Planta (2014) 240:1237–1251
DOI 10.1007/s00425-014-2146-7

ORIGINAL ARTICLE

Proteomics profiling of fiber development and domestication in upland cotton (*Gossypium hirsutum* L.)

Guanjing Hu · Jin Koh · Mi-Jeong Yoo · Dharminder Pathak · Sixue Chen · Jonathan F. Wendel

Received: 9 May 2014 / Accepted: 7 August 2014 / Published online: 26 August 2014

Abstract Comparative proteomic analyses were performed to detail the evolutionary consequences of strong directional selection for enhanced fiber traits in modern upland cotton (*Gossypium hirsutum* L.). Using two complementary proteomic approaches, 2-DE and iTRAQ LC–MS/MS, fiber proteomes were examined for four representative stages of fiber development. Approximately 1,000 protein features were characterized using each strategy, collectively resulting in the identification and functional categorization of 1,223 proteins. Unequal contributions of homoeologous proteins were detected for over a third of the fiber proteome, but overall expression was balanced with respect to the genome-of-origin in the allopolyploid *G. hirsutum*. About 30 % of the proteins were differentially expressed during fiber development within wild and domesticated cotton. Notably, domestication was accompanied by a doubling of protein developmental dynamics for the period between 10 and 20 days following pollination. Expression levels of 240 iTRAQ proteins and 293 2-DE spots were altered by domestication, collectively representing multiple cellular and metabolic processes, including metabolism, energy, protein synthesis and destination, defense and stress response. Analyses of homoeolog-specific expression indicate that duplicated gene products in cotton fibers can be differently regulated in response to selection. These results demonstrate the power of proteomics for the analysis of crop domestication and phenotypic evolution.

Keywords Domestication · Evolutionary proteomics · Fiber development · *Gossypium* · Homoeolog expression · Polyploid

Abbreviations

2-DE	Two-dimensional gel electrophoresis
iTRAQ	Isobaric tag for relative and absolute quantification
LC	Liquid chromatography
MS	Mass spectrometry

Introduction

Lint fibers of upland cotton (*Gossypium hirsutum* L.) provide the most important natural fibers for the textile industry, which support the manufacture of diverse

Electronic supplementary material The online version of this article (doi:10.1007/s00425-014-2146-7) contains supplementary material, which is available to authorized users.

G. Hu · D. Pathak · J. F. Wendel (✉)
Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011, USA
e-mail: jfw@iastate.edu

J. Koh · M.-J. Yoo · S. Chen
Department of Biology, University of Florida, Gainesville, FL 32610, USA

J. Koh · S. Chen
Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL 32610, USA

Present Address:
D. Pathak
Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana 141004, Punjab, India

S. Chen
Genetics Institute, University of Florida, Gainesville, FL 32610, USA

consumer and industrial products throughout the world. In addition to their agronomic and economic importance, cotton fibers provide an excellent single-celled model for the study of basic biological processes in plants (Haigler et al. 2012). Each cotton "fiber" is a single, remarkably elongated cell that differentiates from epidermal cells of the ovular integument. Following cell differentiation, which typically occurs 2–3 days before anthesis, fiber cells become dramatically elongated during the next 20 days post-anthesis (dpa), followed by cell wall thickening via secondary wall synthesis (Haigler et al. 2012). At maturity, fibers of commercial upland cotton lines typically reach 30 mm in length, being one to two thousand times greater in length than in diameter (http://www.cottoninc.com). Because they are readily harvested and separated from other plant cells, cotton fibers provide a superior system for the study of plant cell wall development, maturation, and programmed cell death.

The most important commercial cotton species, *G. hirsutum*, is an allopolyploid ($2n = 4x = 52$) descended from a hybridization and genome doubling event 1–2 million years ago between a diploid, A-genome *Gossypium* species from Africa or Asia, and a D-genome diploid similar to that found in Central and South America (Wendel 1989). This polyploidy event, occurring after 5–10 million years of evolutionary divergence between the diploid progenitors, also gave birth to additional allopolyploid species including *G. barbadense*, whose cultivated form is often known as Egyptian or Pima cotton (Wendel et al. 2012). Both *G. hirsutum* and *G. barbadense* were independently domesticated for their seed hairs, and possibly as oil-seed crops, at least 5,000 years ago in the Yucatan Peninsula and the inter-montane Peruvian Andes areas, respectively (Wendel et al. 2010). As shown in Supplementary material Fig. S1, *G. hirsutum* has a large indigenous range encompassing many of the drier regions of the tropical and subtropical Americas, where it occurs in a continuum of morphological forms ranging from wild to domesticated. Truly wild plants, of race *yucatanense*, are sprawling perennial shrubs often found as a dominant component of native beach strand vegetation in the Yucatan Peninsula (Brubaker and Wendel 1994).

In contrast to the modern crop with greatly enhanced fiber fineness, strength and length, the shorter, coarser, and unspinnable fibers of the wild plants provide the antecedent that can be used in a comparative framework to understand the evolutionary basis of the dramatic morphological transformation that led to modern, agronomically improved cotton. Growth curve analyses of cotton fibers from wild vs. domesticated plants demonstrated that long fibers are associated with increased growth rate during primary wall synthesis and a prolonged period of fiber elongation (Applequist et al. 2001). The modern *G. hirsutum* cultivar TM1 reached a maximum growth rate between 10 and 15 dpa, whereas the wild accession Tx2094 displayed little or no growth between 10 and 15 dpa, followed by a significantly increased rate between 15 and 20 dpa. Microscopic observation indicated that 20 dpa is near the beginning of secondary wall synthesis for both accessions, and that domesticated cotton is only slightly ahead of the wild form in this respect. At maturation, the final fiber length of Tx2094 is about 60 % of that in TM1 (Rapp et al. 2010). Comparative expression profiling using early microarrays demonstrated that the transcriptomes of these examples of wild and domesticated *G. hirsutum* were radically altered by domestication and crop improvement, with nearly a quarter of genes in the genome being differentially expressed (Rapp et al. 2010). A more recent study that included multiple accessions representing wild and domesticated *G. hirsutum* using RNA-seq showed that nearly 5,000 genes have been differentially regulated as a consequence of cotton fiber domestication (Yoo and Wendel 2014). These data suggest that the strong directional selection accompanying domestication of cotton fibers has entailed a complex alteration of the transcriptional developmental network. A number of different metabolic pathways have been implicated in this complex evolutionary process (Chaudhary et al. 2008, 2009; Hovav et al. 2008; Rapp et al. 2010; Bao et al. 2011; Yoo and Wendel 2014).

In addition to genetic and transcriptomic analyses, proteomic investigation of fiber development has proven to be a powerful and complementary approach for the study of cotton domestication (Hu et al. 2013). Between wild and domesticated forms of *G. barbadense*, 190 of 1,317 fiber proteins sampled were differentially expressed at one or more stages of fiber development, whereas few of these expression changes were suggested at the transcriptional level. It is clear that mRNA expression levels only partially predict protein abundances, as well as the cellular and physiological activities mainly executed by proteins. Although a two-dimensional gel electrophoresis (2-DE)-based approach has been established and used to identify more than 200 proteins that change during fiber development in upland cotton (Yao et al. 2006; Li et al. 2007; Yang et al. 2008; Pang et al. 2010; Zhao et al. 2010; Zhang et al. 2013), we are aware of no reports of proteomic transformations accompanying the domestication of *G. hirsutum*.

Here, we analyzed the fiber proteomes of wild and domesticated *G. hirsutum* at four developmental time points using two independent proteomic approaches, the classic two-dimensional gel electrophoresis (2-DE) that was widely used in cotton proteomic studies and the more recently developed isobaric tag for relative and absolute quantification (iTRAQ) technology coupled with LC–MS/

MS. Using each method, we characterized the global expression patterns in both wild and domesticated *G. hirsutum*, and documented proteins expressed during fiber development. We identified differentially expressed proteins as candidates for functional analyses that may yield insight into domestication and future cotton improvement, and diagnosed homoeolog-specific differences between duplicated gene products. Finally, we compared the results produced by each proteomic strategy and the reproducibility of protein profiling by these methods, a subject of current interest in proteomics research (Aggarwal et al. 2006; Chong et al. 2006; Wu et al. 2006; Fenselau 2007; Thelen and Peck 2007; Vercauteren et al. 2007).

Materials and methods

Plant materials, tissue collection and protein extraction

To represent modern upland cotton, we used the genetic and cytogenetic standard, Texas Marker Stock 1 (TM-1). For wild *G. hirsutum*, we chose race *yucatanense*, accession Tx2094 (USDA GRIN PI 501501), an unambiguously wild form (Wendel et al. 2010) that has been used in most molecular analyses of wild *G. hirsutum*. Plants from each accession were grown in the Bessey Hall Greenhouse at Iowa State University. Flowers were tagged at anthesis and developing bolls were harvested at 5, 10, 20 and 25 days post-anthesis (dpa), representing the key developmental stages of primary wall synthesis and elongation (5 and 10 dpa), and the transition to (20 dpa) and continuation (25 dpa) of secondary wall synthesis. Bolls were dissected immediately after harvest and ovules were frozen in liquid nitrogen and stored at −80 °C. Fiber proteins were extracted and purified as previously reported (Hu et al. 2013). Three biological replicates per developmental stage and per cotton accession were prepared. For each replicate, ovules were harvested and pooled from five plants, and subjected to protein extraction.

2-DE and image analysis

Extracted proteins were dissolved in isoelectric focusing (IEF) buffer [8 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 2 % (v/v) Triton X-100, and 50 mM DTT] at room temperature, and quantified using the 2-D Quant Kit (GE Healthcare Biosciences, Pittsburgh, PA, USA) with bovine serum albumin as standard. For each 1 mg of fiber protein, the final volume was adjusted to 450 µl with IEF buffer and then 2.25 µl of IPG buffer 3–10 NL (GE Healthcare Biosciences) was added. After centrifugation at 10,000*g* for 10 min to remove insoluble materials, the supernatants were loaded on 24-cm 3–10 NL immobilized pH gradient (IPG) strips (GE Healthcare Biosciences) for rehydration. As previously described (Hu et al. 2011), first dimensional isoelectric focusing (IEF) and the secondary SDS-PAGE were performed for protein separation with finished gels stained using a colloidal Coomassie Brilliant Blue G-250 solution.

Finished 2-DE gels were scanned using ImageScanner (GE Healthcare Biosciences) at a resolution of 600 dpi and 16-bit grayscale pixel depth. Image and statistical analysis was conducted with the Progenesis SameSpots software version 4.0 (Non Linear Dynamics, Durham NC, USA), using standard procedures and default parameters. Briefly, spot detection, background subtraction, and normalization were performed following automatic alignment of all images with manual inspection. Detected spots were 100 % matched across images, so that all gels contain the same number of spots without missing values. Differential expression of spot volumes that represent protein abundances was assessed using the built-in statistical tool in the SameSpots software for contrasts of interest. Our criteria for significant changes were based on an ANOVA $P < 0.05$ as calculated with a fold change cutoff of >1.2 or <0.8 (same as applied in iTRAQ analysis below). By matching our 2-DE gels against a reference spot map of upland cotton fibers (Zhang et al. 2013), protein identification was retrieved for corresponding spots based on a cross-database approach of gel comparison (Lemkin et al. 2005).

iTRAQ LC–MS/MS, protein identification and quantification

Extracted protein were dissolved in protein buffer [8 M urea, 25 mM triethylammonium bicarbonate (TEAB), 0.2 % (v/v) TX-100, 0.1 % SDS (w/v), pH 8.5], and prepared for iTRAQ labeling as described in Hu et al. (2013). For each sample, 100 µg of protein was reduced, alkylated, and trypsin-digested using the iTRAQ Reagents 8-plex Kit according to the manufacturer's instructions (AB Sciex, Inc., Foster City, CA, USA). The TM1 proteins were labeled with iTRAQ tags 113 (5 dpa), 114 (10 dpa), 115 (20 dpa) and 116 (25 dpa), and Tx2094 proteins were labeled with tags 117 (5 dpa), 118 (10 dpa), 119 (20 dpa) and 121 (25 dpa). After combining labeled samples, the peptide mixture was fractionated with strong cation exchange chromatography and analyzed using a quadrupole time-of-flight QSTAR Elite MS/MS system (AB Sciex Inc.), and data were acquired as described previously (Koh et al. 2012). Reversed-phase C_{18} chromatographic separation of peptides was carried out on a pre-packed BetaBasic C_{18} PicoFrit column (75 µm id × 10 cm length, New Objective, Woburn, MA) at 300 nl/min using the following gradient: 5 % B for 1 min as an equilibration status; 60 %

B for 99 min as a gradient; 90 % B for 5 min as a washing status; 5 % B for 10 min as an equilibration status (solvent A: 0.1 % formic acid in 97 % water, 3 % ACN; solvent B: 0.1 % formic acid in 97 % ACN, 3 % water).

The MS/MS data were processed by a thorough database search considering biological modifications and amino acid substitutions under the Paragon^{TM} algorithm (Shilov et al. 2007) and the Pro Group^{TM} algorithm, using ProteinPilot version 4.5 software (AB Sciex, Inc.). As described in Hu et al. (2013), three protein databases were used for iTRAQ protein and homoeolog identification, including a non-redundant *Gossypium* protein database and two separate A- and D-genome diploid databases. Methylthio-cysteine and amine groups at the N-terminus and lysine were considered for the fixed modifications and variable modifications were included for post-translational modifications (PTMs). The cutoff of protein identification was set to a confidence level of 95 %. The global false discovery rate (FDR) of identified protein lists was determined by performing searches against the reversed protein databases, with estimates derived from both the conventional approach and a non-linear fitting method (Tang et al. 2008) as shown in Supporting Information Table S1. The identified proteins were assigned to protein families using PANTHER (Mi et al. 2010) and functional classes based on the *Arabidopsis* functional catalog (Bevan et al. 1998).

For protein quantification, only MS/MS spectra that were uniquely identified for a particular protein were used to extract peak intensities of iTRAQ labeling tags, which were subsequently normalized across samples using the built-in bias correction function of ProteinPilot (AB Sciex, Inc.). Relative protein quantification for comparisons of interest was generated with *P* values using the software's standard procedures. To be identified as being significantly differentially expressed, a protein must have been quantified with a fold change of >1.2 or <0.8 and $P < 0.05$ in at least two of the biological triplicates, along with a Fisher's combined probability of <0.05 (Fisher 1948).

Functional enrichment test and other statistical analysis

The agriGO tool was used to perform Single Enrichment Analysis (SEA) for differentially expressed proteins derived from the iTRAQ analysis (Du et al. 2010). Based upon the supported *Gossypium* background data, each set of differentially expressed proteins was compared to the reference list of all iTRAQ-identified proteins. Gene Ontology (GO) terms were enriched with multi-test adjustment of the Benjamini–Yekutieli method (FDR <0.05) (Benjamini and Yekutieli 2001). Hierarchical clustering with regular bootstrap (BP) and approximately unbiased bootstrap (AU) *P* values were performed using the R software package pvclust (Suzuki and Shimodaira 2006), specifying average linkage and Pearson's correlation distance metric with 10,000 iterations. The bootstrap *P* values indicated how strongly each cluster is supported by the data. For example, for a cluster with $P > 95$ %, the hypothesis that "the cluster does not exist" is rejected with a significance level of 5 %. The branch length represents the degree of dissimilarity in protein expression among conditions.

Results

Parallel proteomic strategies applied to developing cotton fibers

To compare the fiber proteomes between wild and domesticated *G. hirsutum*, we chose race *yucatanense*, the most likely wild ancestor of upland cotton (Wendel et al. 2010), for fiber proteomic profiling, and the genetic and cytogenetic standard Texas Marker Stock 1 (TM-1) as a representative of modern cultivars. Proteins prepared from cotton fibers collected at four key stages, i.e., 5 (primary wall synthesis), 10 (fast fiber elongation), 20 and 25 dpa (the onset and continuation of secondary wall synthesis, respectively) were analyzed using two independent proteomic approaches (as illustrated in Fig. S2). Electrophoresis of twenty-four 2-DE gels (Fig. S3) and three eight-plex iTRAQ experiments were performed in parallel to profile fiber proteomes in wild (Tx2094) and domesticated (TM1) cotton accessions. Based on wide-range (pH 3–10) IPG-IEF separation of fiber proteomes, a high-resolution 2-DE map was established containing 907 detected spots reproducibly matched across all gels (Table S2). By matching our spot map to a recently published 2-DE reference dataset for upland cotton, which identified 235 differentially expressed protein spots during fiber development (Zhang et al. 2013), protein identification was accomplished for 182 spots on our map representing electrophoretic isoforms of 147 proteins. Using iTRAQ labeling coupled with LC–MS/MS experiments, a total of 1,199 fiber proteins were characterized (identification confidence >95 %, FDR <1 %; Table S1 and S3), among which 558 proteins were diagnosed with homoeolog-specific peptides (Table S4). Collectively, 2-DE and iTRAQ analyses resulted in the identification of 1,223 non-redundant proteins (i.e., from unique *Gossypium* genes), with 113 proteins identified by both methods (Table S2). With respect to the quantitative expression profiles of these commonly identified proteins, interestingly, only a small number of significant changes overlapped between the 2-DE and iTRAQ datasets. One explanation for this observation is the lack of strict correspondence between the protein features identified by different methods; for

example, one single protein is often separated by 2-DE gels to more than one isoform, while an iTRAQ-identified protein actually represents a group of homologs sharing the detected peptides (discussed in detail later). Proteins displaying consistent expression changes during fiber development and domestication by both methods include a fk506 binding-like protein (Gorai.001G230100), translation elongation factor 2 (Gorai.002G036600), phenylcoumaran benzylic ether reductase (Gorai.002G127000), rhamnose biosynthetic enzyme 1 (Gorai.003G163300), serine hydroxymethyltransferase (Gorai.011G153900), ran-binding protein 1 (Gorai.013G269700), one uncharacterized protein (Gorai.004G235700) and two annexins (Gorai.007G239000; Gorai.009G237900).

Functional categorization of the identified proteins revealed that protein families of transferase, isomerase, chaperone and transfer/carrier protein were over-represented by 2-DE proteins, while nucleic acid binding protein, ligase and membrane traffic protein were more highly represented by proteins identified by iTRAQ (Fig. 1a). With respect to physical properties of the identified fiber proteomes, a wide range of isoelectric points (pIs) and molecular masses were recovered from both 2-DE and iTRAQ datasets (Fig. 1b, c). Proteins within the pI range 4–7 dominated the pI spectrum, accounting for 86.4 % (2-DE) and 64.6 % (iTRAQ) of the total identified proteins from cotton fibers. High pI proteins (above 7) were better represented by iTRAQ, and notably, 52 proteins, with pI above 10, were identified outside the separation range of our 2-DE analysis (Fig. 1b). Neither approach identified proteins with pI values below 4. Proteins within the mass range of 10–70 kDa represented 89.1 % (2-DE) and 85.8 % (iTRAQ) of the identified proteins (Fig. 1c). High mass proteins above 100 kDa were only resolved by iTRAQ, and the protein with highest mass was identified as a 566 kDa auxin transport protein (Gorai.001G206500), far beyond the molecular weight range (10–120 kDa) of 2-DE gels. These results suggested that iTRAQ is more inclusive with respect to proteins having high molecular masses and pIs relative to the 2-DE method, despite the fact that only a

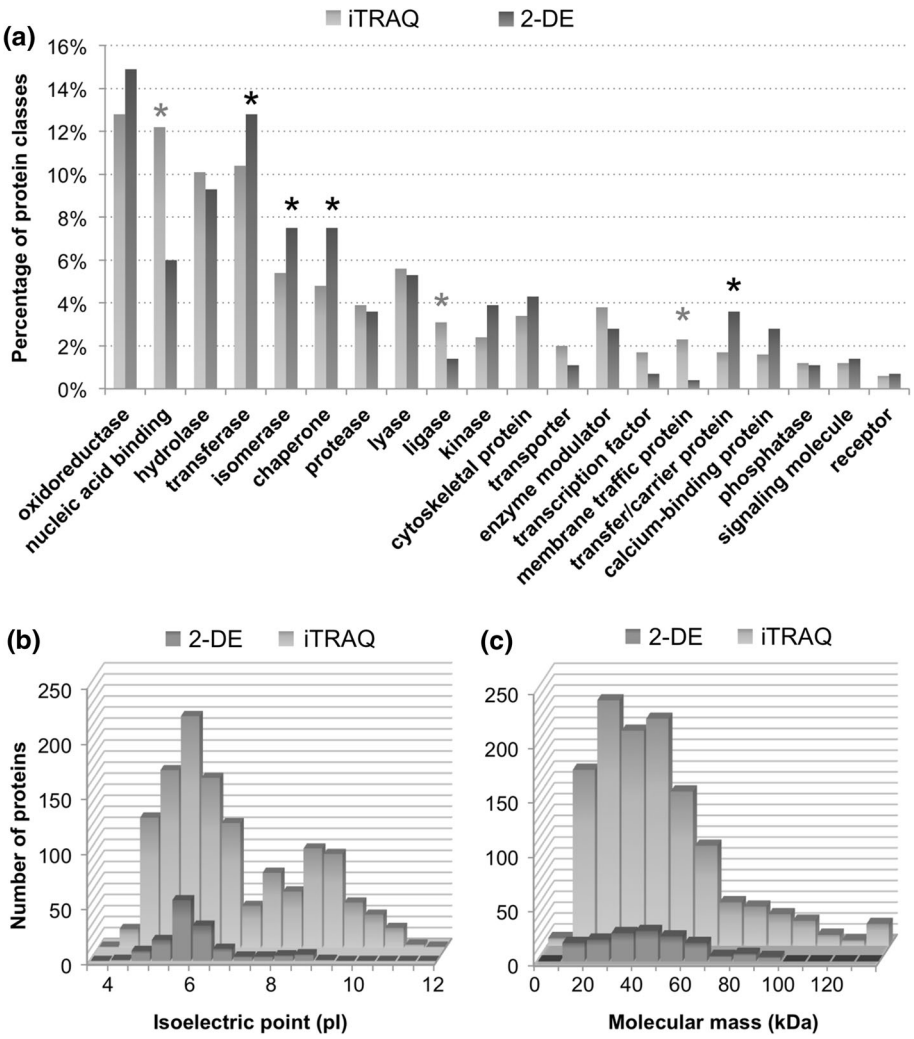

Fig. 1 Classification and biochemical properties of proteins expressed during fiber development. Panther protein functional classes (**a**), theoretical pIs (**b**) and molecular masses (**c**) were compared between proteins identified from 2-DE (*dark columns*) and iTRAQ (*light grey columns*). *Asterisk* significant difference based on binomial tests

small portion of 2-DE spots was identified. Regardless of their different throughput efficiencies and coverage in terms of protein identification, both iTRAQ and 2-DE methods allow quantitative and comparative profiling of global protein features.

Quantitative proteomic changes during fiber development

To study protein expression during fiber development, proteomic changes were examined between adjacent developmental stages (5–10, 10–20, and 20–25 dpa), which revealed that 292 iTRAQ proteins (24.6 % of 1,189 proteins) and 331 2-DE spots (36.4 % of 907 spots; Table S5) were significantly differentially expressed within wild or domesticated cotton fibers. Of these, wild (Tx2094) and domesticated (TM1) cotton displayed nearly identical levels of developmental expression variation (iTRAQ, 202 Tx2094 vs. 203 TM1 proteins; 2-DE, 193 Tx2094 vs. 181 TM1 spots), suggesting that domestication has not substantively changed this feature of protein dynamics during development.

As shown in Fig. 2a (blue arrows), iTRAQ analysis revealed that in Tx2094 the highest number of protein expression changes occurred early during fiber development (5–10 dpa, 156 proteins), twofold higher than changes during later time intervals (10–20 dpa, 45; 20–25 dpa, 55; $P < 0.05$, Fisher's exact test). This pattern differed from that shown by TM1, where large numbers of changes were detected through 20 dpa (5–10 dpa, 121; 10–20 dpa, 132 proteins), with fewer changes observed later (20–25 dpa, 29 proteins; $P < 0.05$, Fisher's exact test). By examining the multivariate clustering of these proteins, the paired time courses of 5–10 and 20–25 dpa were similar for TM1 and Tx2094 (Fig. 2b), suggesting that developmental changes during these periods of fiber elongation and secondary wall synthesis are similar in domesticated and wild cotton. This was not the case for the period of rapid fiber elongation during primary wall synthesis (10–20 dpa). Functional categorization revealed that from 10 to 20 dpa more proteins were differentially regulated in TM1 versus Tx2094 ($P < 0.05$, Fisher's exact test), corresponding to functional activities of "metabolism", "energy", "protein synthesis", "protein destination and storage" and "signal transduction" (Table 1). Although no apparent co-expression patterns were identified as being distinct between wild and domesticated cotton with respect to functional categorization, interestingly, the "protein synthesis" class, including ribosomal protein subunits, translation factors and tRNA synthetases, was predominantly down-regulated from 5 to 20 dpa in both accessions ($P < 0.05$, Fisher's exact test; Table S5). In addition, functional enrichments based on the Singular Enrichment Analysis (SEA) revealed that ribosomal activities and translation were specifically enriched in TM1 (Table S6), also suggesting the functional significance of this biological process accompanying domestication.

Proteomic changes based on 2-DE spot features exhibited similar levels of proteomic variation between early (5–10 dpa) and late (20–25 dpa) developmental stages within each accession (TM1, 69 vs 65 spots; Tx2094, 83 vs 75 spots; Fig. 2c, blue arrows). Consistent with results from the iTRAQ analyses, more proteins were differentially expressed from 10 to 20 dpa in TM1 than in Tx2094 (94 vs 44 spots; $P < 0.05$, Fisher's exact test). In addition, multivariate expression patterns of 2-DE spots also indicated clustering of paired time courses between wild and domesticated cottons for 5–10 and 20–25 dpa ($\sim$80 and 90 % bootstrapping, respectively; Fig. 2d), thereby again suggesting that the most dramatic rewiring of the fiber proteome caused by domestication and crop improvement occurred between 10 and 20 dpa. A total of 82 differentially expressed spots were identified, among which the identified proteins for 69 spots were also detected in iTRAQ analyses (Table S2 and S5). In comparison to the protein functional classes represented by iTRAQ data, no proteins identified by 2-DE analysis were categorized into the classes of "transcription", "intracellular traffic" or "cell structure" (Table 1).

Differential protein expression between wild and domesticated *G. hirsutum*

When comparing protein expression levels between TM1 and Tx2094 at each time point, 240 iTRAQ proteins (20.2 %) and 293 2-DE spots (32.3 %) were differentially expressed at one or more developmental stages (Table S5). As shown in Fig. 2 (brown arrows), the numbers of 2-DE spots displaying expression change were distributed equally among the four developmental time points, while for the proteins profiled by iTRAQ, the highest number of expression changes occurred early in fiber elongation (5 dpa, 137 proteins), followed by fewer changes later during primary wall synthesis (10 dpa, 78 proteins) and the transition to secondary cell wall synthesis (20 dpa, 100 proteins), with the fewest changes during secondary wall synthesis (25 dpa, 48 proteins). The distribution of up-regulation between TM1 and Tx2094 is statistically symmetric, except for 2-DE results where more proteins were up-regulated in TM1 than Tx2094 at 5 dpa, followed by a switch in direction at 10 dpa ($P < 0.05$, Chi-squared test). Functional analysis of these proteins, including all of the proteins from iTRAQ and those identified from 2-DE, indicated that all functional categories are involved in the proteomic differences between TM1 and Tx2094, with more than half of the proteins belonging to the categories

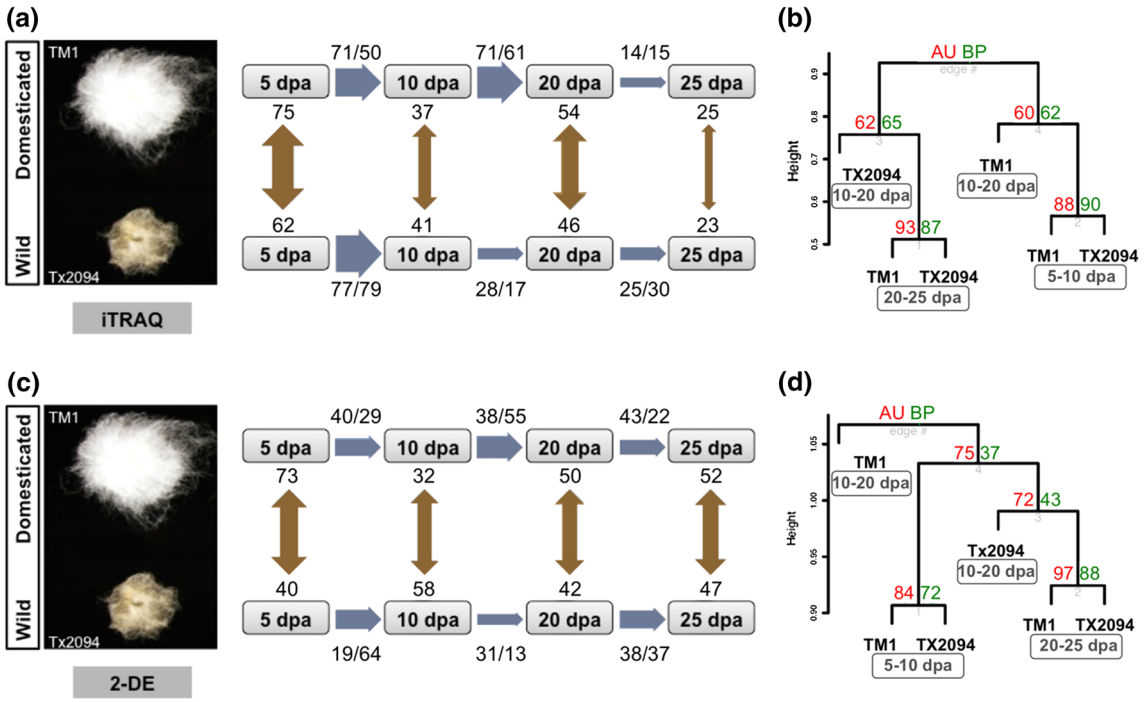

Fig. 2 Number of proteins differentially expressed during fiber development within and between wild and domesticated *Gossypium hirsutum*. Differential expression was independently analyzed for 1,189 iTRAQ proteins (**a**) and 907 2-DE spots (**c**). A representative image of a single seed with attached cotton fibers is shown for each accession. *Numbers by the end of arrows* denote the numbers of proteins up-regulated for the specified comparison, and *numbers by the beginning of arrows* denote numbers of down-regulated proteins. *Arrow thickness* is proportional to the number of differentially expressed proteins in a given comparison. Developmental expression changes within each accession, represented by log2 ratios between adjacent time points, were subjected to hierarchical clustering analysis for iTRAQ (**b**) and 2-DE (**d**) datasets, respectively. A bootstrapping of 10,000 iterations was performed to calculate bootstrap probabilities (BP) and approximately unbiased bootstrap probabilities (AU)

Table 1 Functional categorization of proteins differentially expressed during fiber development

Protein numbers	iTRAQ[a]													2-DE
	5–10 dpa				10–20 dpa				10–20 dpa				Total (%)	Total (%)
	Up		Down		Up		Down		Up		Down			
Functional category	T	X	T	X	T	X	T	X	T	X	T	X		
Metabolism	15	14	5	5	12	3	13	6	4	7	2	5	53 (18.2)	23 (28.0)
Energy	3	6		6	5	1	6		1	2		2	25 (8.6)	8 (9.8)
Cell growth and cytoskeleton	3	6			3	1	1		1	2			10 (3.4)	10 (12.2)
Transcription			1	6	2			2		1			7 (2.4)	
Protein synthesis		3	44	25		1	21	4				1	53 (18.2)	2 (2.4)
Protein destination	4	5	3	9	5	1	7	1	1	4	1	3	30 (10.3)	4 (4.9)
Transporter	3	4			7	2	1				1	2	9 (3.1)	3 (3.7)
Intracellular traffic	1	1									1		2 (0.7)	
Cell structure		2	3	4	3	2	5	3	2	2	1		9 (3.1)	
Signal transduction	7	13		1	8	1				1	1	1	22 (7.5)	7 (8.5)
Defense and stress response	5	8	5	9	7	3	7	4	3	6	1	5	30 (10.3)	7 (8.5)
Secondary metabolism	2	6	3	2	2	2	4	3	2	3	1	4	11 (3.8)	6 (7.3)
Unclassified	7	11	7	10	7		6	5	1	2	5	2	31 (10.6)	12 (14.6)
											SUM		**292**	**82**

[a] Up- and down- regulated protein numbers in the domesticated cotton TM1 (T) and the wild accession Tx2094 (X) and were tabulated for the iTRAQ dataset

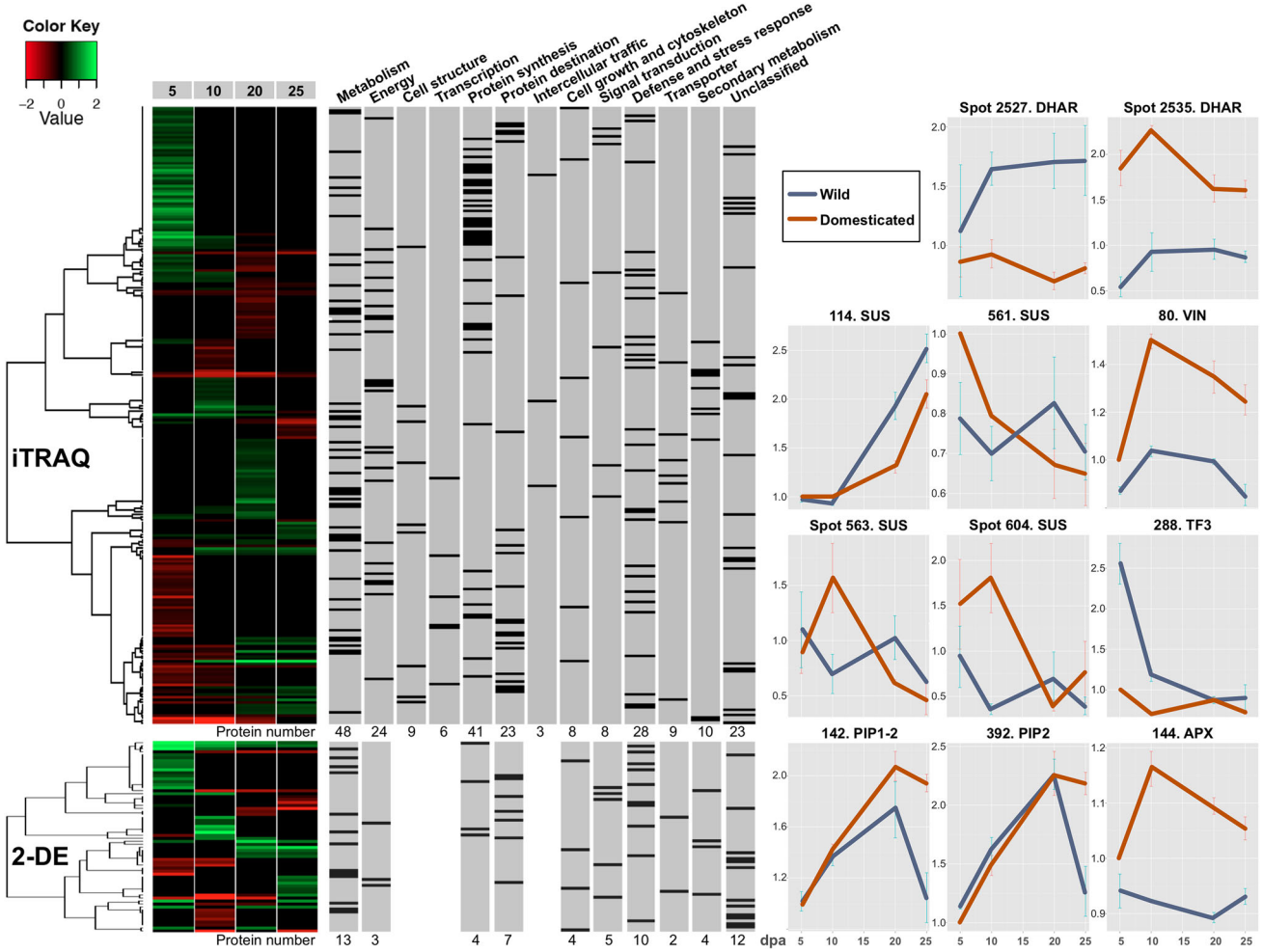

Fig. 3 Differentially expressed proteins in domesticated *G. hirsutum* relative to its wild progenitor. Differential expression patterns of 240 iTRAQ proteins and 64 proteins identified from 2-DE are separately clustered on the vertical axis of the heatmap. Expression ratios of domesticated versus wild are plotted on a $\log_2$ scale for each time point. Up- and down-regulation are shown in *green* and *red colors*, respectively; *black* corresponds to no significant change. Based on Bevan et al. (1998), functional category was assigned to each protein, whose corresponding row is marked *black* in the central *grey* columns. Examples of protein expression profiles are shown on the right. *DHAR* dehydroascorbate reductase, *SUS* sucrose synthase, *VIN* vacuolar invertase, *PIP* plasma membrane intrinsic protein, *APX* ascorbate peroxidase, chloroplastic

of "metabolism", "energy", "protein synthesis", "protein destination", and "defense and stress response" (Fig. 3). Co-expression associated with categorized cellular function was observed for the class of "protein synthesis", where most proteins were up-regulated by domestication in 5 dpa fibers; this observation was confirmed by the SEA enrichment results (Table S6). In addition, 5 of 7 transcription-related proteins in the "unclassified" class were down-regulated in domesticated cotton, including a transcription factor (Gorai.004G264900, expression profile shown in Fig. 3 panel "288. TF3"), a transcription coactivator (Gorai.010G214300) and three RNA binding proteins (Gorai.010G102100, Gorai.004G004300, Gorai.005G052700).

Homoeolog expression in allopolyploid *G. hirsutum*

In addition to studying total protein accumulation from both homoeologs (A_T and D_T, where the subscript indicates the specific genome in the allopolyploid) of allopolyploid cotton (which contains A_T and D_T genomes), we also utilized iTRAQ and 2-DE approaches in an effort to separately study the protein products of individual homoeologous copies of duplicated genes. To accomplish this, protein identification with evidence of genome-of-origin is necessary, an opportunity afforded by MS-based proteomics such as iTRAQ analysis which is capable of diagnosing homoeolog-specific peptides, as shown by Hu et al. (2013).

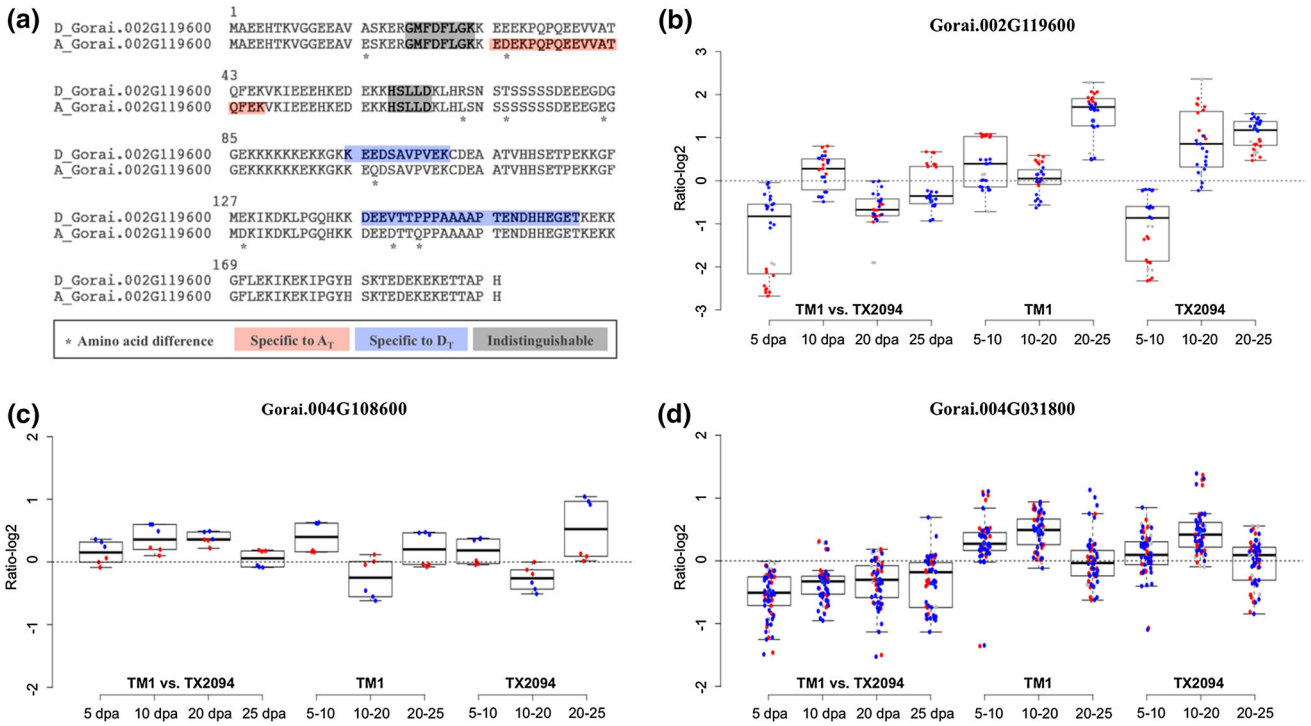

Fig. 4 Homoeolog-specific protein expression. Sequence alignment for a pair of homoeologous proteins is displayed for Gorai.002G119600, marked with their amino acid differences and identified peptides (**a**). *Boxplots* show expression ratios for the homoeologous protein pair of Gorai.002G119600 (**b**), Gorai.004G108600 (**c**), and Gorai.004G031800 (**d**). Individual measurements plotted as points are labeled for homoeolog-specific peptides (*red* A_T, *blue* D_T) and peptides shared by both homologs (*grey*). Expression ratios are calculated on log2 scale for comparisons of TM1 vs. Tx2094 at each developmental time point (i.e., 5, 10, 20 and 25 dpa), and between adjacent time points within both accessions (e.g., 5–10 denotes 10 vs. 5 dpa)

A total of 558 proteins that had genome-diagnostic peptides were identified by iTRAQ. Of these, 137 proteins had peptides diagnostic for both homoeologs (Table S4). Different from the expression profile of total protein accumulation, homoeolog-specific expression was quantified based on only the corresponding A_T- or D_T- specific peptides (as opposed to using all peptides). For example, MS spectra identified for a dehydrin protein (Gorai.002G119600), known to play a role in response to cold and drought stresses (Hanin et al. 2011), were mapped to five distinct peptides (one to multiple spectra can be mapped to one peptide), among which two were diagnosed as specific to D_T, one as specific to A_T, and the other two were shared by both homoeologs (Fig. 4a). As shown in Fig. 4b, measurements specific to A_T and D_T homoeologs appeared to distribute separately in most comparisons; at 25 dpa, for example, log2 ratios of the D_T homoeolog (blue spots) were all below zero, suggesting down-regulation of the protein expression in TM1 compared to Tx2094, whereas all values of the A_T homoeolog (red) indicated up-regulation. Similar distinctions between homoeologs were also found for Beta-hydroxyisobutyryl-CoA hydrolase 1 isoform 2 (Gorai.004G108600), where only its D_T homoeolog appeared to be differentially regulated during domestication and fiber development ($P < 0.05$, Student's t test; Fig. 4c). These results provide evidence of differential expression and variable regulation between protein homoeologs in allopolyploid *G. hirsutum*. Using this comparative framework, homoeolog-specific expression was also inspected for all proteins for which the appropriate data were recovered; this analysis showed that in general, A_T and D_T homoeologs display indistinguishable expression patterns. For example, for aspartic proteinase A1 (Gorai.004G031800), the measurements specific to each homoeolog are intermingled, indicating equivalence in expression (Fig. 4d).

Among the remaining proteins with diagnostic peptides detected for only one of the two homoeologs, 230 and 191 proteins had peptides identified that were specific to A_T and D_T homoeologs, respectively. As suggested and demonstrated in Hu et al. (2013), the solo detection of one of the two homoeologs reflects the non-random nature of MS in detecting high-abundance peptides from the more highly expressed copy, which, therefore, can be used as an indicator of biased protein expression favoring the homoeolog detected. Using this logic, the number of A_T and D_T homoeolog expression biases are statistically equivalent ($P > 0.05$, Chi-squared test), suggesting that although

homoeolog expression bias of fiber proteins appears to be pervasive (421 of 558 proteins), their directions are balanced with respect to genome-of-origin in *G. hirsutum*. To increase the statistical power of this analysis, we applied quantitative threshold (N) to indicate solo detection of either homolog; that is, the biased expression of one homoeolog was considered only when the identified number of homoeolog-specific spectra $\geq N$. By applying the N value from 1 to 101 until no biased expression was allowed, the balanced detection of A_T and D_T expression biases was well maintained.

Discussion

Complementary proteomic approaches applied to cotton domestication

Since the time that "Variation under Domestication" was published (Darwin 1859), comparative analysis of cultivated plants and their wild relatives has been recognized as a useful approach to generate insights into the evolutionary consequences of diversifying and directional selection. Over the past 20 years, much progress has been made to identify the causal mutations and underlying genetic architecture that control morphological transformations during crop domestication and improvement (Doebley et al. 2006; Burke et al. 2007, 2008; Burger et al. 2008; Gross and Olsen 2010; Olsen and Wendel 2013a, b). As reviewed in Olsen and Wendel (2013a, b), the increasing application of genome-scale systems biology approaches (e.g., genomics, transcriptomics, proteomics, and metabolomics) promises to shed qualitatively new light on crop plant evolution. Here we demonstrate the use of advanced proteomic profiling tools in an elite cotton cultivar and a wild accession to gain insight into cotton fiber development and evolution for upland cotton, *G. hirsutum*. We characterized the developmental dynamics of fiber proteomes and conducted intraspecific comparative analyses using 2-DE and iTRAQ approaches, powerful and complementary techniques to study complex protein profiles (Rose et al. 2004; Chen and Harmon 2006; Thelen and Peck 2007).

Following the advent of methodology for the extraction of proteins from cotton fibers (Yao et al. 2006), a number of proteomic studies have been performed using the 2-DE method (Yao et al. 2006; Li et al. 2007; Yang et al. 2008; Pang et al. 2010; Zhao et al. 2010; Zhang et al. 2013), which separates proteins into complex maps of spots by native charge (pI) and molecular mass, followed by characterization of their molecular activities during fiber growth. Approximately 200 proteins differentially expressed during fiber development typically are identified using this approach (Yang et al. 2008; Zhang et al. 2013), as recently reviewed (Zhou et al. 2014). In comparison to 2-DE based studies, our previous iTRAQ analysis characterized 1,317 cotton fiber proteins in *G. barbadense* (Hu et al. 2013), a much higher proteome coverage due to the nature of high-throughput protein identification of MS-based proteomics (Thelen and Peck 2007; Diz et al. 2012). In this study, our iTRAQ results not only accounted for 70 % of the characterized 2-DE proteomes, but also provided a more comprehensive catalog of fiber proteins with a wider range of pI and molecular mass (Fig. 1b, c). These data serve as a valuable resource for studying the molecular mechanisms of fiber elongation and cell wall synthesis.

Despite the limitations of 2-DE in terms of low proteome coverage and data throughput, the ability of this methodology to resolve protein isoforms as different spots is a unique advantage relative to other technologies, in that it provides additional information on post-translational modifications and other protein-level regulatory mechanisms. For example, differentially modified forms of a dehydroascorbate reductase (DHAR; Gorai.012G068600) were detected as separate spots on 2-DE gels, which displayed opposite expression patterns in response to domestication (Fig. 3 panels "Spot 2527" and "Spot 2535"). In iTRAQ analysis, the same protein was uniquely identified, but its expression changes were found to be mostly intermediate between those of 2-DE isoforms (Table S5). This discrepancy in protein characterization can be explained by the peptide-centric nature of iTRAQ analysis: between different protein isoforms, only a few peptides bear distinct point modifications or proteolytic sites, which often are under-represented due to their relatively low abundance, while more abundant and common peptides are more likely to be identified and grouped together as a single identification. Thus, various protein isoforms are less likely to be studied by iTRAQ compared to 2-DE. Overall, the different strengths of 2-DE and iTRAQ underscore the importance of technical diversity in revealing the complexity of biological system and evolutionary studies at protein level.

Fiber proteomic changes accompanying domestication

Over thousands of years of human-mediated selection and agronomic improvement, wild *G. hirsutum* underwent numerous phenotypic modifications to become a modern crop with dramatically enhanced fiber length and fineness (Fig. S1). Transcriptomic analyses of wild and domesticated cotton highlight the fact that the molecular networks of fiber development are highly altered by domestication and crop improvement (Chaudhary et al. 2008; Hovav et al. 2008; Rapp et al. 2010; Yoo and Wendel 2014). Here we extend these initial insights to the proteomic level, thereby providing new perspectives on the molecular basis of

evolutionary transformations and cotton biology. The level of proteomic change (20.2–32.3 %) we observed between wild and domesticated *G. hirsutum* is in agreement with that previously reported at the transcriptome level (23.4 % of 40,430 genes) using the same model accessions (Rapp et al. 2010). In that previous study, the greatest amount of differential gene expression was found early in fiber development (2 dpa), prior to obvious macroscopic differences in fiber morphology. It was suggested that the gene regulatory network was rewired by early developmental events, with these changes propagating through the transcriptional network during subsequent developmental periods. At the protein level, the highest amount of expression change was also observed at the earliest stage sampled in our study (5 dpa).

One caveat to our study is that only one accession each of the wild and domesticated *G. hirsutum* was used for the proteomic comparison. Therefore, some of the observed expression changes might be specific to the choice of accessions, rather than representing the general domestication process. Future studies would benefit from including multiple accessions of wild and domesticated cotton for profiling fiber proteomes, such as recently applied to a transcriptomic survey of cotton domestication (Yoo and Wendel 2014). However, *G. hirsutum* exhibits an extremely narrow gene pool, for both the modern crop cultivars and truly wild races (Wendel et al. 1992; Brubaker and Wendel 1994), which suggests that the overall patterns we observed likely are representative of the proteomic alterations accompanying upland cotton domestication.

In addition to providing a quantitative overview of the complexity of the proteomic changes during development and as a consequence of domestication, the data also implicate key processes altered by these phenomena. Among the differentially expressed fiber proteins, about 30 % function as catalytic enzymes in biochemical pathways of nucleotide, amino acid, fatty acid, lipid, and carbohydrate (Table 1; Fig. 3), offering clues into the regulation of the basal metabolic reactions that sustain fiber cell growth. In particular, carbohydrate metabolism is known to be critical for both energy production and for providing intermediates for cell wall synthesis (Stiff and Haigler 2012). An important enzyme in this process is sucrose synthase (SUS), which catalyzes a reversible reaction but preferentially converts sucrose into fructose and UDP glucose. Two differentially expressed SUS proteins were identified by iTRAQ, displaying distinct expression profiles. As shown in Fig. 3 (right panels), SUS protein 114 (Gorai.010G091800) was up-regulated from 10 dpa and displayed a higher expression level in the wild accession at 20 and 25 dpa, while protein 561 (Gorai.010G092300) was highly expressed at 5 dpa and decreased later in TM1, in contrast to having static expression in Tx2094. According to the 2-DE profile, another SUS protein was represented by two spot isoforms, with up-regulation in domesticated cotton at 10 dpa (Gorai.009G038000; Fig. 3 panels "Spot 563" and "Spot 604"). Consistent with this observation of divergent regulation, it has been shown that SUS gene family members are actively involved in a wide range of tissues and developmental processes in cotton (Chen et al. 2012), and play important roles in fiber initiation (Liu et al. 2012) and cell expansion for both primary (Ruan et al. 2005) and secondary cell wall synthesis (Amor et al. 1995; Salnikov et al. 2003). However, it is not clear how a variety of SUS enzymes are temporally differentiated while function collectively in developing fibers. We speculate that future exploration of these SUS homologs could provide clues to important regulatory machinery governing fiber growth.

The other major route of sucrose degradation in plants is through the reaction of invertase (also known as β-fructofuranosidase), which irreversibly hydrolyzes sucrose to fructose and glucose. In developing cotton fibers, a vacuolar invertase (VIN1) was reported to promote fiber elongation, possibly through osmotic regulation mediated by the accumulation of hexoses as osmotically active solutes promoting water influx into the vacuole for turgor-driven cell expansion (Wang et al. 2010). Here, VIN1 (Gorai.008G216800; Fig. 3 panel "80. VIN") was up-regulated following fiber initiation and remained high through primary cell elongation in both cotton accessions; notably, domesticated cotton exhibited higher levels of VIN1 expression, especially during the period of fiber elongation. In response to the elevated turgor pressure, water channel proteins (aquaporins) are known to play important roles in cotton fibers by supporting rapid water influx into vacuoles (Liu et al. 2008). Increased protein abundance of two plasma membrane intrinsic aquaporins (Gorai.004G212800; Gorai.002G002500) was correlated with fast fiber growth in our data, and maintained at a higher level at later stages in the domesticated cotton (Fig. 3 panels "142. PIP1-2" and "392. PIP2"); these data temporally parallel the prolonged period of fiber elongation in domesticated relative to wild cotton (Applequist et al. 2001). Together, these results suggest that the domestication process may have altered regulatory networks that modulate invertase activity and osmotic regulation for enhanced and/or prolonged fiber elongation.

Parallel domestication of cotton but variable evolutionary trajectories

As an extension to our previous analysis of crop domestication in Pima cotton *G. barbadense* (Hu et al. 2013), the current proteomic data for *G. hirsutum* provides a comparative data set for the elucidation of key regulatory

changes that were either exclusively or commonly targeted by human-mediated selection during these two independent domestication events, in addition to providing information that may be relevant to the fiber qualities that are unique to each species. Taking ribosomal proteins as an example, domesticated representatives from both species exhibit peak expression of a variety of ribosomal proteins during fiber initiation, in contrast to the lower expression levels observed in their wild progenitors. What makes this case more interesting is that different original states were characterized between two wild accessions, seen as down-regulation from 5 to 10 dpa in *G. hirsutum* versus up-regulation in *G. barbadense*. It appears that the diversified developmental genetics of ribosomal proteins in different wild cotton species was later reshaped into congruent expression by convergent evolution under strong artificial selection for advanced fiber traits.

With respect to proteomic changes accompanying fiber development, the most dramatic transformation in *G. hirsutum* fibers is the period from 10 to 20 dpa, where over twice as many expression changes were found in domesticated than in the wild cotton. In the other major cultivated cotton species, *G. barbadense*, regulatory rewiring also appears to have altered the period of primary wall synthesis, but differs in that the proteomic transition occurs earlier in the domesticated cotton versus its primitive progenitor while maintaining a similar amount of expression change (Hu et al. 2013). Therefore, these results present quite different evolutionary trajectories in two independent events of cotton domestication. One potential hypothesis to explain this divergence, given that fiber development proceeds with relatively similar timelines between the domesticated phenotypes, is that the two wild species exhibit different developmental timing of fiber growth. Without direct evidence supporting this hypothesis, an alternative but not exclusive explanation is that regulatory changes selected for improved fiber traits, such as ROS scavenging (Chaudhary et al. 2009), regulation of the profilin gene family (Bao et al. 2011) and cytoskeleton dynamics (Hu et al. 2013), are intermingled with those

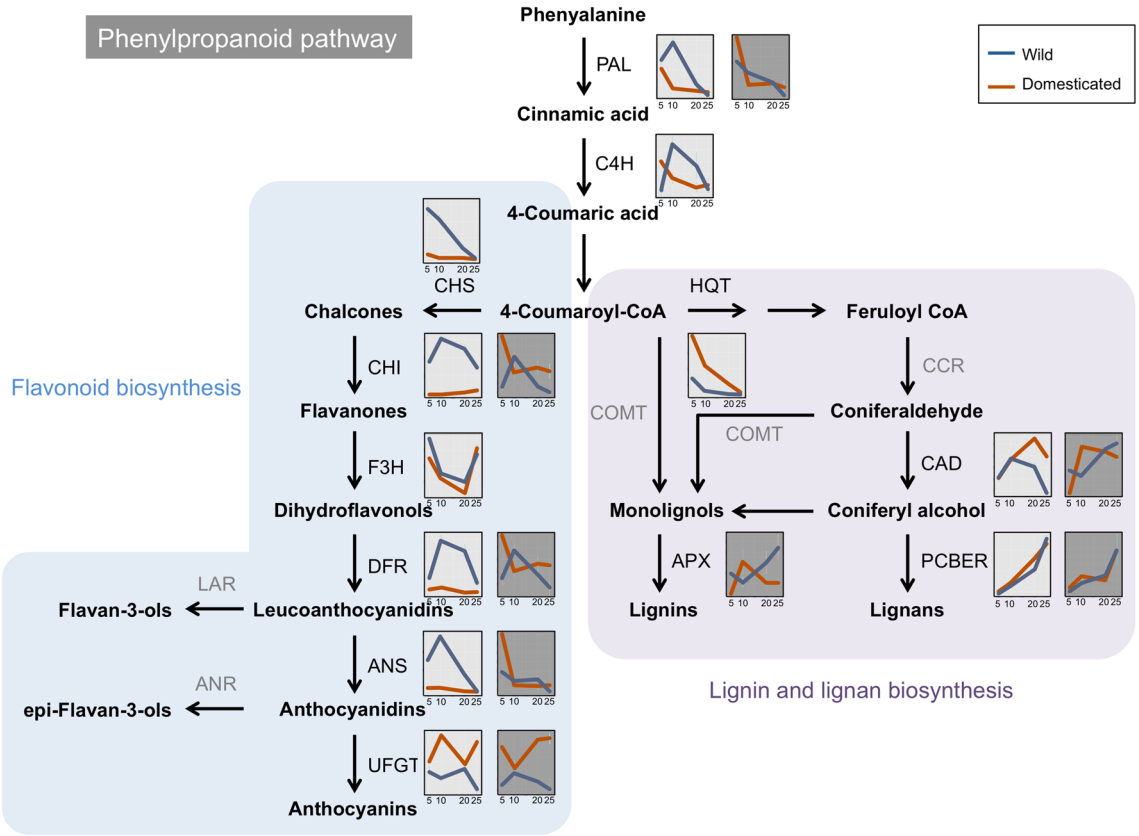

Fig. 5 Expression of phenylpropanoid pathway proteins in wild and domesticated cottons. Expression profiles are shown next to corresponding enzymes. Relative expression levels on a log2 scale at 5, 10, 20 and 25 dpa are presented for wild and domesticated accessions in *G. hirsutum* with *light grey panels* and *G. barbadense* with *dark grey panels*. *ANS* anthocyanidin synthase, *APX* ascorbate peroxidase, cytosolic, *C4H* cinnamate 4-hydroxylase, *CAD* cinnamyl-alcohol dehydrogenase, *CCR* cinnamyl CoA reductase, *CHI* chalcone isomerase, *CHS* chalcone synthase, *COMT* caffeic acid *O*-methyltransferase, *DFR* dihydroflavonol 4-reductase, *HQT* hydroxycinnamoyl CoA quinate transferase, *PAL*, L-phenylalanine ammonia lyase, *PCBER* phenylcoumaran benzylic ether reductases, UFGT, UDP glucose: flavonoid-3-*O*-glucosyltransferase

relevant to fiber physical properties unique to each cultivated species, as exemplified below.

In a comparative transcriptome analysis between *G. hirsutum* and *G. barbadense* cultivars, Al-Ghazi et al. (2009) demonstrated several pathways that may play important roles in determining fiber quality, among which phenylpropanoid biosynthesis leading to production of a wide range of secondary metabolic compounds responsible for developmental signaling, cell wall structure, pigments, plant protection and stress response, was highlighted. The difference in gene expression and metabolic contents seen between *G. hirsutum* and *G. barbadense* could directly impact cell wall structure or affect auxin signaling pathways involved in fiber development and lead to the various physical and chemical properties of different cotton fibers (Al-Ghazi et al. 2009; Fan et al. 2009; Tan et al. 2013). Interestingly, differential expression of proteins from the phenylpropanoid pathway was observed in both, parallel cotton domestication events, as shown in Fig. 5. In *G. hirsutum* (light grey panels), protein expression of major phenylpropanoid and flavonoid pathway enzymes was coordinately down-regulated under domestication. These observations are consistent with previously reported results at the transcriptome level, and support the hypothesis that the domestication process may have reallocated resources from secondary metabolic pathways to more fundamental processes for fiber growth, including carbohydrate and fatty acid synthesis (Yoo and Wendel 2014). However, in *G. barbadense* (dark grey panels), the flavonoid pathway was up-regulated during early fiber elongation and during the transition to secondary wall synthesis (Hu et al. 2013). In addition to flavonoid and anthocyanin biosynthesis, the sub-branch of the phenylpropanoid pathway leading to lignin synthesis appears to be up-regulated following both domestication events, but at different developmental stages. Unlike many other plant cells, cotton fibers are long known to contain little to no lignin in their secondary cell walls (Fan et al. 2009). By examining the molecular and biochemical activities of cinnamyl alcohol dehydrogenase (CAD)—a key enzyme catalyzing the final step of monolignol biosynthesis, Fan et al. (2009) pointed out that the previously overlooked monolignol and lignin compounds are bound to cell walls and potentially of central importance to wall mechanical strength and cotton fiber growth; this protein is inferred to have been differentially affected by domestication in *G. hirsutum* and *G. barbadense*.

Detection and analysis of duplicated gene products

We characterized expression differences between wild and domesticated cotton at the homoeolog-specific level, as in our previous work (Hu et al. 2013). Although there were few cases where A_T and D_T homoeologs displayed strikingly different expression patterns, our data reveal evidence of unequal contributions of homoeologous proteins to the total protein pool, and how this total may obscure subtleties at the homoeolog level.

In the context of crop domestication, a wealth of indirect evidence suggests a role for polyploidization in generating adaptive plasticity and novel phenotypic variation for domestication-related traits (reviewed by (Paterson 2005; Udall and Wendel 2006). Given that all crop plants are polyploids, one promising application of evolutionary proteomics described here is to document the modification pattern of homoeolog-specific protein expression, which may lead to new evidence that links gene duplication to functional and phenotypic significance.

Acknowledgments We thank Kara Grupp and Anna Tuchin for help in tissue collection. We acknowledge the Protein and Proteomics Facility of the Iowa State University for technical assistance in 2-DE analysis. We acknowledge the Proteomics and Mass Spectrometry Facility of the University of Florida's Interdisciplinary Center for Biotechnology Research for assistance in LC–MS/MS analysis. DP was supported by a training grant from Punjab Agricultural University, Ludhiana, India. The LC–MS/MS system was funded by National Institute of Health grant 1S10RR025418-01 to SC. This work was funded by Cotton Incorporated grant 09-558 and by the NSF Plant Genome Research Program, both to JFW.

References

Aggarwal K, Choe LH, Lee KH (2006) Shotgun proteomics using the iTRAQ isobaric tags. Brief Funct Genomic Proteomic 5:112–120

Al-Ghazi Y, Bourot S, Arioli T, Dennis ES, Llewellyn DJ (2009) Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. Plant Cell Physiol 50:1364–1381

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. Proc Natl Acad Sci USA 92:9353–9357

Applequist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. Evol Dev 3:3–17

Bao Y, Hu G, Flagel LE, Salmon A, Bezanilla M, Paterson AH, Wang Z, Wendel JF (2011) Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). Proc Natl Acad Sci USA 108:21152–21157

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188

Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W et al (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391:485–488

Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am J Bot 81:1309–1326

Burger JC, Chapman MA, Burke JM (2008) Molecular insights into the evolution of crop plants. Am J Bot 95:113–122

Burke JM, Burger JC, Chapman MA (2007) Crop evolution: from genetics to genomics. Curr Opin Genet Dev 17:525–532

Chaudhary B, Hovav R, Rapp R, Verma N, Udall JA, Wendel JF (2008) Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. Evol Dev 10:567–582

Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF (2009) Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). BMC Genom 10:378

Chen S, Harmon AC (2006) Advances in plant proteomics. Proteomics 6:5504–5516

Chen A, He S, Li F, Li Z, Ding M, Liu Q, Rong J (2012) Analyses of the sucrose synthase gene family in cotton: structure, phylogeny and expression patterns. BMC Plant Biol 12:85

Chong PK, Gan CS, Pham TK, Wright PC (2006) Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: implication of multiple injections. J Proteome Res 5:1232–1240

Darwin C (1859) On the origin of the species by means of natural selection: or, the preservation of favoured races in the struggle for life. John Murray, London

Diz AP, Martinez-Fernandez M, Rolan-Alvarez E (2012) Proteomics in evolutionary ecology: linking the genotype with the phenotype. Mol Ecol 21:1060–1080

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38:W64–W70

Fan L, Shi WJ, Hu WR, Hao XY, Wang DM, Yuan H, Yan HY (2009) Molecular and biochemical evidence for phenylpropanoid synthesis and presence of wall-linked phenolics in cotton fibers. J Integr Plant Biol 51:626–637

Fenselau C (2007) A review of quantitative methods for proteomic studies. J Chromatogr B Analyt Technol Biomed Life Sci 855:14–20

Fisher RA (1948) Questions and answers #14. Am Stat 2:30–31

Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. Trends Plant Sci 15:529–537

Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci 3:104

Hanin M, Brini F, Ebel C, Toda Y, Takeda S, Masmoudi K (2011) Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. Plant Signal Behav 6:1503–1509

Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu G, Wendel JF (2008) The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet 4:e25

Hu G, Houston NL, Pathak D, Schmidt L, Thelen JJ, Wendel JF (2011) Genomically biased accumulation of seed storage proteins in allopolyploid cotton. Genetics 189:1103–1115

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

Koh J, Chen S, Zhu N, Yu F, Soltis PS, Soltis DE (2012) Comparative proteomics of the recently and recurrently formed natural allopolyploid *Tragopogon mirus* (Asteraceae) and its parents. New Phytol 196:292–305

Lemkin PF, Thornwall GC, Evans J (2005) Comparing 2-D electrophoretic gels across internet databases. In: Walker J (ed) The proteomics protocols handbook. Humana Press, Totowa, pp 279–306

Li HB, Qin YM, Pang Y, Song WQ, Mei WQ, Zhu YX (2007) A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. New Phytol 175:462–471

Liu D, Tu L, Wang L, Li Y, Zhu L, Zhang X (2008) Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. Plant Cell Rep 27:1385–1394

Liu K, Han M, Zhang C, Yao L, Sun J, Zhang T (2012) Comparative proteomic analysis reveals the mechanisms governing cotton fiber differentiation and initiation. J Proteomics 75:845–856

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 38:D204–D210

Olsen KM, Wendel JF (2013a) A bountiful harvest: genomic insights into crop domestication phenotypes. Annu Rev Plant Biol 64:47–70

Olsen KM, Wendel JF (2013b) Crop plants as models for understanding plant adaptation and diversification. Front Plant Sci 4:1–16

Pang CY, Wang H, Pang Y, Xu C, Jiao Y, Qin YM, Western TL, Yu SX, Zhu YX (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

Paterson AH (2005) Polyploidy, evolutionary opportunity, and crop adaptation. Genetics of adaptation. Genetica 123:191–196

Rapp RA, Haigler CH, Flagel L, Hovav RH, Udall JA, Wendel JF (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. BMC Biol 8:139

Rose JK, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS (2004) Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant J 39:715–733

Ruan YL, Llewellyn DJ, Furbank RT, Chourey PS (2005) The delayed initiation and slow elongation of fuzz-like short fibre cells in relation to altered patterns of sucrose synthase expression and plasmodesmata gating in a lintless mutant of cotton. J Exp Bot 56:977–984

Salnikov VV, Grimson MJ, Seagull RW, Haigler CH (2003) Localization of sucrose synthase and callose in freeze-substituted secondary-wall-stage cotton fibers. Protoplasma 221:175–184

Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics 6:1638–1655

Stiff RM, Haigler HC (2012) Recent advances in cotton fiber development. In: Oosterhuis DM, Cothren JT (eds) Flowering and fruiting in cotton. The Cotton Foundation, Tennessee, pp 163–192

Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540–1542

Tan J, Tu L, Deng F, Hu H, Nie Y, Zhang X (2013) A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. Plant Physiol 162:86–95

Tang WH, Shilov IV, Seymour SL (2008) Nonlinear fitting method for determining local false discovery rates from decoy database searches. J Proteome Res 7:3661–3667

Thelen JJ, Peck SC (2007) Quantitative proteomics in plants: choices in abundance. Plant Cell 19:3339–3346

Udall JA, Wendel JF (2006) Polyploidy and crop improvement. Crop Sci 46:S-3–S-14

Vercauteren FG, Arckens L, Quirion R (2007) Applications and current challenges of proteomic approaches, focusing on two-dimensional electrophoresis. Amino Acids 33:405–414

Wang L, Li X-R, Lian H, Ni D-A, He Y-k, Chen X-Y, Ruan Y-L (2010) Evidence that high activity of vacuolar invertase Is required for cotton fiber and arabidopsis root elongation through osmotic dependent and independent pathways, respectively. Plant Physiol 154:744–756

Wendel JF (1989) New World tetraploid cottons contain Old World cytoplasm. Proc Natl Acad Sci USA 86:4132–4136

Wendel JF, Brubaker CL, Percival AE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310

Wendel JF, Brubaker CL, Seelanan T (2010) The origin and evolution of *Gossypium*. In: Stewart JM, Oosterhuis DM, Heitholt JJ, Mauney JR (eds) Physiology of cotton. Springer, The Netherlands, pp 1–18

Wendel JF, Flagel LE, Adams KL (2012) Jeans, genes, and genomes: cotton as a model for studying polyploidy. In: Soltis PS, Soltis DE (eds) Polyploidy and genome evolution. Springer, Berlin, Heidelberg, pp 181–207

Wu WW, Wang G, Baek SJ, Shen RF (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. J Proteome Res 5:651–658

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

Yao Y, Yang YW, Liu JY (2006) An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. Electrophoresis 27:4559–4569

Yoo MJ, Wendel JF (2014) Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. PLoS Genet 10:e1004073

Zhang B, Yang YW, Zhang Y, Liu JY (2013) A high-confidence reference dataset of differentially expressed proteins in elongating cotton fiber cells. Proteomics 13:1159–1163

Zhao PM, Wang LL, Han LB, Wang J, Yao Y, Wang HY, Du XM, Luo YM, Xia GX (2010) Proteomic identification of differentially expressed proteins in the Ligon *lintless* mutant of upland cotton (*Gossypium hirsutum* L.). J Proteome Res 9:1076–1087

Zhou M, Sun G, Sun Z, Tang Y, Wu Y (2014) Cotton proteomics for deciphering the mechanism of environment stress response and fiber development. J Proteomics 105:74–84