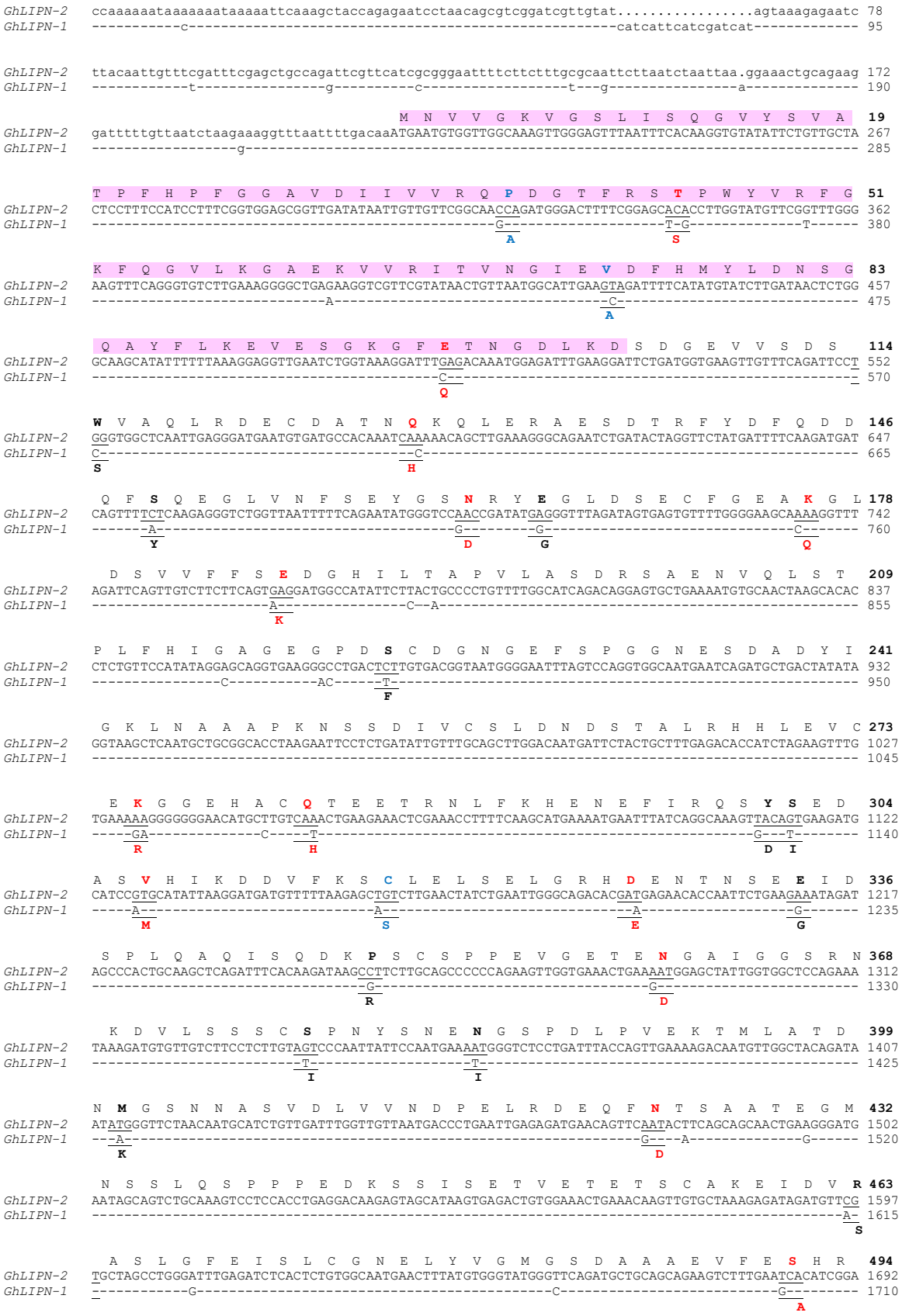
Figure S1: Genomic sequence containing the *GhLIPN-1* transcript and promoter. The genomic sequences preceding and following the putative *GhLIPN-1* transcription start and end sites are in lowercase and the 1 kb promoter sequence analysed for *cis*-elements is underlined. The 11 exons are shaded and numbered 1-11 and the 10 introns are numbered 1-10, with the numbers underlined. The C1B10 EST is indicated with an orange wavy underline (). Translation start and end codons are in large, underlined bold font and putative poly(A) transcript termination signals are double-underlined. A 17 bp insert in exon 1, not present in the *GhLIPN-2* transcript, which was used for homoeolog-specific amplification of the *GhLIPN-1* promoter is in upper case red font.

Figure S2

```
-1289 gtaggctcttcttcctcttcccatcacccctattttctcttttactctcccaaagcatttcactgtctt
-1219 ttccctcacaccttctttccttgaaaacctgtaactcctcaggtacatctgcctacacgagaaactgtcc
-1149 accactctactcggtcggcctgagctccttgacctcaaaaactcagcatccgattctggccacttgtata
-1079 gattcacgtatgttgccctaacagggaccagagggtccgtagcatcggacaagcacctagaaatgcaaac
-1009 agagctcatttttcttaacgctgattatgagattgtgagagaacgagtatatatagagtgaaaaaagagag
 -939 atgatgatgtaacgtaattaggcaacgagtggataggctaacgaggtttatttttattttgaccagaaggg
 -869 gccgcctttaaaagtagattggagtatgatataagtaggagttgcgatgtgaagcaggggatgaatgtgt
 -799 aataaaaaagaaagtaaaattcctgccttctcctcatctcaaccagtcgggtaaaacataaaatctaaat
 -729 tatttagtatttatttaaatgagatttggttggtgatcatggtattatagttttggtagaagaagatggg
 -659 tcttacaaaaaacaaacatcccaaactataatttagattttaatttctattaattacgtgttaaatgcca
 -589 tcttagccacgacacagagcatgatatcttaaacacgggaattaaatctttacttctaaattcacttatt
 -519 tggtcgcatttaattttaatcatttctttttcttttccaaaacaagctcaatatttgttatgtttgaatta
 -449 agctcaggaggtttcttttcacgtgtttgaaacaactaacagcttaattcaaacctaacaaatattgagc
 -379 ttgttttggaattatatgcgaccaaatttttctaataatagacaagaaacaaatgtgaaatgaagggtaat
 -309 ttagtcaatttagatgtaaagatttttttttagcattactgtagaagtttcaataattataatgtttgtt
 -239 cattccggcggctctaccatccttaagcctgaaaaagcgggacaatatttcaaatatcctaaattaccct
 -169 tccattaatcacaaatttacacctcatatttctcgcttcgtagagtctttctaagcaaaactaagggacc
  -99 tatttatgtcttttcacaaggttacttgtttattgcatggcggttgcggttagtcgccgttgccaacttc
  -29 tgactcgttttcaaaatttccatttttttt
    1 CCAAAAAATAAAAAAATAAAAATTCAAAGCTACCAGAGAATCCTAACAGCGTCGGATCGTTGTATAGTAA 1
   71 AGAGAATCTTACAATTGTTTCGATTTCGAGCTGCCAGATTCGTTCATCGCGGGAATTTTCTTCTTTGCGC
  141 AATTCTTAATCTGTAAGCTTACCGACTTTGACTGTATAGTTATATTTTAGTTAAATTTTGAAATTAGAAA 1
  211 GTTCGATTTTTTTTCCCACAATTCTCGTCGTTTAATCTGTATGTTATCAATCTTGTTTGATTGCTTTTAC
  281 TGTATACTGGACATTTATTTTATGTTTGGTTGCTGAGAAAATTGGAGGTTTTTTTTCCCTTGTGTTTCTA
  351 AAGTTTAAGCTTTTCTTCTTTTACGATAAATATGTTTTAGCATTCTATATTGGGATCATACTTCATATAA
  421 TATTCGTTAAACTTGGATTTCTATATTTGGTTCCAAGTGGCAGCTTATAGATTCCCATACATGCTTTCTT
  491 CCTTTTTTTCCCCATTTTATTTTCTTGCTGAGCTAAAGGGAACATGTTATGGTTTAAGTTGGCTTCGTTT
  561 AATTTTGAACTGAACGCAGATACCTGAAAGTAAAGAACAAAGATCCTTGATTAGAACGATTTTGGATGTT
  631 TATATTGATTTGTGGATTCATTTTTTGATAGAATTAAGGAAACTGCAGAAGGATTTTTGTTAATCTAAGA 2
  701 AAGGTTTAATTTTGACAAATGAATGTGGTTGGCAAAGTTGGGAGTTTAATTTCACAAGGTGTATATTCTG
  771 TTGCTACTCCTTTCCATCCTTTCGGTGGAGCGGTTGATATAATTGTTGTTCGGCAACCAGATGGGACTTT
  841 TCGGAGCACACCTTGGTATGTTCGGTTTGGGAAGTTTCAGGGTGTCTTGAAAGGGGCTGAGAAGGTCGTT
  911 CGTATAACTGTTAATGGCATTGAAGTAGATTTTCATATGTATCTTGATAACTCTGGGCAAGCATATTTTT
  981 TAAAGGAGGTTGAATCTGGTAAAGGATTTGAGACAAATGGAGATTTGAAGGATTCTGATGGTGAAGTTGT
 1051 TTCAGATTCCTGGGTGGCTCAATTGAGGGATGAATGTGATGCCACAAATCAAAAACAGCTTGAAAGGGCA
 1121 GAATCTGATACTAGGTTCTATGATTTTCAAGATGATCAGTTTTCTCAAGAGGGTCTGGTTAATTTTTCAG
 1191 AATATGGGTCCAACCGATATGAGGGTTTAGATAGTGAGTGTTTTGGGGAAGCAAAAGGTTTAGATTCAGT
 1261 TGTCTTCTTCAGTGAGGATGGCCATATTCTTACTGCCCCTGTTTTGGCATCAGACAGGAGTGCTGAAAAT
 1331 GTGCAACTAAGCACACCTCTGTTCCATATAGGAGCAGGTGAAGGGCCTGACTCTTGTGACGGTAATGGGG
 1401 AATTTAGTCCAGGTGGCAATGAATCAGATGCTGACTATATAGGTAAGCTCAATGCTGCGGCACCTAAGAA
 1471 TTCCTCTGATATTGTTTGCAGCTTGGACAATGATTCTACTGCTTTGAGACACCATCTAGAAGTTTGTGAA
 1541 AAAGGGGGGGAACATGCTTGTCAAACTGAAGAAACTCGAAACCTTTTCAAGCATGAAAATGAATTTATCA
 1611 GGCAAAGTTACAGTGAAGATGCATCCGTGCATATTAAGGATGATGTTTTTAAGAGCTGTCTTGAACTATC
 1681 TGAATTGGGCAGACACGATGAGAACACCAATTCTGAAGAAATAGATAGCCCACTGCAAGCTCAGATTTCA
 1751 CAAGATAAGCCTTCTTGCAGCCCCCCAGAAGTTGGTGAAACTGAAAATGGAGCTATTGGTGGCTCCAGAA
 1821 ATAAAGATGTGTTGTCTTCCTCTTGTAGTCCCAATTATTCCAATGAAAATGGGTCTCCTGATTTACCAGT
 1891 TGAAAAGACAATGTTGGCTACAGATAATATGGGTTCTAACAATGCATCTGTTGATTTGGTTGTTAATGAC
 1961 CCTGAATTGAGAGATGAACAGTTCAATACTTCAGCAGCAACTGAAGGGATGAATAGCAGTCTGCAAAGTC
 2031 CTCCACCTGAGGACAAGAGTAGCATAAGTGAGACTGTGGAAACTGAAACAAGTTGTGCTAAAGAGATAGA
```

Figure S2: Genomic sequence containing the *GhLIPN-2* transcript and promoter. The genomic sequences preceding the putative *GHLIPN-2* transcription start site (TSS) is in lowercase and the 1 kb promoter sequence analysed for *cis*-elements is underlined. The 11 exons are shaded and numbered 1-11 and the 10 introns are numbered 1-10, with the numbers underlined. The C1B10 EST is indicated with a wavy underline (____). Translation start and end codons are in large bold font and underlined. A 13 bp insert located in intron 6 (6) and not present in the *GhLIPN-1* transcript, which was used for homoeolog-specific amplification of *GhLIPN-2* is in upper case red font.

Figure S3


GhLIPN-2 ccaaaaaataaaaaaataaaaattcaaagctaccagagaatcctaacagcgtcggatcgttgtat.....agtaaagagaatc 78
GhLIPN-1 ----------c----------catcattcatcgatcat---------- 95
GhLIPN-2 ttacaattgtttcgatttcgagctgccagattcgttcatcgcgggaattttcttctttgcgcaattcttaatctaattaa.ggaaactgcagaag 172
GhLIPN-1 ----------t----------g----------c----------t---g----------a---------- 190
M N V V G K V G S L I S Q G V Y S V A 19
GhLIPN-2 gatttttgttaatctaagaaaggtttaattttgacaaATGAATGTGGTTGGCAAAGTTGGGAGTTTAATTTCACAAGGTGTATATTCTGTTGCTA 267
GhLIPN-1 ----------g---------- 285
T P F H P F G G A V D I I V V R Q P D G T F R S T P W Y V R F G 51
GhLIPN-2 CTCCTTTCCATCCTTTCGGTGGAGCGGTTGATATAATTGTTGTTCGGCAACCAGATGGGACTTTTCGGAGCACACCTTGGTATGTTCGGTTTGGG 362
GhLIPN-1 ----------G----------T-G----------T------ 380
A S
K F Q G V L K G A E K V V R I T V N G I E V D F H M Y L D N S G 83
GhLIPN-2 AAGTTTCAGGGTGTCTTGAAAGGGGCTGAGAAGGTCGTTCGTATAACTGTTAATGGCATTGAAGTAGATTTTCATATGTATCTTGATAACTCTGG 457
GhLIPN-1 ----------A----------C---------- 475
A
Q A Y F L K E V E S G K G F E T N G D L K D S D G E V V S D S 114
GhLIPN-2 GCAAGCATATTTTTTAAAGGAGGTTGAATCTGGTAAAGGATTTGAGACAAATGGAGATTTGAAGGATTCTGATGGTGAAGTTGTTTCAGATTCCT 552
GhLIPN-1 ----------C---------- 570
Q
W V A Q L R D E C D A T N Q K Q L E R A E S D T R F Y D F Q D D 146
GhLIPN-2 GGGTGGCTCAATTGAGGGATGAATGTGATGCCACAAATCAAAAACAGCTTGAAAGGGCAGAATCTGATACTAGGTTCTATGATTTTCAAGATGAT 647
GhLIPN-1 C----------C---------- 665
S H
Q F S Q E G L V N F S E Y G S N R Y E G L D S E C F G E A K G L 178
GhLIPN-2 CAGTTTTCTCAAGAGGGTCTGGTTAATTTTTCAGAATATGGGTCCAACCGATATGAGGGTTTAGATAGTGAGTGTTTTGGGGAAGCAAAAGGTTT 742
GhLIPN-1 ------A----------G----------G----------C---------- 760
Y D G Q
D S V V F F S E D G H I L T A P V L A S D R S A E N V Q L S T 209
GhLIPN-2 AGATTCAGTTGTCTTCTTCAGTGAGGATGGCCATATTCTTACTGCCCCTGTTTTGGCATCAGACAGGAGTGCTGAAAATGTGCAACTAAGCACAC 837
GhLIPN-1 ----------A----------C--A---------- 855
K
P L F H I G A G E G P D S C D G N G E F S P G G N E S D A D Y I 241
GhLIPN-2 CTCTGTTCCATATAGGAGCAGGTGAAGGGCCTGACTCTTGTGACGGTAATGGGGAATTTAGTCCAGGTGGCAATGAATCAGATGCTGACTATATA 932
GhLIPN-1 ----------C----------AC------T---------- 950
F
G K L N A A A P K N S S D I V C S L D N D S T A L R H H L E V C 273
GhLIPN-2 GGTAAGCTCAATGCTGCGGCACCTAAGAATTCCTCTGATATTGTTTGCAGCTTGGACAATGATTCTACTGCTTTGAGACACCATCTAGAAGTTTG 1027
GhLIPN-1 ---------- 1045
E K G G E H A C Q T E E T R N L F K H E N E F I R Q S Y S E D 304
GhLIPN-2 TGAAAAAGGGGGGGAACATGCTTGTCAAACTGAAGAAACTCGAAACCTTTTCAAGCATGAAAATGAATTTATCAGGCAAAGTTACAGTGAAGATG 1122
GhLIPN-1 -----GA----------C-----T----------G---T-------- 1140
R H D I
A S V H I K D D V F K S C L E L S E L G R H D E N T N S E E I D 336
GhLIPN-2 CATCCGTGCATATTAAGGATGATGTTTTTAAGAGCTGTCTTGAACTATCTGAATTGGGCAGACACGATGAGAACACCAATTCTGAAGAAATAGAT 1217
GhLIPN-1 -----A----------A----------A----------G------- 1235
M S E G
S P L Q A Q I S Q D K P S C S P P E V G E T E N G A I G G S R N 368
GhLIPN-2 AGCCCACTGCAAGCTCAGATTTCACAAGATAAGCCTTCTTGCAGCCCCCCAGAAGTTGGTGAAACTGAAAATGGAGCTATTGGTGGCTCCAGAAA 1312
GhLIPN-1 ----------G----------G---------- 1330
R D
K D V L S S S C S P N Y S N E N G S P D L P V E K T M L A T D 399
GhLIPN-2 TAAAGATGTGTTGTCTTCCTCTTGTAGTCCCAATTATTCCAATGAAAATGGGTCTCCTGATTTACCAGTTGAAAAGACAATGTTGGCTACAGATA 1407
GhLIPN-1 ----------T----------T---------- 1425
I I
N M G S N N A S V D L V V N D P E L R D E Q F N T S A A T E G M 432
GhLIPN-2 ATATGGGTTCTAACAATGCATCTGTTGATTTGGTTGTTAATGACCCTGAATTGAGAGATGAACAGTTCAATACTTCAGCAGCAACTGAAGGGATG 1502
GhLIPN-1 ---A----------G----A----------G------ 1520
K D
N S S L Q S P P P E D K S S I S E T V E T E T S C A K E I D V R 463
GhLIPN-2 AATAGCAGTCTGCAAAGTCCTCCACCTGAGGACAAGAGTAGCATAAGTGAGACTGTGGAAACTGAAACAAGTTGTGCTAAAGAGATAGATGTTCG 1597
GhLIPN-1 ----------A- 1615
S
A S L G F E I S L C G N E L Y V G M G S D A A A E V F E S H R 494
GhLIPN-2 TGCTAGCCTGGGATTTGAGATCTCACTCTGTGGCAATGAACTTTATGTGGGTATGGGTTCAGATGCTGCAGCAGAAGTCTTTGAATCACATCGGA 1692
GhLIPN-1 ----------G----------C----------G---------- 1710
A

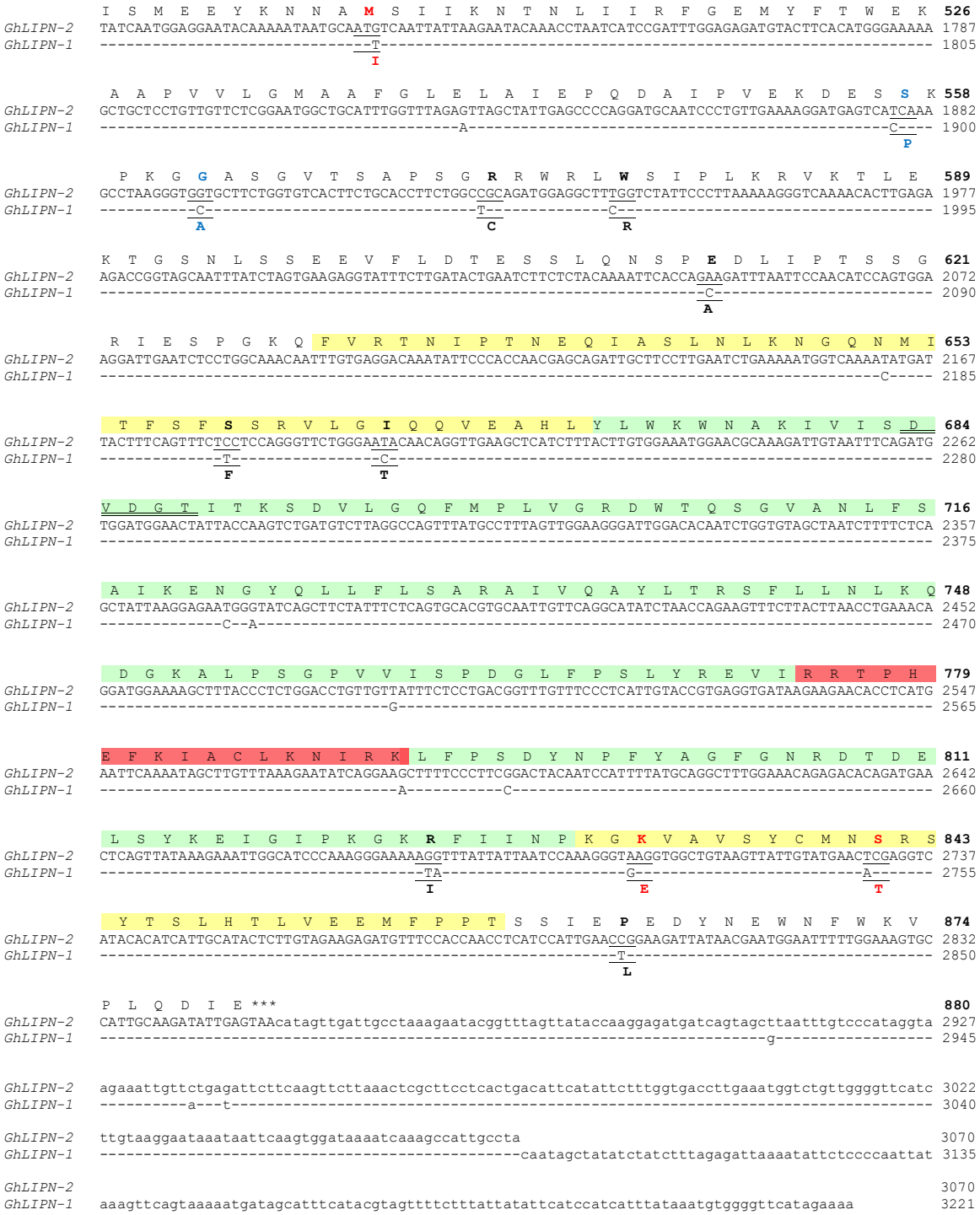


Figure S3: Comparison of the full-length cDNAs and deduced amino acid sequences of *GhLIPN-2* and *GhLIPN-1* encoding lipin proteins in *Gossypium hirsutum*. Conserved, semi-conserved and radical (non-conserved) non-synonymous amino acid substitutions are indicated in blue, red and black font, respectively. The N-terminal and C-terminal lipin domains are shaded in pink and yellow, respectively. The HAD domains are shaded in green and the bipartite nuclear localization signals (BNLS) are shaded in red. The DxDxT/V haloacid dehalogenase catalytic motif involved with phosphatase activity is double-underlined.

Supplementary Table S2:

Predicted putative post-translational modification sites in the GhLIPN homoeologs.

Modification/site	Position on aa sequence	Sequence (in order of appearance)
Protein kinase C (PKC) phosphorylation sites	40-42, 93-95, 161-163, 198-200, **492-494**[I], 556-558, **571-573**[I], 620-622, **658-660**[I], 729-731, 813-815	TFR, SGK, SNR[I]/SDR[2], SDR, **SHR**[I], SSK[I]/SPK[2], **SGR**[I], SGR, **SSR**[I], SAR, SYK
Casein kinase II (CKII) phosphorylation sites	99-102, 106-109, 236-239, 300-303, 316-319, 330-333, 351-354, 445-448, 496-499, 596-599, 611-614, 683-686, 690-693, 777-780, 813-816, 849-852, 859-862	TNGD, SDGE, SDAD, SYSE[I]/SDIE[2], SCLE[I]/SSLE[2], TNSE, SPPE, SISE, SMEE, SSEE, SPED[I]/SPAD[2], SDVD, TKSD, TPHE, SYKE, TLVE, SSIE
Tyrosine kinase phosphorylation sites	133-141, 807-814	RAEsDTRfY, RDTdELSY
N-Glycosylation sites	155-158, 234-237, 251-254, 261-264, 379-382, 405-408, **423-426**[I], 432-435, 594-597	NFSE, NESD, NSSD, NDST, NYSN, NASV, **NTSA**[I], NSSL, NLSS
N-Myristoylation sites	8-13, 39-44, 55-60, **165-170**[2], 232-237, 276-281, 361-366, 402-407, 430-435, 482-487, 538-543, 561-566, 592-597, 764-769, 805-810, 818-823	GSLISQ, GTFRST[I]/GTFRSS[2], GVLKGA, **GGLDSE**[2], GGNESD, GGEHAC, GAIGGS, GSNNAS, GMNSSL, GSDAAA, GLELAI, GGASGV[I]/GAASGV[2], GSNLSS, GLFPSL, GNRDTD, GIPKGK

Homeoelog specific sites: [I] GhLIPN-1 only, [2] GhLIPN-2 only.

Additional file 5.

Supplementary Figure S4: ClustalW alignment of the GhLIPN gene copies with proteins from *Vitis vinifera* and *Arabidopsis thaliana*, the top scoring lipin domain-containing proteins from a NCBI PHI-BLAST. Identical amino acids (*), conserved substitutions (:), semi-conserved substitutions (.) and gaps (-) are indicated in the alignment. Amino acids that are conserved among representatives across all eukaryotic kingdoms, based on a PANTHER (protein analysis through evolutionary relationships, www.pantherdb.org) multiple alignment of the lipin family (PTHR12181), are indicated in large bold font. The N-terminal and C-terminal lipin domains are shaded in pink and yellow, respectively. The HAD domains (672-829 aa) are shaded in green and the BNLSs are shaded in red. The DxDxT/V HAD catalytic motif involved with phosphatase activity is underlined. Amino acid differences between GhLIPN-1 and -2 homoeologs are in bold red font. Highly conserved residues, which affect the phosphatidate phosphatase (PAP) activity of lipins, are shaded in blue and a conserved phosphoserine is shaded in orange.

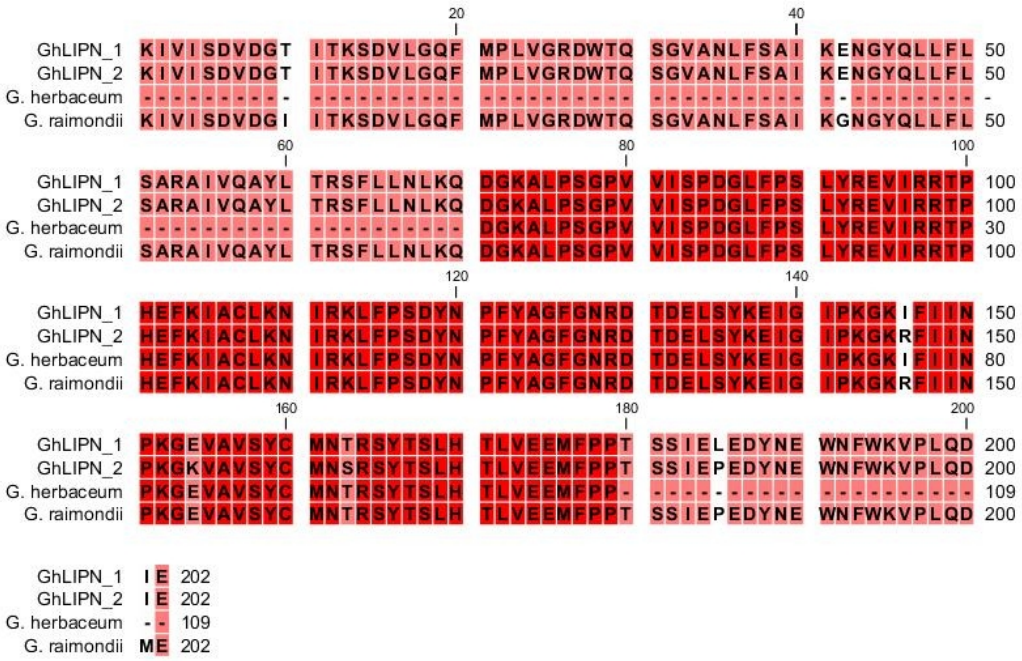
```
GhLIPN-2        MNVVGKVGSLISQGVYSVATPFHPFGGAVDIIVVRQPDGTFRSTPWYVRFGKFQGVLKGAEKVVRITVNGIEVDFHMYLDNSGQAYFLKEVESGKGFETNGDLKDSDGEVVSDSW-----   115
GhLIPN-1        MNVVGKVGSLISQGVYSVATPFHPFGGAVDIIVVRQADGTFRSSPWYVRFGKFQGVLKGAEKVVRITVNGIEADFHMYLDNSGQAYFLKEVESGKGFQTNGDLKDSDGEVVSDSS-----   115
Vitis           MNVVGIVGSLISQGVYSVATPFHPFGGAVDVIVVQQQDGTFRTTPWYVRFGKFQGVLKGAEKMVRISVNGVEAKFHMYLDNSGEAYFIREVSS-EGKGTNGIIKESDGLEVIDDSSKDNG   119
Arabidopsis     MSLVGRVGSLISQGVYSVATPFHPFGGAIDVIVVQQQDGSFRSTPWYVRFGKFQGVLKGAEKFVRISVNGTEADFHMYLDNSGEAYFIREVDPAANDTNNLISGSENNNGNQNNG-----   115
                *.:** ********************:*:***:* **:**::******************.***:*** *..*********:***::**..  .  .*    ..:.    :.

GhLIPN-2        ------------------VAQLRDECDATN-QKQLERAESDTRFYDFQDDQFSQEGLVNFSEYGSNRYEGLDSECFG--EAKGLDSVVFFSEDGHILTAPVLASDRSAENVQLSTPLFHI   214
GhLIPN-1        ------------------VAQLRDECDATN-HKQLERAESDTRFYDFQDDQFYQEGLVNFSEYGSDRYGGLDSECFG--EAQGLDSVVFFSKDGHILTAPVLASDRSAENVQLSTPLFHI   214
Vitis           DNVTVNTCKLESSVSDPGVVQIRDECASSGGWLERVESDNDRRFYEFQDDQSSHEGSVELSEYGSNQYESFDHVGHFGESRALDSEVVLVSVDGHILTAPISSSEGNTENLQLITPQFHL   223
Arabidopsis     -----VTYRLEHSLSDSGTGELREGFDPLS-RLERTESDCNRRFYDFQDDPPS-----PTSEYGSARFDNLNVESYG-DSQGSDSEVVLVSIDGHILTAPVSVAEQEAENLRLNTPQFHL   223
                                      . ::*:   . .   :  .:: : ***:****     ***** :: .::   .   .    ..**:.* ********:  :: .:**::* ** **:

GhLIPN-2        GAGEGPDSCDGNGEFSPGGNESDADYIGKLN--AAAPKNSSDIVCSLDNDSTALRHHLEVCEKGGEHACQTEETRNLFKHENEFIRQSYSEDASVHIK-DDVFKSCLELSELGRHDENTN   331
GhLIPN-1        GAGEGPDFCDGNGEFSPGGNESDADYIGKLN--AAAPKNSSDIVCSLDNDSTALRHHLEVCERGGEHACHTEETRNLFKHENEFIRQSDIEDASMHIK-DDVFKSSLELSELGRHEENTN   331
Vitis           GPGEGTDFCEGNEEFSAGEGPWAAGYLNELD--SASANVDSQNVCSVNNDNSAFGHQLEVCEGEKEKASLADRTQDVATQGRGPSMQSNLEDKNISIERKDVFRSCLELTELATQVVNGD   357
Arabidopsis     APGDGTEFCEGNTEFASSETPWDTEYIDKVEESSDTANIASDKVDAINDERNDLDSHSRDNAEKDSHDAERDLLGSCLEQSELTKTSENVKSEEPGPTFEDRNLKEGEFPLRTIMENDRS   343
                ..*:*.: *:** **:..     : *:.:::  : :.:  *: * ::::: . :  : .      .: .  :   .   : .    ..  :. .     .*   .  *:.       : .

GhLIPN-2        SEEIDSPLQAQISQDKPSCSPPE----VGETENGAIGGSRNKDVLS---SSCSPN----YSNENGSPDLPVEKTMLATDNMGSNNASVDLVVNDPELRDEQFNTSAATEGMNSSLQSPPP   440
GhLIPN-1        SEGIDSPLQAQISQDKRSCSPPE----VGETEDGAIGGSRNKDVLS---SSCIPN----YSNEIGSPDLPVEKTMLATDNKGSNNASVDLVVNDPELRDEQFDTSAATEGMNSSLQSPPP   440
Vitis           IRHLNSSLKVQEGMENSQEKSPQGLRAVDDTEHGHVVQFSNDDELS---SCNPESPWNTTSPDLCVEVEPNEKNELSMEHIELDNMSVPSVRNDPEWKDEQFGMLAVEGTNGSPQRPAPE   474
Arabidopsis     EDEVTIESIDTLVDSFESSTTQITIEEVKTTEGSRISVDSNADSECKDEQTSAETAILFNNQESSISVDSNADSECKDEQPRISAETAILINNQEGGIIESEDQDSERVSIDSTREEVDK   463
```

```
                    :          .  . ..     *  ** . :    * *  .   .    .. :     .  ..    ::   .  :.  : *:     *. .  :     .*. .

GhLIPN-2        EDKSSISETVETETSCAKEIDVRASLGFEISLCGNELYVGMGSDAAAEVFESHRISMEEYKNNAMSIIKNTNLIIRFGEMYFTWEKAAPVVLGMAAFGLELAIEPQDAIPVEKDESSKPK   560
GhLIPN-1        EDKSSISETVETETSCAKEIDVSASLGFEISLCGNELYVGMGSDAAAEVFEAHRISMEEYKNNAISIIKNTNLIIRFGEMYFTWEKAAPVVLGMAAFGLELAIEPQDAIPVEKDESPKPK   560
Vitis           DACSKSETVETQATISCEGIQTDSSIRFEISLCGKELRAGMGLVAAAEAFEAQRISEEEFKTSAPSIIKNENLIIRFREKYLTWDKAAHIVLGMAAFGLDLPVEPKDAIPVEQDETPKAR   594
Arabidopsis     DNEDRKTVVSVGVTSSVDEGEPDTDQRYELSLCKDELRQGMGLSAAAEVFDAHMISKEEYINSATSILESENLVVRIRETYMPWTKAARIVLGKAVFDLDLDIQPDDVISVEENESPKPK   583
                :  .    .    * . .  :  :.  :*:*** .**  ***  ****.*::: ** **: ..* **::. **::*: * *:.* *** :*** *.*.*:* ::*.*.*.**::*:.*.:

GhLIPN-2        GGASGVTSAPSGRRWRLWSIPLKRVKTLEKTGSNLSSEEVFLDTESSLQNSPEDLIPTSSGRIESPGKQFVRTNIPTNEQIASLNLKNGQNMITFSFSSRVLGIQQVEAHLYLWKWNAKI   680
GhLIPN-1        GAASGVTSAPSGCRWRLRSIPLKRVKTLEKTGSNLSSEEVFLDTESSLQNSPADLIPTSSGRIESPGKQFVRTNIPTNEQIASLNLKNGQNMITFSFFSRVLGTQQVEAHLYLWKWNAKI   680
Vitis           GGDSKIAATSSGRRWRLWPIPFRRVKTLQHTDSNSSSEDVFVDSESGSQSTHVEPIPPSPGGSETPKKQLGRTNIPTTEQIASLNLKEGQNMVTFSFSTRVLGTQQVDAHIYLWKWNARI   714
Arabidopsis     DDETTITPSSSGTRWRLWPIPFRRVKTVEHTGSNSSSEEDLFVDSEPGLQNSPETQSTTESRHESPRRQLVRTNVPTNEQIASLNLKDGQNMITFSFSTRVLGTQQVDAHIYRWRWDTKI   703
                .  : ::.:.** **** .**::****:::*.** ***: :.  ..   ..  :  ..: .  *:* :*: ***:**.*********:****:**** :**** ***:**:* *:*:::*

GhLIPN-2        VISDVDGTITKSDVLGQFMPLVGRDWTQSGVANLFSAIKENGYQLLFLSARAIVQAYLTRSFLLNLKQDGKALPSGPVVISPDGLFPSLYREVIRRTPHEFKIACLKNIRKLFPSDYNPF   800
GhLIPN-1        VISDVDGTITKSDVLGQFMPLVGRDWTQSGVANLFSAIKENGYQLLFLSARAIVQAYLTRSFLLNLKQDGKALPSGPVVISPDGLFPSLYREVIRRTPHEFKIACLKNIRKLFPSDYNPF   800
Vitis           VISDVDGTITKSDVLGQFMPLVGKDWTQSGVARLFSAIKENGYQLLFLSARAIVQAYLTRSFLLNLKQDGKALPNGPIVISPDGLFPSLYREVIRRAPHEFKIACLEDIRALFPSDYNPF   834
Arabidopsis     VISDVDGTITKSDVLGQFMPFIGKDWTQSGVAKLFSAIKENGYQLLFLSARAIVQAYLTRNFLNNLKQDGKALPTGPVVISPDGLFPALYREVIRRAPHEFKIACLEDIRKLFPTDYNPF   823
                ********************::*:********.**************************.** **********.**:*********:********:*********::** ***:*****

GhLIPN-2        YAGFGNRDTDELSYKEIGIPKGKRFIINPKGKVAVSYCMNSRS-YTSLHTLVEEMFPPTSSIEPEDYNEWNFWKVPLQDIE-                                          880
GhLIPN-1        YAGFGNRDTDELSYKEIGIPKGKIFIINPKGEVAVSYCMNTRS-YTSLHTLVEEMFPPTSSIELEDYNEWNFWKVPLQDIE-                                          880
Vitis           YAGFGNRDTDELSYRKIGIPKGKIFIINPKGEVAISHRIDVKS-YTSLHTLVNDMFPPTSLVEQEDFNSWNFWKMPLPDIEL                                           915
Arabidopsis     YAGFGNRDTDELSYRKLGIPKGKIFIINPKGEVATGHRIDVKKSYTSLHTLVNDMFPPTSLVEQEDYNPWNFWKLPIEEVE-                                          904
                **************:::****** *******:** .: :: :. ********::****** :* **:* *****:*: ::*
```

Additional file 6 Supplementary Figure S5: Alignment of the *GhLIPN* gene copies with a homologous partial mRNA sequence (Genbank: CO087195) from *Gossypium raimondii* and an amplified region from *Gossypium herbaceum*, to determine their putative genomic origin. **A:** nucleotide alignment. **B:** amino acid alignment.

A

B



Supplementary Figure S5: Alignment of the *GhLIPN* gene copies with a homologous partial mRNA sequence (Genbank: CO087195) from *Gossypium raimondii* and an amplified region from *Gossypium herbaceum*, to determine their putative genomic origin. A: nucleotide alignment. B: amino acid alignment.

Additional file 7.

Fig S6: Alignment of the experimentally obtained promoter regions of *GhLIPN-1* and *-2*. One 1 kb of the *GhLIPN-2* promoter, preceding the putative transcription start site (TSS), and 1441 bp of the *GhLIPN-1* promoter are shown. Identical nucleotides are highlighted and gaps are indicated with dashes. The different classes of putative *cis*-elements are indicated in bold coloured font and underlined as follows: CAAT-box *cis*-elements (orange), gibberellin-responsive *cis*-elements (green), DNA-binding with one finger (DOF)-related *cis*-elements (pink), W-box *cis*-elements (purple), plant defence-related and elicitor-responsive *cis*-elements (red) and abiotic stress-responsive *cis*-elements (blue).

```
GhLIPN-2_prom     TTTCTTAACGCTGATTATGAGATTGTGAGAGAACGAGTATATATAGAGTGAAAAAAGCA  -943
GhLIPN-1_prom     TTGGTTCACGTTGATTATGTGATTGTGAGAAAAAGAGAATATATAGAGTGAAAAAAGG  -1383

GhLIPN-2_prom     GAGATGATGATGTAACGTAATTAGGCAACGAGTGGATAGGCTAACGAGGTTTATTTTA  -885
GhLIPN-1_prom     GAGATGATGATGTAACGTAATTAGGCAAAGAGTGGATAGGCTAATGAGGTTTATTTTA  -1325

GhLIPN-2_prom     TTTTGACCAGAAGGGGCCGCCTTTAAAAGTAGATTGGAGTATGATATAAGTAGGAGTT  -827
GhLIPN-1_prom     TTTTGACCAGAATGGGCCGCCTTTAAAAGTAGAATGGAGTATGATATAAGTAGGAGTT  -1267

GhLIPN-2_prom     GCGATGTGAAGCAGGGGATGAATGTGTAATAAAAAAGAAAGTAAAATTCCTGCCTTCT  -769
GhLIPN-1_prom     GCGATATGAAGCAGGGGATGAATGTGTAATAAGAAAGAAAATAAAATTCATGGCTTCT  -1209

GhLIPN-2_prom     CCTCATCTCA-----ACCAGTCGGGTAAAACATAAAATC----TAAATTAT--TTAG-  -723
GhLIPN-1_prom     CCTCATCTCAACTCAACCCGTAAGGTAAAACATAAAAACAAACTAAAATATAGTTAAG  -1151

GhLIPN-2_prom     -TATTTATTTAAATGAGATTTGGTTGGTGATCATGGTA-TTATAGTTTTGGTAGAAGA  -667
GhLIPN-1_prom     TTAATAATCTAAATGA-ATTAG--TGGTGATCATGGTTGTTGTAGTTTTGGTAGAAGA  -1096

GhLIPN-2_prom     AGATGGGTC---TT--ACAAAAAACAAACATCCCAAACTATAATTTAGATTTTAATTT  -614
GhLIPN-1_prom     AGATGGGTGAGTTTCAACAAAAACAAAACATCCCAAACTATAATTTAGAATTTA----  -1042

GhLIPN-2_prom     CTATTAATTACGTGTTAAATGCCATCTTAGCCACGACACAGAGCATGATATCTTAAAC  -556
GhLIPN-1_prom     ---------------TAAATG---TTTAATTCAATGCTCACCG-------TCTTAAAC  -1009

GhLIPN-2_prom     ACGGGAATTAAATC----------TTTACTTCTAA-----ATTCACTTATTTGGTCGC  -513
GhLIPN-1_prom     ACCGGAATTAAATCTAAATTCACTTTTGTTTCTTATTATTATTACAAAATTTTGTCGC  -951

GhLIPN-2_prom     ATTTAATTTTAATCATTTCTTTTCTTTTCCAAAACAAGCTCAATATTTGTTATGTTTG  -455
GhLIPN-1_prom     ATTTAATTTTAATCATTTCTTTTCTTTTCCAAAACAAGCTCAATATTTGGTATGTTTG  -893

GhLIPN-2_prom     AATTAAGCT---------------------------------------------------  -446
GhLIPN-1_prom     AATTAAGCTTTTAGTTGTTTCAAACACTTACAAAAATAACAGAATTTTATGTTGGTTC  -835

GhLIPN-2_prom     ---------------------------------------------------------  -445
GhLIPN-1_prom     ACTGGAAAGTGTATCACTAACACTCATTTTTTATTTTTGTTATTTTGTCACATGTAAG  -777

GhLIPN-2_prom     ---------------------------------------------------------  -445
GhLIPN-1_prom     ATTTGAAATATACAAGTTTCAAGGTTTTTATTATTCAATCAGAATAAATATGAATAAT  -719

GhLIPN-2 prom     ---------------------------------------------------------  -445
GhLIPN-1_prom     TTTCACTAATATGATAGAAAATACTCCTGAATAATGTCTTTTATTTCGATCTCATTAT  -661

GhLIPN-2_prom     ---------------------------------------------------------  -445
GhLIPN-1_prom     AGAATGATTAGATTTTTGAATTTCAATTTAGATATGTAGGAAAATATAAAATGTAAG  -603

GhLIPN-2_prom     ---------------------------------------------------------  -445
GhLIPN-1_prom     GGAGGAATTTTTGAGGAAAACTAACAAATTAGATGTTAGACTTTGATGTTTGAGATAG  -545

GhLIPN-2 prom     ------------------------------------------------------------  -445
GhLIPN-1_prom     ATTATAATCTACATTTAAAAAAAAAAAAGAAAAAGAAAAAGAATTTCGACTTCTTTTT  -487

GhLIPN-2_prom     ------------------------------------------CAGGAGGTTTCTTTTC-A  -429
GhLIPN-1_prom     AAACAAAAGTTGATTAGTGTTAGTGATACACTGTCATTCTCAGGAGGTTACTTTTCTA  -429

GhLIPN-2_prom     CGTGTTTGAAACAACTAACAGCTTAATTCAAACCTAACAAATATTGAGCTTGTTTTGG  -371
GhLIPN-1_prom     AGTGTTTGAAACAACTGAAAGCTTAATTCAAACATAACAAATATTGAGCTTGTTTTGG  -371

GhLIPN-2_prom     AA---------------TTATATGCGACCAAATTTTCTAATAATAGACAAGAAACAAA  -328
GhLIPN-1_prom     AAAAAGAAATGATTTAAATTATATGCGACAAAATTTTCTAATAATATATAAGAAACAAA  -313

GhLIPN-2_prom     TGTGAAATGAAGGGTAATTTAGTCAATTTAGATGTAAAGATT-------TTTTTTTAG  -277
GhLIPN-1_prom     AGTGAAATGAAGGGTAATTTAGTCAATTTAGAAGTTAATAATTTTTTTCTTTTTTTCG  -255
```

```
GhLIPN-2_prom      CATTACTGTAGAAGTTTCAATAATTATAATGTTTGTTCATTCCGGCGGCTCTACCATC  -219
GhLIPN-1_prom      GAATAAAACTCAAGTTTCAATAATTATAATGTTTCTTCATCCCGGCGGCTCTACCATC  -197

GhLIPN-2_prom      CTTAAG-CCTGAAAAAGCGGGACAATATTTCAAATATCCTAAATTACCCTTCCATTAA  -162
GhLIPN-1_prom      CTTAAGACCGGAAAAAGCGGGACAATATTTCAAATATCCAAAATTACCCTTCCATTAA  -139

GhLIPN-2_prom      TCACAAATTTACACCTCATATTTCTCGCTTCGTAGAGTCTTTCTAAGCAAAACTAAGG  -104
GhLIPN-1_prom      TCACAAATTTACACGTCATATTTCTCG----------TATTTTTAAGCAAAACTAAGG  -91

GhLIPN-2_prom      GACCTATTTATGTCTTTTCACAAGGTTACTTGTTTATTGCATGGCGGTTGCGGTTAG-  -47
GhLIPN-1_prom      GACCTATTTATGTCTTTTCGCA--------------TTGCATGGCGGTTGCGGTTAGG  -47

GhLIPN-2_prom      TCGCCGTTGCCAACTTCTGACTCGTTTTCAAAATTTCCATTTTTTT              -1
GhLIPN-1_prom      TCGCCGTTGCCAACTTCTGACTCGTTTTCTAAATTTCCATTTTTCC              -1
```

Table S3 A-C: *In silico* analysis of selected *cis*-elements in the promoters of the *GhLIPN* homoeologs. The promoter regions were obtained by genome walking upstream from the *GhLIPN-1* and *-2* transcription start sites. Homoeolog-specific promoter amplification was used to distinguish between the two promoters of *GhLIPN-1* and *-2*. The analysed promoter sequences were limited to 1 kb from the predicted TSSs. Identical *cis*-elements occurring in the same positions in the promoters of the two homoeologs are indicated by the position number in bold font.

Table S3.A: Analysis of hormone-responsive *cis*-elements in the promoters of the *GhLIPN* homoeologs.

Both homoeologs						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
AGCBOXNPGLB / GCC box	-	-233[2], -211[1]	AGCCGCC	Binding site for ethylene-response factors (ERFs), which act as stress signal response factors, conserved in most PR protein genes	*Arabidopsis thaliana, Nicotiana plumbaginifolia, Nicotiana sylvestris, Oryza sativa*	PLACE, PlantCARE
ARR1AT	+ -	-989[2], -981[2], -854[2], -778[1], -708[2], -656[1], -651[1], -625[2], -546[1], -476[1], -362[1], -290[2] -999[1], -940[1], -740[1], -738[2], -545[2], -539[1], -502[2], -163[2], -140[1]	NGATT	Binds cytokinin-regulated ARR1	*Arabidopsis thaliana*	PLACE
ERE	+ -	-193[2], -170[1] -775[1]	ATTTCAAA	Ethylene-responsive element	*Dianthus caryophyllus*	PlantCARE
GARE2OSREP1	+	-930[2], -856[1]	TAACAGA	Gibberellin-responsive element (GARE) for gibberellin-up-regulated proteinase expression	*Oryza sativa*	PLACE

GAREAT	+ -	-581[2], **-394[B]** -799[1], -467[2]	TAACAAR	GARE for gibberellin-response	*Arabidopsis thaliana*	PLACE
GCCCORE	+ -	-869[2] -233[2], -211[1]	GCCGCC	Core of GCC-box found in many pathogen-responsive genes eg: plant defensin 2.1 (PDF1.2), Thionin 2.1 (Thi2.1) and PR-4. Also an ethylene- and jasmonate-responsive element	*Arabidopsis thaliana, Solanum lycopersicum*	PLACE
MYBGAHV	+ -	-581[1], **-394[B]** -799[1], -467[2]	TAACAAA	Central element of gibberellin (GA) response complex (GARC), specific binding site for Gamyb	*Hordeum vulgare, Oryza sativa*	PLACE

GhLIPN N

Name of *cis*-element	Strand	Position/s	Sequence	Predicted function	Organism/s	Database
ASF1MOTIFCAMV/ TGACG-motif	-	-125	TGACG	SA responsive, binds TGA1 in a non-expresser of pathogenesis-related 1 (NPR1)-dependent manner for transcriptional activation of defence-related genes during SAR	*Arabidopsis thaliana*, CaMV, *Hordeum vulgare, Nicotiana tabacum*	PLACE, PlantCARE
CGTCA-motif	+	-125	CGTCA	Involved in MeJA-responsiveness	*Hordeum vulgare*	PlantCARE

GhLIPN I

Name of *cis*-element	Strand	Position/s	Sequence	Predicted function	Organism/s	Database
ABRE	+	-605, -430	YACGTG	Involved in ABA responsiveness	*Arabidopsis thaliana*	PlantCARE
CAREOSREP1	-	-831	CAACTC	Binds gibberellin up-regulated elements	*Oryza sativa*	PLACE
DPBFCOREDCDC3	+ -	-557 -430	ACACNNG	Binds a novel class of bZIP transcription factors, DPBF-1 and 2, embryo-specific and induced by ABA	*Arabidopsis thaliana, Daucus carota*	PLACE
TATCCACHVAL21	-	-911	TATCCAC	Part of GARC for gibberellin response	*Hordeum vulgare*	PLACE
TCA-element	+ -	-10 -670	$CCATCT_5$	Involved in salicylic acid responsiveness	*Nicotiana tabacum*	PlantCARE

[B] Both homoeologs, [1]*GhLIPN-1*, [2]*GhLIPN-2*.

Table S3.B Analysis of elicitor/defence-responsive *cis*-elements in the promoters of the *GhLIPN* homoeologs.

Both homoeologs						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
AGCBOXNPGLB / GCC box	-	-233[2], -211[1]	AGCCGCC	Binding site for ERFs, which act as stress signal response factors, conserved in most PR protein genes	*Arabidopsis thaliana, Nicotiana plumbaginifolia, Nicotiana sylvestris, Oryza sativa*	PLACE, PlantCARE
BOXLCOREDCPAL	+ -	-224[2], -202[1] -702[2]	ACCWWCC	Binds DCMYB1 transcription factor in the core promoter region of PAL1 in response to elicitation	*Daucus carota*	PLACE
GCCCORE	+ -	-869[2] -233[2], -211[1]	GCCGCC	Core of GCC-box found in many pathogen-responsive genes eg: PDF1.2, Thi2.1 and PR4, also an ethylene- and jasmonate -responsive element	*Arabidopsis thaliana, Solanum lycopersicum*	PLACE
GT1GMSCAM4	+ -	-951[2], -516[1], -510[1], -186[1], -29[2] -269[1], -261[1], -7[1]	GAAAAA	"GT-1 motif" Plays a role in pathogen- and salt-induced SCaM-4 (CaM isoform) gene expression	*Glycine max*	PLACE
HSE	+ -	-507[1] -357[2]	AAAAAATTTC	Involved in heat stress-responsiveness	*Brassica oleracea*	PlantCARE
MYB1LEPR	-	-584[1], -416[2]	GTTAGTT	Regulates defence-related gene expression, binds tomato transcription factor Pti4	*Arabidopsis thaliana, Solanum lycopersicum*	PLACE

GhLIPN N						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
ASF1MOTIFCAMV/ TGACG-motif	-	-125	TGACG	Essential for SAR, binds TGA1 in a NPR1-dependent manner for transcriptional activation of defence-related genes	*Arabidopsis thaliana,* CaMV, *Hordeum vulgare, Nicotiana tabacum*	PLACE, PlantCARE
BIHD1OS	+	-789, -455	TGTCA	Binding site of OsBIHD1, a rice BELL homeodomain transcription factor - in disease resistance responses.	*Oryza sativa*	PLACE
SEBFCONSSTPR10A	+	-790	YTGTCWC	Binds the single-stranded DNA binding protein SEBF, causes repression of the defence gene PR-10a	*Solanum tuberosum*	PLACE
GhLIPN I						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
ABRERATCAL	+	-430	MACGYGB	"ABRE-related sequence", Calcium-responsive	*Arabidopsis thaliana*	PLACE
Box-W1	+	-882	TTGACC	Fungal elicitor-responsive element	*Petroselinum crispum*	PlantCARE
CACGTGMOTIF	+ and -	-430	CACGTG	G-box, binds G-box-binding factors (GBFs), involved in phytochrome A-regulated and elicitor-inducible gene expression	*Antirrhinum majus, Arabidopsis thaliana, Brassica napus, Catharanthus roseus, Phaseolus vulgaris, Solanum lycopersicum, Triticum aestivum, Zea mays*	PLACE
MYB26PS	+ -	-394 -467	TAACAAA	MYB26 binding site, present in promoter regions of several phenylpropanoid biosynthetic genes	*Pisum sativum*	PLACE
MYBPLANT	+ -	-398 -702	MACCWAMC	Binds AmMYB308 and AmMYB330, which regulate phenylpropanoid and lignin biosynthesis	*Antirrhinum majus, Arabidopsis thaliana, Petroselinum crispum, Petunia hybrida , Phaseolus vulgaris, Zea mays*	PLACE

[B] Both homoeologs, [1]*GhLIPN-1*, [2]*GhLIPN-2*

Table S3.C: Analysis of abiotic stress-responsive *cis*-elements in the promoters of the *GhLIPN* homoeologs.

Both homoeologs						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
ABRELATERD1	+ -	-604[2], -429[2] -430[2], -127[1]	ACGTG	ABRE-like sequence required for etiolation-induced expression of early responsive to dehydration 1 (erd1)	*Arabidopsis thaliana*	PLACE
ACGTATERD1	+ and -	-928[2], -604[2], -429[2], -126[1]	ACGT	Required for etiolation-induced expression of erd1	*Arabidopsis thaliana*	PLACE
ANAERO1CONSENSUS	+ -	-650[2], -486[1], -334[2], -319[1] -985[1]	AAACAAA	Consensus sequence found in *cis*-elements of genes involved with anaerobic fermentation	*Arabidopsis thaliana, Hordeum vulgare, Solanum lycopersicum, Oryza sativa, Petunia hybrida, Pisum sativum, Zea mays*	PLACE
EECCRCAH1	+	-831[2], -643[1], -504[1]	GANTTNC	Binding site of CO_2-responsive MYB transcription factor LCR1	*Chlamydomonas reinhardtii*	PLACE
HSE	+ -	-507[1] -357[2]	AAAAAATTTC	Involved in heat stress-responsiveness	*Brassica oleracea*	PlantCARE
MYBCORE	+ -	-60[1], -59[2], -54[1], -53[2], **-43[B]** -856[1], -417[1], -413[2]	CNGTTR	Binding site for all animal MYB and at least two plant MYB proteins, ATMYB1 and ATMYB2	*Arabidopsis thaliana , Petunia hybrida*	PLACE
MYB2CONSENSUSAT	+ -	-417[1] **-43[B]**	YAACKG	MYB recognition site in the promoters of dehydration-responsive gene *rd22* and other genes in Arabidopsis	*Arabidopsis thaliana*	PLACE
MYCCONSENSUSAT	+ and -	-786[1],-430[2], -417[1], -331[2]	CANNTG	Dehydration and cold-responsive, binds to ATMYC2	*Arabidopsis thaliana*	PLACE

GhLIPN N						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
LTRE1HVBLT49	-	-259	CCGAAA	Core of low temperature responsive element (LTRE)	*Hordeum vulgare*	PLACE
MBS	+	-417	CAACTG	MYB binding site involved in drought-inducibility	*Arabidopsis thaliana*	PlantCARE
MYCATERD1	-	-786	CATGTG	Water stress, MYC recognition sequence necessary for expression of erd1	*Arabidopsis thaliana*	PLACE
MYCATRD22	+	-786	CACATG	Binding site for MYC (rd22BP1) in promoter of dehydration-responsive gene, *rd22*	*Arabidopsis thaliana*	PLACE
PREATPRODH	+	-813	ACTCAT	Hypoosmolarity-responsive element	*Arabidopsis thaliana*	PLACE
LTR	-	-259	CCGAAA	Involved in low-temperature responsiveness	*Hordeum vulgare*	PlantCARE
GhLIPN I						
Name of *cis*-element	**Strand**	**Position/s**	**Sequence**	**Predicted function**	**Organism/s**	**Database**
LTRECOREATCOR15	-	-754	CCGAC	Core of LTRE involved in cold-, drought- and ABA-regulated gene expression	*Arabidopsis thaliana, Brassica napus*	PLACE

[B] Both homoeologs, [1]*GhLIPN-1*, [2]*GhLIPN-2.*