

Two-dimensional gel electrophoresis-based analysis provides global insights into the cotton ovule and fiber proteomes

Xiang Jin^{1,2}, Limin Wang^{1,3}, Liping He³, Weiqiang Feng^{1,3} & Xuchu Wang^{1,2*}

¹Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China;

²The State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China;

³College of Horticulture and Landscape, Hainan University, Haikou 570228, China

Received May 14, 2015; accepted July 20, 2015; published online January 22, 2016

Proteomic analysis of upland cotton was performed to profile the global detectable proteomes of ovules and fibers using two-dimensional electrophoresis (2DE). A total of 1,203 independent protein spots were collected from representative 2DE gels, which were digested with trypsin and identified by matrix-assisted laser desorption and ionization-time-of-flight/time-of-flight (MALDI-TOF/TOF) mass spectrometry. The mass spectrometry or tandem mass spectrometry (MS or MS/MS) data were then searched against a local database constructed from *Gossypium hirsutum* genome sequences, resulting in successful identification of 975 protein spots (411 for ovules and 564 for fibers). Functional annotation analysis of the 975 identified proteins revealed that ovule-specific proteins were mainly enriched in functions related to fatty acid elongation, sulfur amino acid metabolism and post-replication repair, while fiber-specific proteins were enriched in functions related to root hair elongation, galactose metabolism and D-xylose metabolic processes. Further annotation analysis of the most abundant protein spots showed that 28.96% of the total proteins in the ovule were mainly located in the Golgi apparatus, endoplasmic reticulum, mitochondrion and ribosome, whereas in fibers, 27.02% of the total proteins were located in the cytoskeleton, nuclear envelope and cell wall. Quantitative real-time polymerase chain reaction (qRT-PCR) analyses of the ovule-specific protein spots P61, P93 and P198 and fiber-specific protein spots 230, 477 and 511 were performed to validate the proteomics data. Protein-protein interaction network analyses revealed very different network cluster patterns between ovules and fibers. This work provides the largest protein identification dataset of 2DE-detectable proteins in cotton ovules and fibers and indicates potentially important roles of tissue-specific proteins, thus providing insights into the cotton ovule and fiber proteomes on a global scale.

cotton, 2DE, MALDI-TOF/TOF, GO enrichment, protein-protein interaction networks

Citation: Jin, X., Wang, L., He, L., Feng, W., and Wang, X. (2016). Two-dimensional gel electrophoresis-based analysis provides global insights into the cotton ovule and fiber proteomes. *Sci China Life Sci* 59, 154–163. doi: 10.1007/s11427-016-4999-4

INTRODUCTION

Cotton fiber is the most prevalent natural material used in the textile industry, and upland cotton (*Gossypium hirsutum*, Gh) is the most widely cultivated cotton species. As one of the longest single cells, cotton fibers also serve as an excellent single-cell model for studying cell elongation and

related biological processes (Ji et al., 2002; Ji et al., 2003; Qin et al., 2011). Over the past 10 years, numerous studies have reported important functional genes and biological processes for cotton fiber development, including ethylene signaling, very long-chain fatty acids, biosynthesis of pectic precursors, ascorbate peroxidase, sucrose synthase gene family, calcium-dependent protein kinase, DNA methylation and miRNAs (Huang et al., 2008; Jin et al., 2013b; Li et al., 2007; Li et al., 2012; Pang et al., 2010; Pei, 2015; Qin

*Corresponding author (email: xchwanghainan@163.com)

et al., 2007; Shi et al., 2006; Wang et al., 2011; Zhou et al., 2013). However, studies on the cotton proteome have not been as abundant as those of the transcriptome, partly due to a lack of information about cotton genome sequences (Du et al., 2013; Liu et al., 2012; Yang et al., 2008; Zheng et al., 2012). Expressed sequence tags (EST) and next-generation sequencing data of the transcriptome collected in the last few years have been used to construct a database to identify the mass spectrometric (MS) data of proteins (Jin et al., 2013a; Wang et al., 2015).

Substantial progress has been made on cotton genome sequencing in recent years. The draft genomes of two diploid cotton species (the ancestors of tetraploid cotton Gh), *Gossypium ramosum* (Gr) and *Gossypium arboreum* (Ga), were reported in 2012 and 2014, respectively (Wang et al., 2012a; Li et al., 2014). Furthermore, the complex allotetraploid genome sequences of Gh TM-1 were reported in 2015 (Cao, 2015; Li et al., 2015; Zhang et al., 2015). Based on the cotton genome sequences, significant advances have been made in the last few years in the development of gel-free and high-throughput MS protein identification technologies (He, 2013; Hu et al., 2013; Meng et al., 2011). Comparative proteomics research of fiber initiation based on isobaric tags for relative and absolute quantification (iTRAQ) has been reported, and 5,884 proteins from 0 day post-anthesis (DPA) wild-type and *fuzzless-lintless* (*fl*) ovules, of which 2,927 proteins preferentially accumulated in wild-type ovules, have been identified (Wang et al., 2015). Furthermore, the lipid metabolic process has been characterized as an enriched pathway, and the *GhPAS2* gene, an inhibitor of very long-chain fatty acid biosynthesis, was shown to significantly suppress fiber initiation. Additionally, quantitative proteomic studies at the post-transcriptional modification level, such as protein phosphorylation, have been performed using iTRAQ coupled with phosphorylated peptide enrichment. These studies identified 619 and 1,020 phosphoproteins in cotton ovules and leaves, respectively, revealing some important roles of protein phosphorylation during cotton fiber initiation (Ma et al., 2014) and nitric oxide-responsiveness (Fan et al., 2014).

However, large-scale data of the two-dimensional gel electrophoresis (2DE) detectable proteomes of cotton tissues have not been reported. Compared with gel-free MS technologies, a 2DE gel-based proteomic approach still provides certain advantages in research, such as protein spot visualization on electrophoresis gel maps, reliable evidence for existing protein isoforms and abundant information from individual protein spots. Here, we report a global investigation of the upland cotton ovule and fiber proteomes using 0 DPA ovules and 10 DPA fibers. A total of 975 protein spots were successfully identified (95% peptide counts \geq 2 and peptides coverage $>$ 5%). Further functional annotation analysis provided global insights into cotton ovule and fiber proteomes.

RESULTS

2DE and MALDI-TOF/TOF identified 975 protein spots

Two-dimensional electrophoresis (2DE) of 0 DPA ovules and 10 DPA fibers were performed to investigate the global proteomes of these tissues. Three biological duplicates using independent cotton plants were performed for both ovules and fibers to generate reproducible, representative 2DE gels (Figure 1). In total, 1,895 protein spots were detected for ovules (1,328 from pH 4–7 and 567 from pH 7–10) and 1,742 protein spots for fibers (1,284 from pH 4–7 and 458 from pH 7–10). A total of 1,203 independent protein spots were excised from 2DE gels and subjected to MALDI-TOF/TOF analysis after in-gel trypsin digestion. Protein spots were considered successfully identified only when 95% confident peptide counts \geq 2, $P<0.05$, and peptides coverage $>$ 5%. As a result, 975 protein spots were successfully identified by searching against a self-constructed upland cotton database derived from the genome sequence data of Gh (Li et al., 2015), in which 411 and 564 hits were found for ovules and fibers, respectively. The protein spots were numbered as P1–P521 for ovules and 1–682 for fibers (Figure 1), and only successfully identified protein spots were labeled. Detailed information regarding these identified protein spots are listed in Tables S1 and S2. A total of 661 protein spots were identified using peptide mass fingerprinting data, and 314 were identified by MS/MS assays. To date, no other 2DE gel-based proteomic studies on cotton have reported protein identifications on such a large scale (Table 1). Thus, our work provides the most comprehensive data set of 2DE-based cotton ovule and fiber proteomes.

Annotation of tissue-specific proteins yielded differential distributions of GO terms between ovules and fibers

All of the 975 identified protein spots were mapped by MASCOT on 717 independent upland cotton genome loci, representing 466 unique UniProt ACC numbers, in which 119 loci were expressed in both ovules and fibers; 154 were ovule-specific (not identified in fiber) and 193 were fiber-specific (Figure 2A).

Of the 466 unique proteins, 387 were successfully mapped to GO annotations by BLAST2GO (Conesa et al., 2005). The distributions of the GO terms mapped according to the biological process category are shown in pie charts for commonly expressed proteins (Figure 2B), ovule-specific proteins (Figure 2C) and fiber-specific proteins (Figure 2D). In general, 119 commonly expressed proteins were mainly represented by GO terms for carbohydrate metabolic (22%), cellular amino acid metabolic (11%) and lipid metabolic (9%) processes, while 154 ovule-specific

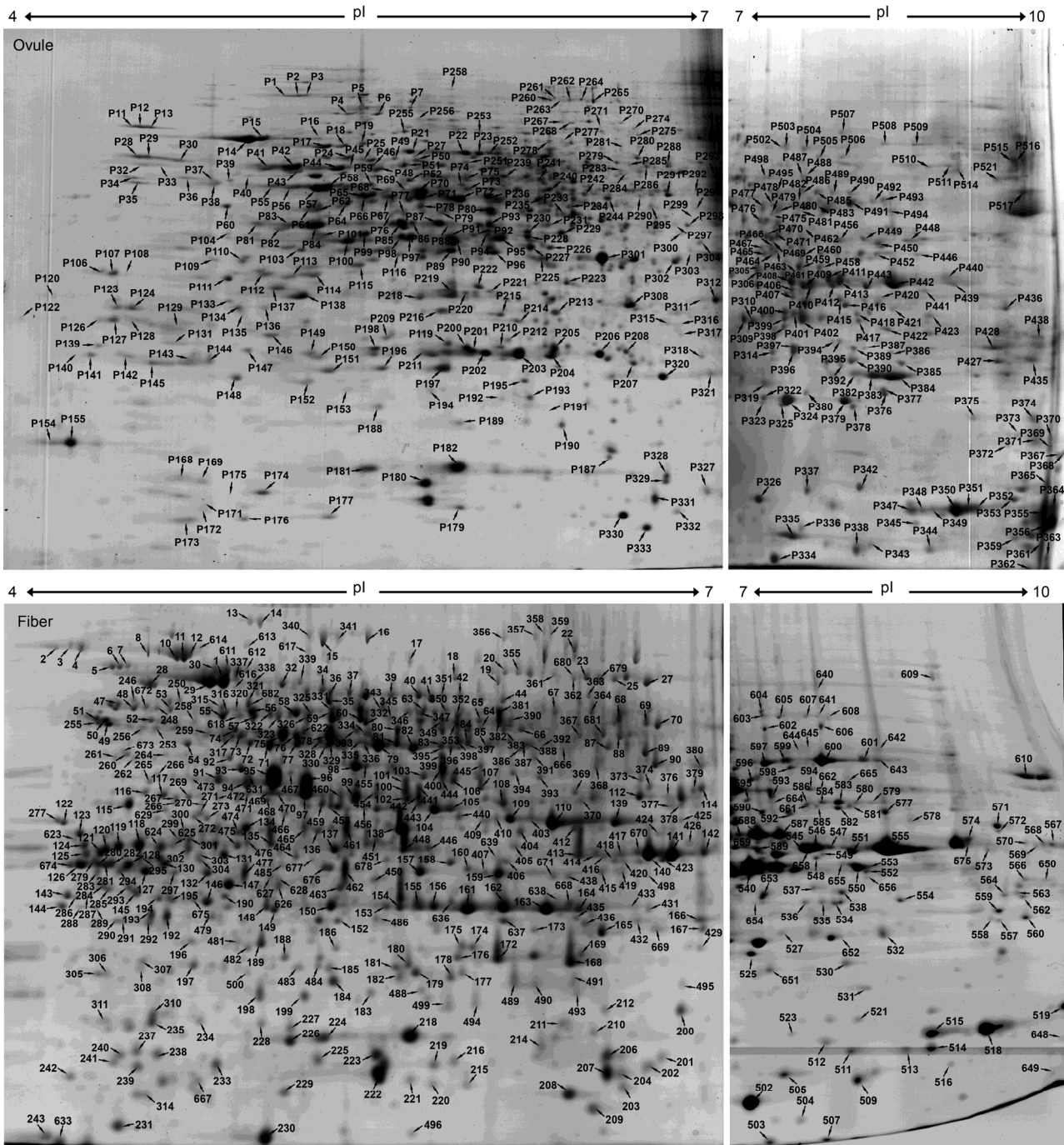


Figure 1 Representative 2DE maps of cotton ovules (top) and fibers (bottom). Both gels are shown as sections in the range from pH 4–7 and pH 7–10 of pH 3–10 strips. A total of 1,895 (1,328 from pH 4–7 and 567 from pH 7–10) and 1,742 (1,284 from pH 4–7 and 458 from pH 7–10) protein spots were detected in the 2DE gels of ovules and fibers, respectively. All protein spots examined by MS were numbered from P1 to P521 for ovules and from 1 to 682 for fiber, while only successfully identified protein spots (95% confident peptides > 2, $P < 0.05$, and peptide coverage > 5%) were labeled on the maps.

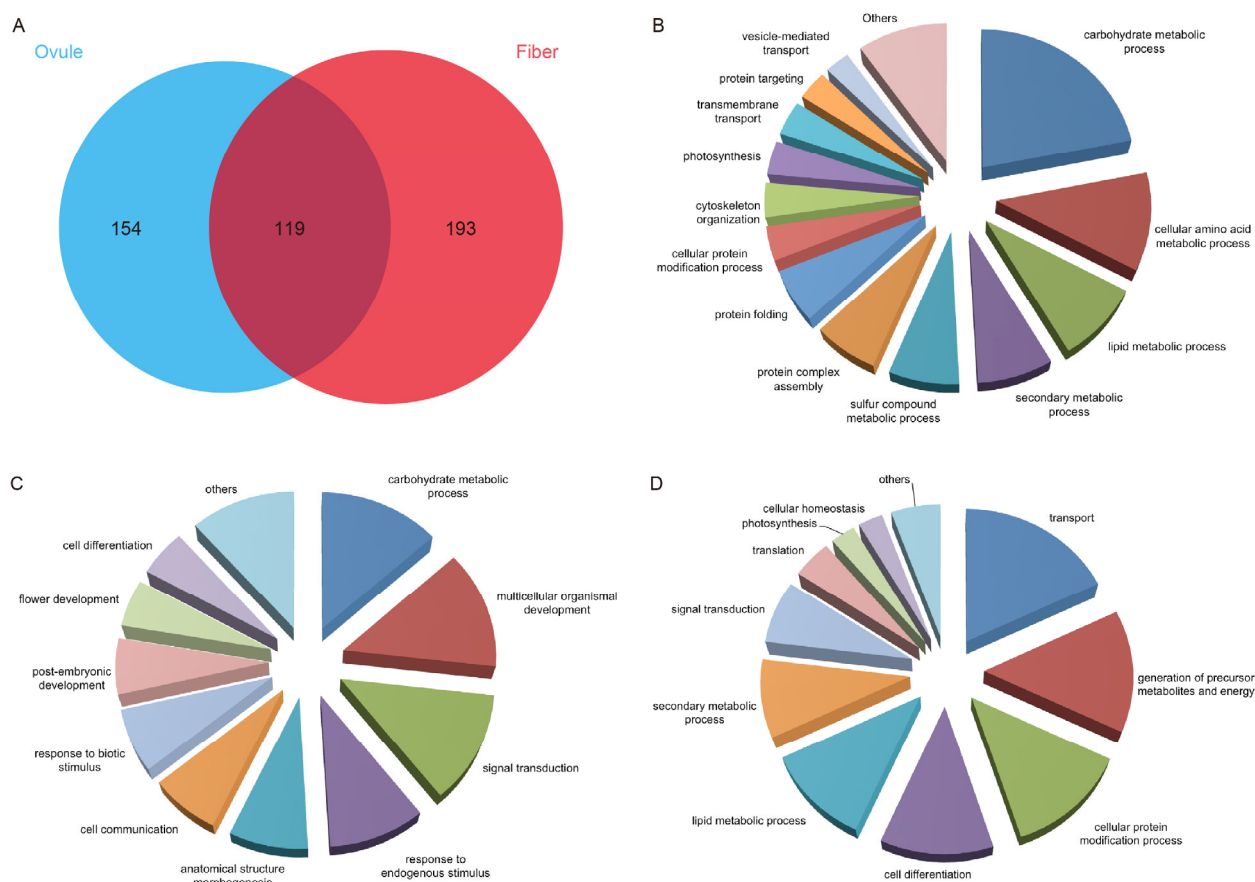
proteins were represented by carbohydrate metabolic (14%), multicellular organismal development (13%) and signal transduction (13%) processes, and 193 fiber-specific proteins were represented by transport (18%), generation of precursor metabolites and energy (14%) and cellular protein modification (13%) processes. These different GO term distribution patterns between ovules and fibers may be re-

lated to their biological functions and may indicate some important biological processes in ovules and fibers.

Furthermore, GO term enrichment analyses were performed using the GOEAST software toolkit (Zheng et al., 2008). GO IDs with FDR-corrected P -values < 0.05 were considered significantly enriched. The resulting GO tree maps are extremely complex; thus, they have been provided

Table 1 Brief summary of 2DE-based proteomic studies on cotton.

MS type	IPG strips	Identified proteins	Sample type	Reference
MS, MS/MS	3–10	10	fiber	Yao et al., 2006
MS, MS/MS	3–10	106	ovule, fiber	Yang et al., 2008
MS/MS	4–7	58	root	Coumans et al., 2009
MS	4–7, 3–10	96	fiber	Pang et al., 2010
MS/MS	4–7	81	fiber	Zhao et al., 2009
MS, MS/MS	3–10	37	fiber	Zheng et al., 2012
MS, MS/MS	4–7, 6–11	46	ovule	Liu et al., 2012
MS	4–7	57	root	Zhao et al., 2012
MS/MS	4–7	21	leaf	Deeba et al., 2012
MS, MS/MS	3–10	71	ovule, fiber	Du et al., 2013
MS/MS	4–7	188	root	Gao et al., 2013
MS, MS/MS	3–10	132	fiber	Zheng et al., 2014
MS	4–7	22	leaf	Cui et al., 2015
MS, MS/MS	4–7, 3–10	975	ovule, fiber	This work

**Figure 2** Functional classification of 975 identified protein spots. All of the 975 identified protein spots represented 466 unique UniProt ACC numbers. The Venn diagram of the overlapping proteins expressed in both ovules and fibers is shown (A). The biological process categories of the GO term distribution of the 119 commonly expressed proteins (B), 154 ovule-specific proteins (C) and 193 fiber-specific proteins (D) are shown.

as online supplementary materials (Figures S1–S3 in Supporting Information). The branch end terms in the tree maps of significantly enriched GO terms of the biological process categories indicate potentially important pathways represented by different protein sets of commonly expressed, ovule-specific and fiber-specific proteins. GO terms for

fatty acid elongation, sulfur amino acid metabolic processes and post-replication repair were enriched in ovule-specific proteins. Notably, for GO:0030497, fatty acid elongation was enriched in ovule-specific proteins, which is consistent with our previous work (Qin et al., 2007; Wang et al., 2015). Additionally, several previously reported biological

processes, such as root hair elongation, galactose metabolic process, and D-xylose metabolic process were attributed to the fiber-specific proteins (Figure S1–S3 in Supporting Information).

Top-abundant proteins showed different GO distribution patterns between ovules and fibers

To investigate the most abundant gene production in ovules and fibers, protein spots with gray volumes greater than 0.15% of total protein abundance were selected. A total of 82 identified protein spots on 2DE gels from ovules and 87 protein spots from fibers were used in subsequent analyses. All of the protein spots with gray volumes greater than 0.15% had similar distribution patterns in ovules and fibers (Figure 3A), representing 28.96% and 27.02% of the total protein abundances in ovules and fibers, respectively (Figure 3B).

Proteins are the main nitrogen-containing biological molecules in plants. Investigating the top-abundant proteins in different tissues could help determine how plants utilize nitrogen in particular tissues. Annotation analysis revealed

that 82 top-abundant ovule protein spots represented 77 unique gene productions, while 87 top-abundant fiber protein spots represented 70 unique gene productions, of which 48 were commonly expressed in both ovules and fiber, 29 proteins were ovule-specific and 22 were fiber-specific. The tree maps of the GO cellular component categories for ovule (Figure 3C) or fiber (Figure 3D) tissue-specific top abundant proteins are shown. Notably, in ovules, 28.96% of the total proteins are mainly located in the Golgi apparatus, endoplasmic reticulum, mitochondrion and ribosome, while in fibers, 27.02% of the total proteins are mainly located in the cytoskeleton, nuclear envelope and cell wall (GO terms enlarged and outlined with red circles in Figures 3C and D). These results indicate that ovules and fibers utilize nitrogen in different sub-cellular areas to satisfy different biological functions. The tree maps of the GO biological process categories for tissue-specific top-abundant proteins are also shown in Figures S4–S6. Notably, GO:0030154 in cell differentiation was only represented by fiber-specific proteins (Figure S6 in Supporting Information).

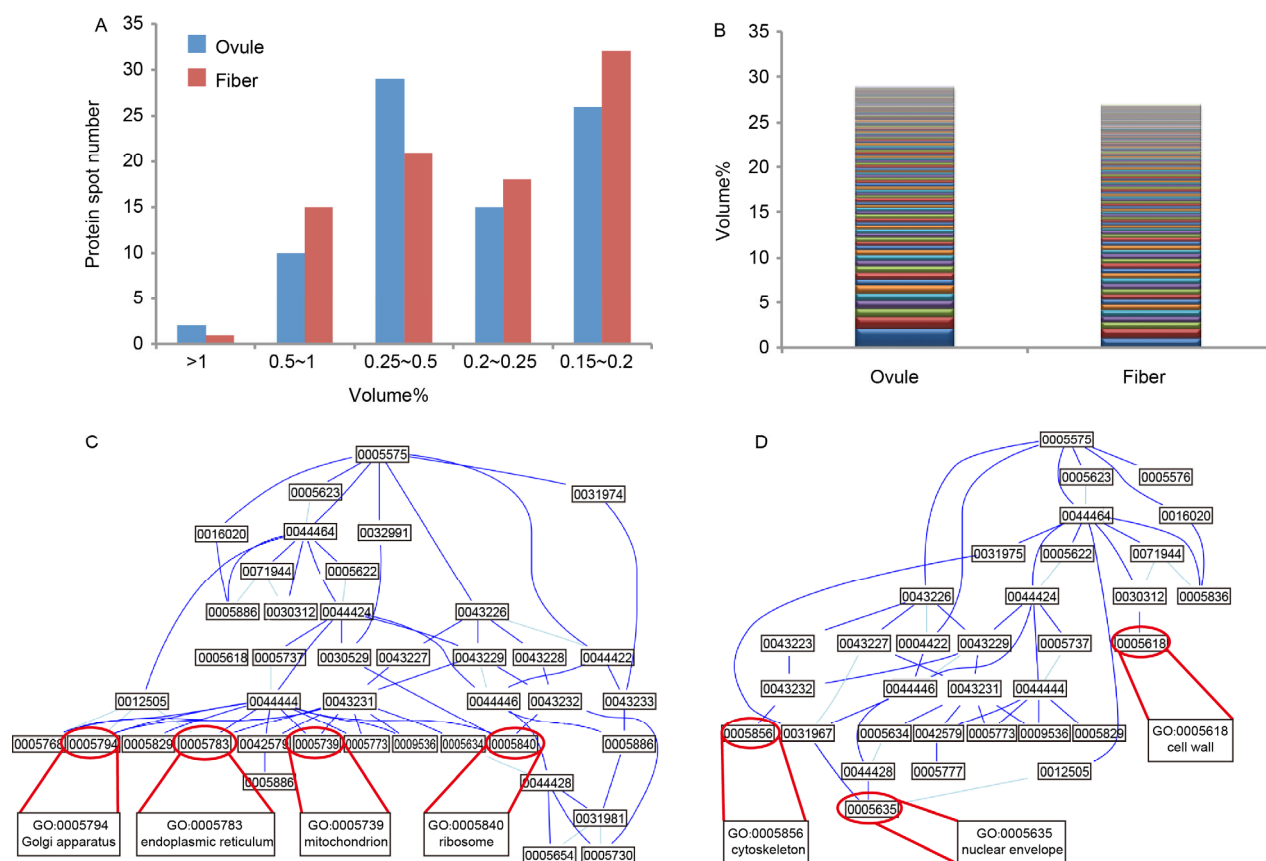


Figure 3 Gray volume and functional distributions of the most abundant (gray volume greater than 0.15%) protein spots in 2DE gels of ovules and fibers. A, Protein spot gray volume (percentage of total proteins) distribution of top-abundant protein spots in ovules and fibers. B, The total percentages of the most abundant protein spots were 28.96% in ovules and 27.02% in fibers. Different colors indicate the gray volumes of individual protein spots. C, Tree map of the GO cellular component category classifications of the most abundant proteins in ovules. D, Tree map of GO cellular component category classifications of the most abundant proteins in fibers. For C and D, potentially important tissue-specific subcellular locations are enlarged and highlighted with red circles. Each box is labeled with a specific GO ID. Lines represent relationships between two GO IDs; blue lines indicate “*is a*” relationships and light blue lines indicate “*part of a*” relationship.

qRT-PCR validation of representative tissue-specific proteins

Further qRT-PCR assays of several ovule- or fiber-specific proteins previously reported to be important in cotton fiber development were performed to validate the proteomics results. The enlarged gel maps for ovule-specific protein spots P61, P93 and P198 are shown, which represent eukaryotic initiation factor 4A-2 (eIF4A2), S-adenosylmethionine synthase 4 (METK4) and proteasome subunit alpha (PAB1), respectively (Figure 4A, top panel). The gray volume of each protein spot (Figure 4A, middle panel) and the qRT-PCR results represented in fold change of their transcripts (Figure 4A, bottom panel) are also shown. The transcript expression level in the ovule template was set to 1. The same experiments were performed for three fiber-specific protein spots, 230, 477 and 511, which represent profiling-3 (PRO3), actin-1 (ACT1) and ethylene-responsive family protein (ERFP), respectively. The enlarged gels, gray volumes and qRT-PCR fold changes are also shown (Figure 4B). The representative MS spectra of these protein spots or MS/MS spectra of particular peptides are also provided as representative data to show the reliability of our MS data (Figure S7 in Supporting Information). The primers used in the qRT-PCR assays are listed in Table S3.

Protein-protein interaction network analyses of tissue-specific proteins indicate potentially important proteins in ovules and fibers

To determine the potential key proteins in tissue-specific protein sets, significantly tissue-enriched proteins were subjected to STRING v10 analysis to identify the protein-protein interaction networks (Szklarczyk et al., 2015).

The results showed that the ovule-specific proteins had two clusters of protein-protein interaction networks. One cluster was mainly centered on aspartate carbamoyltransferase (PYRB), serine hydroxymethyltransferase 2 (SHM2), GDP-mannose 3,5-epimerase (At5g28840) and delta-1-pyrroline-5-carboxylate dehydrogenase 12A1 (ALDH12-A1), while the other cluster consisted of proliferating cell nuclear antigen 2 (PCNA2) and ubiquitin-conjugating enzyme E2 35 (UBC35), which are primarily represented in carboxylic acid metabolic processes and post-replication repair pathways (Figure 5A). The fiber-specific proteins contributed to four sub-clusters of protein-protein interaction networks. One cluster mainly centered on small ubiquitin-related modifier 1 (SUMO1); the second one mainly consisted of pyruvate dehydrogenase E1 component subunit alpha-2 (IAR4); the third one was centered on UDP-glucose-6-dehydrogenase 2 (UGD2) and UTP-glucose-1-phosphate uridylyltransferase (UGP2); and the last one mainly consisted of eukaryotic translation initiation factor 2 (eIF2), 26S protease regulatory subunit 6B homolog (RPT3) and OB-fold nucleic acid binding domain-containing protein (At2g04520), which are represented in the biological

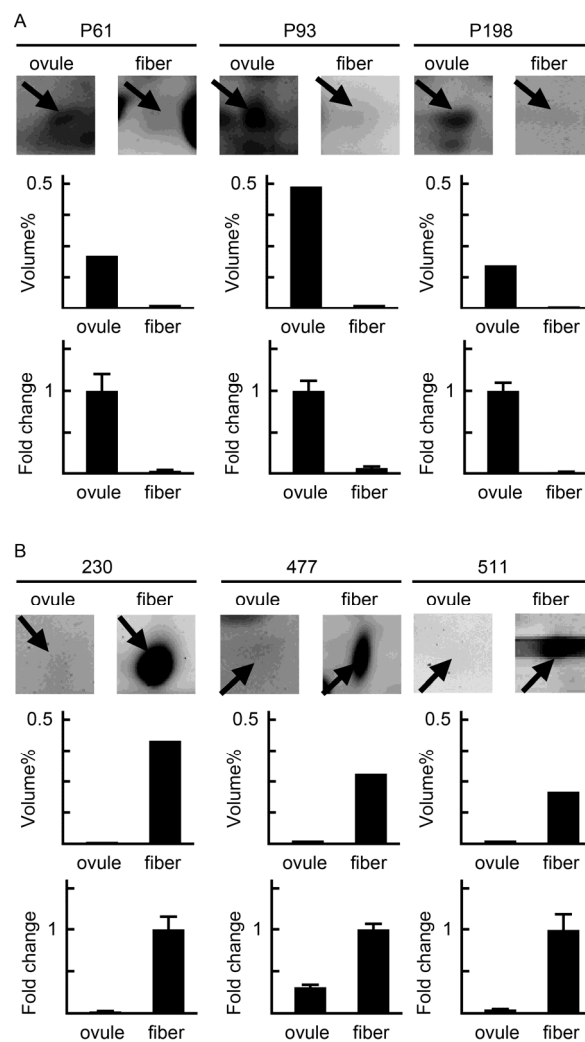


Figure 4 qRT-PCR validation of tissue-specific proteins. A, Enlarged 2DE gels (top panel), gray volumes (middle panel) and qRT-PCR assay results (bottom panel) of representative ovule-specific protein spots P61, P93 and P198, which represent eIF4A2, METK4 and PAB1, are shown. The transcript expression levels in ovules were set to 1. B, Enlarged 2DE gels (top panel), gray volumes (middle panel) and qRT-PCR assay results (bottom panel) of fiber-specific protein spots 230, 477 and 511, which represent PRO3, ACT1 and ERFP, are shown. The transcript expression levels in fibers were set to 1.

processes of root hair elongation, galactose metabolic process, and D-xylose metabolic process (Figure 5B).

DISCUSSION

The most comprehensive 2DE profile data set of cotton ovules and fibers

Cotton proteomic studies have made significant progress over the past ten years. Limited by the throughput and cost of 2DE-based approaches, all studies on the cotton proteome have identified no more than 200 proteins on 2DE gels (Table 1). Most previous works only identified proteins that were preferentially accumulated at a particular develop-

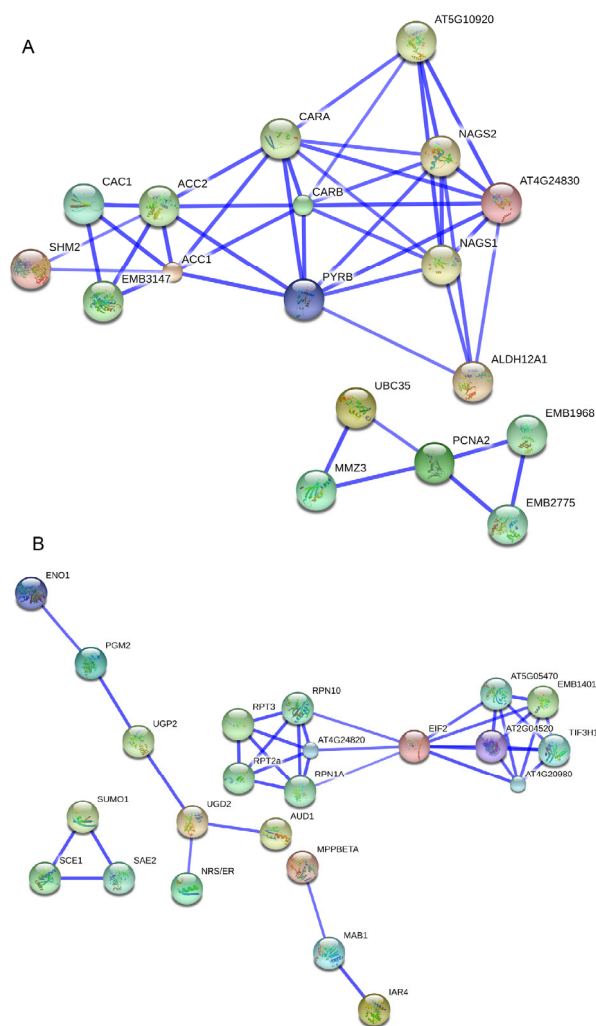


Figure 5 Protein-protein interaction network analyses of tissue-specific proteins. A, Protein-protein interaction networks of ovule-specific proteins. B, Protein-protein interaction networks of fiber-specific proteins. Colored lines between the proteins indicate various types of interaction evidence. Protein nodes that are enlarged indicate the availability of three dimensional (3D) protein structure information. AT5G10920, argininosuccinate lyase. CARA, carbamoyl phosphate synthetase small subunit. CARB, carbamoyl phosphate synthetase large subunit. NAGS1, N-acetyl-I-glutamate synthase 1. NAGS2, N-acetyl-I-glutamate synthase 2. CAC1, chloroplastic acetyl coenzyme a carboxylase 1. ACC1, acetyl-CoA carboxylase 1. ACC2, acetyl-CoA carboxylase 2. AT4G24830, argininosuccinate synthase. SHM2, serine hydroxymethyltransferase 2. PYRB, pyrimidine B. EMB3147, embryo defective 3147. ALDH12A1, aldehyde dehydrogenase 12A1. UBC35, ubiquitin conjugating enzyme 35. MMZ3, MMS ZWEI homologue 3. PCNA2, proliferating cell nuclear antigen 2. EMB1968, embryo defective 1968. EMB2775, embryo defective 2775. SUMO1, small ubiquitin-like modifier 1. SCE1, SUMO conjugating enzyme 1A. SAE2, SUMO conjugating enzyme 2. ENO1, enolase 1. PGM2, phosphoglucomutase 2. UGP2, UDP-glucose pyrophosphorylase 2. NRS/ER, nucleotide-rhamnose synthase/epimerase-reductase UER1. AUD1, UDP-glucuronic acid decarboxylase 2. MPPBETA, metalloendopeptidase. MAB1, macci-bou. IAR4, IAA-conjugate-resistant 4. RPT3, regulatory particle triple-a ATPase 3. RPT2a, regulatory particle AAA-ATPase 2A. RPN10, regulatory particle non-ATPase 10. AT4G24820, 26S proteasome regulatory subunit Rpn7. RPN1A, 26S proteasome regulatory subunit S2 1A. EIF2, eukaryotic translation initiation factor 2. AT5G05470, eukaryotic translation initiation factor 2 A2. AT2G04520, translation initiation factor. AT4G20980, eukaryotic translation initiation factor 3 subunit 7. EMB1401, eukaryotic translation initiation factor 2 beta subunit. TIF3H1, translation initiation factor 3 subunit H1.

mental stage or tissue. However, scarce information was available about the largest portion of 2DE-detectable protein spots, usually more than 1,000 protein spots, beyond differentially expressed proteins. Gel-free high-throughput MS approaches have been rapidly developed in recent years and have identified 1,000–5,000 or more proteins in cotton tissues (Hu et al., 2013; Meng et al., 2011; Wang et al., 2015). However, these data were still lacking global proteome information in some aspects that could be provided by 2DE maps, such as protein spot visualization, reliable evidence for existing protein isoforms and the abundances of individual protein spots.

This work provided the most comprehensive data set of 2DE profiling of cotton ovules and fibers by successfully identifying 975 protein spots on 2DE gels. The representative 2DE maps coupled with the detailed MS or MS/MS identity information constructed a useful database for the 2DE-based proteome to further study cotton ovule and fiber proteomes (Figure 1, Tables S1 and S2 in Supporting Information).

Several protein spots on the 2DE gels were represented by the same transcript. This finding is very common in 2DE gels; however, few works have discussed this phenomenon. The possible reasons include post-translational modification, alternative splicing, site-specific truncated degradation of proteins and protein-protein interactions (especially for protein spots with experimental molecular weights substantially greater than theoretical molecular weights). However, few experiments have been conducted to validate any of these possibilities. The 975 identified protein spots represented 717 independent cotton genome loci and 466 unique UniProt ACC numbers, providing a large data set of protein species for investigating the different gene transcriptional patterns between ovules and fibers. The possibility exists that a particular gene product can be commonly expressed in both ovules and fibers, even though the protein spot was not successfully identified in either ovules or fibers, providing false-positive results for annotation analysis. However, we performed qRT-PCR validation of transcripts (Figure 4) and narrowed the range of tissue-specific proteins by using the unique UniProt ACC numbers instead of independent cotton genome loci. A combination of 2DE gel-based and gel-free high-throughput MS technologies may be helpful to address these concerns.

Annotation and network analyses of tissue-specific proteins and commonly expressed proteins provided global insight into the 2DE-detectable proteomes of cotton ovules and fibers

The annotation and network analyses of tissue-specific proteins provided potential functional roles for these proteins. Most of the results were reasonable and consistent with previous findings. Lipid metabolic processes have been reported to be important during fiber initiation (Qin et al., 2007; Wang et al., 2015), and we found that 119 commonly

expressed proteins were classified with GO terms of carbohydrate metabolic, cellular amino acid metabolic and lipid metabolic processes (Figure 2B). The biosynthesis of pectic precursors has been reported to play important roles during fiber elongation (Pang et al., 2010), and we also found that galactose metabolic and D-xylose metabolic processes were enriched in fiber-specific proteins. Protein-protein interaction network analyses also recognized two previously reported proteins, UGD2 and UGP2 (Pang et al., 2010), as the centers of the sub-network clusters (Figures S1–S3 in Supporting Information, Figure 4).

In conclusion, our work has provided the most comprehensive data set of 2DE-detectable proteins of cotton ovules and fibers. A total of 975 protein spots were successfully identified on the same coverage scale as gel-free LC-MS/MS technologies. Taking advantage of the 2DE approach, our data provided visualization of 2DE maps, information about protein isoforms and abundant volumes of individual protein spots, thus providing a global level of insight into the cotton ovule and fiber proteomes, which will enable further investigations.

MATERIALS AND METHODS

Cotton materials

Cotton plants (Gh cv. Xuzhou 142) were grown in fully automated walk-in growth rooms with 60% relative humidity at 34°C in the light and 28°C in the dark (12 h light/dark cycle). Ovules were harvested on the day post anthesis (0 DPA), and fibers were separated from the ovules after the cotton bolls were harvested on 10 DPA.

Protein extraction and 2DE

Total protein extraction was performed as described previously (Wang et al., 2007). The concentrations of protein samples were determined by a Bradford assay, and 1,200 µg of proteins from ovules or fibers were loaded onto 24 cm IPG strips (immobilized pH gradient, GE Healthcare, USA) with linear pH gradients from 4–7 or 3–10, respectively. The SDS-PAGE gels were visualized by GAP staining methods (Wang et al., 2012b) and analyzed using Image Master 2D Platinum Software (Version 5.0, GE Healthcare, USA). Three biological duplicates of 2DE using proteins extracted from independent cotton plants were performed for both ovules and fibers to generate reproducible 2DE maps of each.

In-gel digestion and MALDI-TOF/TOF analysis

Protein spots of interest were manually excised from 2-DE gels and washed with MilliQ water three times for 30 min each to remove possible contaminants on the surfaces of the gels. Protein spots were washed another three times with destaining solution containing 50 mmol L⁻¹ NH₄HCO₃ and 50% ACN for 30 min each at 37°C and then incubated in

100 µL of 100% ACN until the gel pieces turned white. The gel pieces were air-dried at 25°C for 1 h.

Proteins were digested in-gel with modified bovine trypsin (Roche, Cat. 11418025001) as described previously (Wang et al., 2009). Trypsin-digested peptide extracts were identified using an AB SCIEX MALDI-TOF/TOF 5800 system (AB SCIEX, USA) with a laser wavelength of 349 nm as previously described (Wang et al., 2013). The peptide mass fingerprinting peak maps were internally calibrated with trypsin autolysis peaks, and all known contaminants were excluded. An additional MS/MS assay was performed on the protein spots that were undetermined by peptide mass fingerprinting.

Protein identification by MASCOT

The MS and MS/MS raw data were combined and searched against a self-constructed database derived from original Gh genome data (Li et al., 2015; Zhang et al., 2015) that included 76,943 protein sequences using an in-house MASCOT server (Matrix Science, USA). Proteins were considered to be successfully identified only in the case of 95% confidence with peptide counts ≥ 2, $P < 0.05$, and peptide coverage > 5%.

Protein annotation and Gene Ontology (GO) enrichment analyses

For successfully identified proteins from cotton genome loci hits, a local BLAST search against the UniProt (<http://www.uniprot.org/>) database was performed using genome-predicted cotton protein sequences to map the identified protein spots to functional annotations. The identified proteins were then categorized by searching against Gene Ontology (<http://www.geneontology.org>) using BLAST2-GO software (Conesa et al., 2005). Mapped proteins were then submitted to GO Slim Viewer (<http://www.agbase.msstate.edu/index.html>) to obtain the GO ID distribution (McCarthy et al., 2006). The visualized GO ID tree maps were generated using the Visualize tool of Amigo (Carbon et al., 2009). The GO ID enrichment analyses were performed using the GOEAST (<http://omicslab.genetics.ac.cn/GOEAST/>) software toolkit (Zheng et al., 2008). GO IDs with false discovery rate (FDR)-corrected P -values < 0.05 were considered significantly enriched.

RNA extraction and qRT-PCR

Cotton ovules and fibers were first frozen in liquid nitrogen before being ground to a fine powder with a mortar and pestle using a modified hot borate method. Total RNA was extracted from the cotton ovules as previously reported (Shi et al., 2006), and cDNA was reverse-transcribed from 1 µg of total RNA. We used ubiquitin (UBQ) as a housekeeping gene for all qRT-PCR experiments (primers listed in Table S3). All qRT-PCR experiments were performed in triplicate using independent RNA samples prepared from different cotton plants.

Protein-protein interaction network analyses

Proteins mapped to significantly tissue enriched GO IDs (recognized by GOEAST) were further analyzed using the STRING V.10 (<http://string-db.org>) database (Szklarczyk et al., 2015) to analyze the protein-protein interaction networks, which provided information on the systematic structures of functionally important proteins.

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

Acknowledgements This work was supported by the Special Fund for Agro-scientific Research in the Public Interest of the People's Republic of China (201403075), Major Technology Project of Hainan (ZDZX2013010-1), Program for Top Young Talents in the Chinese Academy of Tropical Agricultural Sciences (ITBB130102), and China Postdoctoral Science Foundation (20110490003). We thank Professor Zhu Yu-Xian from Peking University for providing the cotton materials and directive advice for this work. We also thank Professor Feng Ji-Dong from China Agriculture University for providing generous help on MS identification of proteins.

- Cao, X. (2015). Whole genome sequencing of cotton—a new chapter in cotton genomics. *Sci China Life Sci* 58, 515–516.
- Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., AmiGO Hub, and Web Presence Working Group. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288–289.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.
- Coumans, J.V., Poljak, A., Raftery, M.J., Backhouse, D., and Pereg-Gerk, L. (2009). Analysis of cotton (*Gossypium hirsutum*) root proteomes during a compatible interaction with the black root rot fungus *Thielaviopsis basicola*. *Proteomics* 9, 335–349.
- Cui, Y., Lu, X., Wang, D., Wang, J., Yin, Z., Fan, W., Wang, S., and Ye, W. (2015). Comparative analysis of salinity-induced proteomic changes in cotton (*Gossypium hirsutum* L.). *Agr Sci* 6, 78–86.
- Deeba, F., Pandey, A.K., Ranjan, S., Mishra, A., Singh, R., Sharma, Y.K., Shirke, P.A., and Pandey, V. (2012). Physiological and proteomic responses of cotton (*Gossypium herbaceum* L.) to drought stress. *Plant Physiol Bioch* 53, 6–18.
- Du, S., Dong, C., Zhang, B., Lai, T., Du, X., and Liu, J. (2013). Comparative proteomic analysis reveals differentially expressed proteins correlated with fuzz fiber initiation in diploid cotton (*Gossypium arboreum* L.). *J Proteomics* 82, 113–129.
- Fan, S., Meng, Y., Song, M., Pang, C., Wei, H., Liu, J., Zhan, X., Lan, J., Feng, C., Zhang, S., and Yu, S. (2014). Quantitative phosphoproteomics analysis of nitric oxide-responsive phosphoproteins in cotton leaf. *PLoS One* 9, e94261.
- Gao, W., Long, L., Zhu, L., Xu, L., Gao, W., Sun, L., Liu, L., and Zhang, X. (2013). Proteomic and virus-induced gene silencing (VIGS) analyses reveal that gossypol, brassinosteroids, and jasmonic acid contribute to the resistance of cotton to *Verticillium dahliae*. *Mol Cell Proteomics* 12, 3690–3703.
- He F. (2013). Lifeomics leads the age of grand discoveries. *Sci China Life Sci* 56, 201–212.
- Hu, G., Koh, J., Yoo, M.J., Grupp, K., Chen, S., and Wendel, J.F. (2013). Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytol* 200, 570–582.
- Huang, Q., Wang, H., Gao, P., Wang, G., and Xia, G. (2008). Cloning and characterization of a calcium dependent protein kinase gene associated with cotton fiber development. *Plant Cell Rep* 27, 1869–1875.
- Ji, S., Lu, Y., Li, J., Wei, G., Liang, X., and Zhu, Y. (2002). A beta-tubulin-like cDNA expressed specifically in elongating cotton fiber induces longitudinal growth of fission yeast. *Biochem Biophys Res Commun* 296, 1245–1250.
- Ji, S., Lu, Y., Feng, J., Wei, G., Li, J., Shi, Y., Fu, Q., Liu, D., Luo, J., and Zhu, Y. (2003). Isolation and analyses of gene specific during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res* 31, 2534–2543.
- Jin, X., Li, Q., Xiao, G., and Zhu, Y. (2013a). Using genome-referenced expressed sequence tag assembly to analyze the origin and expression patterns of *Gossypium hirsutum* transcripts. *J Integr Plant Biol* 55, 576–585.
- Jin, X., Pang, Y., Jia, F., Xiao, G., Li, Q., and Zhu, Y. (2013b). A potential role for CHH DNA methylation in cotton fiber growth patterns. *PLoS One* 8, e60547.
- Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., Chen, W., Liang, X., Shang, H., Liu, W., Shi, C., Xiao, G., Gou, C., Ye, W., Xu, X., Zhang, X., Wei, H., Li, Z., Zhang, G., Wang, J., Liu, K., Kohel, R., Percy, R., Yu, J., Zhu, Y., Wang, J., and Yu, S. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat Genet* 46, 567–572.
- Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R., Ma, Z., Shang, H., Ma, X., Wu, J., Liang, X., Huang, G., Percy, R., Liu, K., Yang, W., Chen, W., Du, X., Shi, C., Yuan, Y., Ye, W., Liu, X., Zhang, X., Liu, W., Wei, S., Huang, G., Zhang, X., Zhu, S., Zhang, H., Sun, F., Wang, X., Liang, J., Wang, J., He, Q., Huang, L., Wang, J., Cui, J., Song, G., Wang, K., Xu, X., Yu, J., Zhu, Y., and Yu, S. (2015). Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat Biotechnol* 33, 524–530.
- Li, H., Qin, Y., Pang, Y., Song, W., Mei, W., and Zhu, Y. (2007). A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fiber cell development. *New Phytol* 175, 462–471.
- Li, Q., Jin, X., and Zhu, Y. (2012). Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid *G. raimondii* genome. *J Genet Genomics* 39, 351–360.
- Liu, K., Han, M., Zhang, C., Yao, L., Sun, J., and Zhang, T. (2012). Comparative proteomic analysis reveals the mechanisms governing cotton fiber differentiation and initiation. *J Proteomics* 75, 845–856.
- Ma, Q., Wu, M., Pei, W., Li, H., Li, X., Zhang, J., Yu, J., and Yu, S. (2014). Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. *BMC Genomics* 15, 466.
- McCarthy, F.M., Wang, N., Magee, G.B., Nanduri, B., Lawrence, M.L., Camon, E.B., Barrell, D.G., Hill, D.P., Dolan, M.E., Williams, W.P., Luthe, D.S., Bridges, S.M., and Burgess, S.C. (2006). AgBase: a functional genomics resource for agriculture. *BMC Genomics* 7, 1–13.
- Meng, Y., Liu, F., Pang, C., Fan, S., Song, M., Wang, D., Li, W., and Yu, S. (2011). Label-free quantitative proteomics analysis of cotton leaf response to nitric oxide. *J Proteome Res* 10, 5416–5432.
- Pang, C., Wang, H., Pang, Y., Xu, C., Jiao, Y., Qin, Y., Western, T.L., Yu, S., and Zhu, Y. (2010). Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and *Arabidopsis* root hair elongation. *Mol Cell Proteomics* 9, 2019–2033.
- Pei, Y. (2015). The homeodomain-containing transcription factor, GhHOX3, is a key regulator of cotton fiber elongation. *Sci China Life Sci* 58, 309–310.
- Qin, Y., Hu, C., Pang, Y., Kastaniotis, A.J., Hiltunen, J.K., and Zhu, Y. (2007). Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *Plant Cell* 19, 3692–3704.
- Qin, Y., and Zhu, Y. (2011). How cotton fibers elongate: a tale of linear cell-growth mode. *Curr Opin Plant Biol* 14, 106–111.
- Shi, Y., Zhu, S., Mao, X., Feng, J., Qin, Y., Zhang, L., Cheng, J., Wei, L., Wang, Z., and Zhu, Y. (2006). Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell*, 18, 651–664.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., and von Mering, C. (2015). STRING

- v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43, D447–D452.
- Wang, H., Mei, W., Qin, Y., and Zhu, Y. (2011). 1-Aminocyclopropane-1-carboxylic acid synthase 2 is phosphorylated by calcium-dependent protein kinase 1 during cotton fiber elongation. *Acta Biochim Biophys Sin* 43, 654–661.
- Wang, K., Wang, Z., Li, G., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., Zou, C., Li, Q., Yuan, Y., Lu, C., Wei, H., Gou, C., Zheng, Z., Yin, Y., Zhang, X., Liu, K., Wang, B., Song, C., Shi, N., Kohel, R., Percy, R., Yu, J., Zhu, Y., Wang, J., and Yu, S. (2012a). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet* 44, 1098–1103.
- Wang, X., Chang, L., Wang, B., Wang, D., Li, P., Wang, L., Yi, X., Huang, Q., Peng, M., and Guo, A. (2013). Comparative proteomics of *Thellungiella halophila* leaves from plants subjected to salinity reveals the importance of chloroplastic starch and soluble sugars in halophyte salt tolerance. *Mol Cell Proteomics*, 12, 2174–2195.
- Wang, X., Fan, P., Song, H., Chen, X., Li, X., and Li, Y. (2009). Comparative proteomic analysis of differentially expressed proteins in shoots of *Salicornia europaea* under different salinity. *J Proteome Res* 8, 3331–3345.
- Wang, X., Li, Q., Jin, X., Xiao, G., Liu, G., Liu, N., and Qin, Y. (2015). Quantitative proteomics and transcriptomics reveal key metabolic processes associated with cotton fiber initiation. *J Proteomics* 114c, 16–27.
- Wang, X., Li, X., Deng, X., Han, H., Shi, W., and Li, Y. (2007). A protein extraction method compatible with proteomic analysis for the euhalophyte *Salicornia europaea*. *Electrophoresis* 28, 3976–3987.
- Wang, X., Wang, D., Wang, D., Wang, H., Chang, L., Yi, X., Peng, M., and Guo, A. (2012b). Systematic comparison of technical details in CBB methods and development of a sensitive GAP stain for comparative proteomic analysis. *Electrophoresis* 33, 296–306.
- Yang, Y., Bian, S., Yao, Y., and Liu, J. (2008). Comparative proteomic analysis provides new insights into the fiber elongating process in cotton. *J Proteome Res* 7, 4623–4637.
- Yao, Y., Yang, Y., and Liu, J. (2006). An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. *Electrophoresis* 27, 4559–4569.
- Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., Hulse-Kemp, A.M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., Chang, L., Zhang, W., Song, Q., Kirkbride, R.C., Chen, X., Dennis, E., Llewellyn, D.J., Peterson, D.G., Thaxton, P., Jones, D.C., Wang, Q., Xu, X., Zhang, H., Wu, H., Zhou, L., Mei, G., Chen, S., Tian, Y., Xiang, D., Li, X., Ding, J., Zuo, Q., Tao, L., Liu, Y., Li, J., Lin, Y., Hui, Y., Cao, Z., Cai, C., Zhu, X., Jiang, Z., Zhou, B., Guo, W., Li, R., and Chen, Z. (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* 33, 531–537.
- Zhao, F., Fang, W., Xie, D., Zhao, Y., Tang, Z., Li, W., Nie, L., and Lv, S. (2012). Proteomic identification of differentially expressed proteins in *Gossypium thurberi* inoculated with cotton *Verticillium dahliae*. *Plant Sci* 185, 176–184.
- Zhao, P., Wang, L., Han, L., Wang, J., Yao, Y., Wang, H., Du, X., Luo, Y., and Xia, G. (2009). Proteomic identification of differentially expressed proteins in the ligan lintless mutant of upland cotton (*Gossypium hirsutum* L.). *J Proteome Res* 9, 1076–1087.
- Zheng, M., Wang, Y., Liu, K., Shu, H., and Zhou, Z. (2012). Protein expression changes during cotton fiber elongation in response to low temperature stress. *J Plant Physiol* 169, 399–409.
- Zheng, M., Meng, Y., Yang, C., Zhou, Z., Wang, Y., and Chen, B. (2014). Protein expression changes during cotton fiber elongation in response to drought stress and recovery. *Proteomics* 14, 1776–1795.
- Zheng, Q., and Wang, X. (2008). GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res* 36, W358–W363.
- Zhou, C., Lu, C., Shang, H., Jing, X., Cheng, H., Zhang, Y., and Song, G. (2013). Genome-wide analysis of the *Sus* gene family in cotton. *J Integr Plant Biol* 55, 643–653.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

SUPPORTING INFORMATION

Figures S1–S3 GO enrichment analysis of 119 commonly expressed proteins (Figure S1), ovule-specific proteins (Figure S2) and fiber-specific proteins (Figure S3).

Figures S4–S6 Tree maps of biological process categories of GO classification for 48 commonly expressed top-abundant proteins (Figure S4), 29 ovule-specific proteins (Figure S5) and 22 fiber-specific proteins (Figure S6).

Figure S7 Representative MS and MS/MS spectra of ovule-specific protein spots P61, P93 and P198 and fiber-specific protein spots 230, 477 and 511.

Table S1 Detailed information regarding the 411 identified protein spots on 2DE gels of cotton ovule.

Table S2 Detailed information regarding the 564 identified protein spots on 2DE gels of cotton fiber.

Table S3 Primers used in this work.

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