EVOLUTION & DEVELOPMENT 10:5, 567–582 (2008)

Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*

Bhupendra Chaudhary,[a,b,1] Ran Hovav,[a,1] Ryan Rapp,[a] Neetu Verma,[a] Joshua A. Udall,[c] and Jonathan F. Wendel[a,*]

[a]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA
[b]Department of Botany, Visva-Bharati, Santiniketan 731 235 WB, India
[c]Department of Plant and Wildlife Science, Brigham Young University, Provo, UT, 84042 USA
*Author for correspondence (email: jfw@iastate.edu)
[1]Equally contributed to this study

SUMMARY *Gossypium barbadense* is widely cultivated because of its extra-long staple cotton with superior luster, silkiness and high yield. These economically important traits were selected during initial domestication of an agronomically inferior wild ancestor, followed by millennia of human-mediated selection. To reveal the effects of this history on the cotton fiber transcriptome, we conducted comparative expression profiling on mechanically isolated fiber cells at three different stages encompassing early, mid, and late fiber elongation in wild (K101) and domesticated (Pima S-7) accessions, using a microarray platform that interrogates 42,429 unigenes. The distribution of differentially expressed genes across developmental stages was different in the two accessions, with a shift toward greater change earlier in cultivated than in wild *G. barbadense*. Approximately 4200 genes were differentially expressed between wild and domesticated accessions at one or more of the stages studied. Domestication appears to have led to enhanced modulation of cellular redox levels and the avoidance or delay of stress-like processes. Prolonged fiber growth in cultivated relative to wild *G. barbadense* is associated with upregulation of signal transduction and hormone signaling genes and down-regulation of cell wall maturation genes. Clues are provided into the processes and genes that may unwittingly have been selected by humans during domestication and development of modern elite lines. Several of the transcriptomic differences between wild and domesticated *G. barbadense* described here appear to have parallels in a second domesticated cotton species, *Gossypium hirsutum*, suggesting that replicated domestication of two different species has resulted in overlapping, parallel, metabolic transformations.

INTRODUCTION

The cotton genus (*Gossypium* L.) includes 45 diploid species and five allotetraploid species, the latter originating from a polyploidy event $\sim$ 1–2 Ma (Wendel and Albert 1992; Seelanan et al. 1997; Wendel and Cronn 2003). The five allotetraploid (2 $n = 52$) species carry "A" and "D" genomes, derived from their Old World A-genome and New World D-genome diploid progenitors (Wendel and Cronn 2003) (Fig. 1). Two allotetraploid species, *Gossypium hirsutum* L. (the source of "Upland" cotton) and *Gossypium barbadense* L. (the source of "Pima" or "Egyptian" cotton) were independently domesticated at least 4000 years ago (Wendel 1995; Dillehay et al. 2007) for their single-celled, epidermal trichomes, colloquially known as "cotton fiber," and cultivars derived from these two species now dominate world cotton commerce. Both species have large indigenous (= wild plus pre-Columbian) ranges in the New World tropics and subtropics (Fryxell 1979), with *G. hirsutum* predominantly distributed in Mesoamerica and the Caribbean, and *G. barbadense* distributed more southerly in South America and the Caribbean, where it overlaps and hybridizes with *G. hirsutum* (Brubaker et al. 1993). Archeological (Dillehay et al. 2007) and molecular (Percy and Wendel 1990; Westengen et al. 2005) evidence indicate that *G. barbadense* was first domesticated on the western slopes of the northern Peruvian Andes, with subsequent dispersal into regions east of the Andes, the Caribbean, and finally in post-Columbian times to Africa and elsewhere in the world. This complex evolutionary history thus involved multiple cultures and thousands of years of human-mediated selection, resulting in myriad forms spanning the wild-to-highly-improved continuum, the latter including common designators such as Egyptian and Pima cotton. Accompanying changes associated with domestication and crop improvement include annualization, photoperiod insensitivity, early flowering, larger boll size, and of particular relevance to this article, greatly

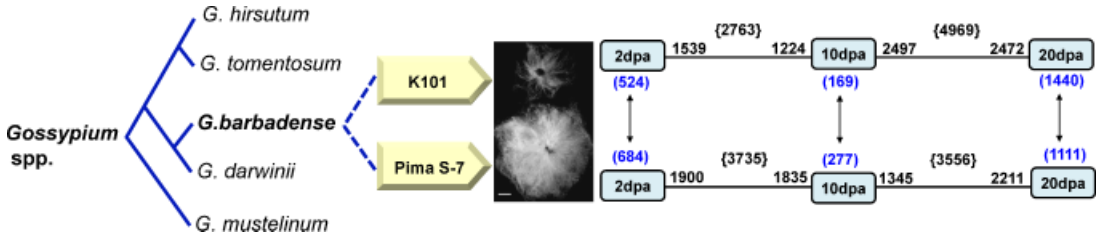


Fig. 1. Schematic representation of the phylogenetic history of allotetraploid (AD genome) *Gossypium*, which radiated into the three lineages and five species shown following polyploid formation in the mid-Pleistocene (branch lengths are arbitrary). Studied here are a wild form (K101) and a modern elite cultivar (Pima S-7) of *G. barbadense*. A representative image of a single seed from each, with attached trichomes ("cotton fiber"), is shown for the period of mid-elongation, 10 days post-anthesis (bar = 5 mm). Microarray analysis was performed for three replicates of each of three stages (represented by boxes) for both accessions. Numbers marked on the line designate the number of genes up-regulated at least 2-fold ($P < 0.05$ and FDR < 0.01) relative to their adjacent developmental stage, with the total in the contrast given in brackets {}. Numbers marked in brackets ( ) denote the numbers of differentially expression genes between wild and cultivated *G. barbadense* at each stage.

enhanced fiber yield and quality, including greater strength, fineness, and length.

Presently little is understood regarding the genetic changes associated with enhanced cotton fiber phenotypes. Notwithstanding striking discoveries of the genetic basis of morphological traits in some plants (Doebley et al. 1997; Doebley 2004, 2006; Konishi et al. 2006; Li et al. 2006; Simons et al. 2006; Sweeney et al. 2006; Komatsuda et al. 2007), relatively little is understood about the manner in which gene networks and biological processes are altered during phenotypic evolution. Here our goal was to understand the system-wide effects that underlie phenotypic change in *G. barbadense*, thereby contributing to our understanding of the genetics of plant domestication (Paterson et al. 1995; Paterson 2002; Doebley 2006) and morphological evolution (Kramer and Irish 1999; Cubas et al. 2001; Doebley 2004, 2006; Kellogg 2004; Vollbrecht et al. 2005; Wang et al. 2005; Clark et al. 2006; Ehrenreich and Purugganan 2006; Konishi et al. 2006; Li et al. 2006; Simons et al. 2006; Sweeney et al. 2006; Komatsuda et al. 2007; Sliwinski et al. 2007). Our approach was to use comparative expression profiling with a recently developed microarray platform containing probes for over 42,000 unigenes and accessions representing both wild and cultivated *G. barbadense*. Expression differences were tabulated across fiber developmental stages in each accession, as were differences between accessions, illuminating possible changes in developmental programs wrought by human selection. This approach is useful as a prelude to identify candidate genes and processes responsible for the difference in fiber quality between wild and cultivated accessions.

MATERIALS AND METHODS

Plant materials and tissue collection

Two accessions of *G. barbadense* were studied, the elite cultivar Pima S-7 and a wild accession from Bolivia, K101, the latter shown by previous allozyme analysis (Percy and Wendel 1990) to be representative of a wild progenitor. Seeds of both accessions were sown in sterilized potting mix in the Pohl Conservatory at Iowa State University, Ames, IA, and three biological replicates were grown for 3–5 months. Flowers were tagged at anthesis and harvested at three time points, 2, 10, and 20 days postanthesis (hereafter, dpa). For each of three biological replicates, ovules were excised, frozen in liquid nitrogen, and stored at -80°C. Additionally, ovules were visually inspected for cell damage and the fibers were inspected for contaminating tissue.

Isolation of total RNA from fibers

Fibers were mechanically separated from ovular tissue using a modification of a liquid N_2 and glass bead shearing procedure (Taliercio and Boykin 2007), as detailed in Hovav et al. (2008). RNAs were extracted using a modified hot-borate procedure optimized for cotton (Wan and Wilkins 1994). All purified RNA samples were quantified on a NanoDrop spectrophotometer and assayed for integrity and purity with a BioAnalyzer (Agilent, Palo Alto, CA, USA).

Amplification of RNA and microarray hybridizations

A total of 18 fiber samples from the two *G. barbadense* accessions were used for RNA isolations. RNA samples were treated with DNase following the manufacturer's protocol (New England Biolabs, Ipswich, MA, USA) and amplified with TargetAmp™ 1-Round aRNA Amplification kit from Epicentre Biotechnologies (Madison, WI, USA). Amplified RNAs were quantified and visually assessed for degradation and DNA contamination via a Bioanalyzer chip (Agilent Technologies, Santa Clara, CA, USA). A total of 12 µg amplified RNA in each of the samples were shipped to NimbleGen Systems (Madison, WI, USA) for cDNA synthesis, labeling, and hybridization to 18 custom microarrays. This custom microarray was developed following a new EST assembly (Udall et al. 2007), building on our first-generation arrays (Udall et al. 2006). The chip features 283,003 oligonucleotide probes, 60 bases in length, designed to interrogate 42,429 genes (for details www.cottonevolution.info), yielding an average redundancy of seven probes per gene, designed where possible to be nonspecific to the two

duplicates (homoeologs) of each gene that resulted from polyploidization between A and D genome diploids approximately 1–2 Ma.

Statistical analysis

A balanced developmental design for microarray analysis was used (Fig. 1). Three developmental time-points were sampled at 2, 10, and 20 dpa, for each of the two accessions, with three biological replicates of each, for a total of 18 microarrays. Statistical analyses were performed using R and SAS statistical software. Raw data values for each microarray were natural log transformed, median centered, and scale normalized across all arrays. All values of the 42,429 genes with seven redundant probes for per gene were median polished with Tukey's biweight method (Tukey 1977; Velleman and Hoaglin 1981) before performing a standard mixed linear model (LM) analysis of the data for expression differences between each fiber developmental stage within and between accessions. We considered a standard mixed LM for the data for each gene as:

$$y_{ijk} = \mu + \delta_i + \tau_j + s_k + \delta\tau_{ij} + e_{ijkm}$$

where y_{ijk} denotes the normalized log-scale signal intensity for genotype i, time-point j, biological replication k; μ denotes an intercept parameter; δ_i denotes the fixed effect of genotype i; τ_j denotes the fixed effect of time-point j; s_k denotes the random effect of replication k; $\delta\tau_{ij}$ denotes the fixed effect of the interaction between genotype i and time-point j; e_{ijk} denotes a random error term intended to capture all other sources of variability. Contrasts for differential expression between genotypes, time points and the interaction genotype × time-points were conducted using this model. For each gene, differences were calculated using the following pairwise contrasts: (K10-K2); (K20-K10); (K20-K2) and (P10-P2); (P20-P10); (P20-P2), where letters denote genotype (K101 or Pima S-7) and numbers denote developmental time-point (2,10 and 20 dpa). The 42,429 *P*-values from each comparison were converted to *q*-values using the method of (Storey and Tibshirani 2003). These *q*-values were used to identify the number of differentially expressed genes for a given comparison when controlling the false discovery rate (FDR) at various levels.

Blast2GO (http://www.blast2go.de/) was used to posit biochemical pathways involved in a given comparison and to calculate the statistical significance of each. Blast2GO includes the Gossip package (Conesa et al. 2005) for statistical assessment of annotation differences between two sets of sequences, using Fisher's exact test for each GO term. FDR controlled *p*-values (FDR <0.05) were used for the assessment of differentially significant metabolic pathways.

Quantitative RT-PCR analyses

Amplified aRNA was used as a template for first strand cDNA synthesis with the Super-script II preamplification system reverse transcriptase kit (Invitrogen, Carlsbad, CA, USA) at 42°C according to the manufacturer's instructions. We used conserved sequences to generate specific primer pairs for the reference sequences (Table S1), using RNA helicase (Q9ZS12) as the reference gene, as it was found to be approximately equally expressed in developing fibers and other plant tissues (Hovav et al. 2007). Quantitative PCR amplification was performed on cDNA using the GeneAmp 5700 Sequence Detection System (PE Biosystems, Foster City, CA, USA) with SYBR Green Master Mix containing AmpliTaq Gold, according to the manufacturer's instructions. All reactions were calibrated against a standard curve.

RESULTS

Intraspecific and temporal variation in gene expression in *G. barbadense* fibers

To explore the impact of domestication on global patterns of gene expression in developing fibers, we compared mRNA expression levels in developing fibers derived from wild (K101) and domesticated (Pima S-7) *G. barbadense* at three developmental stages. This analysis revealed the numbers of differentially expressed genes between adjacent time-points during fiber development within and between accessions (Fig. 1). The distribution of differentially expressed genes was relatively even among developmental stages in cultivated cotton, whereas in the wild accession K101, expression changes were biased toward later in development. In Pima S-7, 3735 (2 vs. 10 dpa) and 3556 (10 vs. 20 dpa) genes (~ 9% and 8% of total genes, respectively) were differentially expressed ($P<0.05$ and FDR <0.01, >2-fold expression change), whereas in the wild accession K101, 2763 (2 vs. 10 dpa) and 4969 (10 vs. 20 dpa) genes (~ 6.5% and 12%, respectively) were differentially expressed (Fig. 1). When comparing accessions at each stage, a total of 684 and 277 genes were up-regulated in Pima S-7 at early (2 dpa) and mid-stages (10 dpa) of elongation, with lower numbers up-regulated in K101 at these same two stages (524 and 169, respectively). By 20 dpa, during the transition to secondary wall synthesis, a much larger number (2551 total) of genes were differentially expressed between wild and cultivated cotton, with a bias toward up-regulation in K101 (1440 vs. 1111 up-regulated in Pima). Thus, domestication may have led to enhanced and/or earlier expression of some genes.

Categories of differentially expressed genes at different stages of fiber development in *G. barbadense*

Assessment of biological processes for the differentially expressed genes revealed several categories that are shared between wild and cultivated accessions, including "chromatin assembly" "transcriptional activator activity" "porin activity" and "lignin biosynthetic process" (Table S2). In addition, some biological processes were differentially over-represented at each stage of fiber development (Table 1).

The following functional categories were differentially up-regulated early (2 dpa) in Pima S-7: (i) "response to hormone stimulus," including genes encoding catalase, peroxidase and ascorbate peroxidase which have putative roles in scavenging

Table 1. Catalog of biological processes statistically over-represented during fiber developmental stages of Pima S-7 and K101 ($P < 0.05$ and FDR < 0.01, > 2-fold expression change)

Developmental contrast	Process up-regulated in	GO annotation	Functional categories
2 versus 10 dpa Processes up in 2 dpa	PimaS-7	GO:0009725	Response to hormone stimulus
		GO:0048731	System development
		GO:0007167	Enzyme linked receptor protein signaling pathway
		GO:0018117	Protein amino acid adenylylation
		GO:0047763	Caffeate O-methyltransferase activity
		GO:0004321	Fatty-acyl-CoA synthase activity
		GO:0009744	Response to sucrose stimulus
		GO:0004019	Adenylosuccinate synthase activity
		GO:0005496	Steroid binding
		GO:0019748	Secondary metabolic process
		GO:0007166	Cell surface receptor linked signal transduction
	K101	GO:0042586	Peptide deformylase activity
		GO:0019988	Charged-tRNA modification
		GO:0008891	Glycolate oxidase activity
2 versus 10 dpa Processes up in 10 dpa	PimaS-7	GO:0006120	Mitochondrial electron transport, NADH to ubiquinone
		GO:0006564	L-serine biosynthetic process
		GO:0046466	Membrane lipid catabolic process
		GO:0006119	Oxidative phosphorylation
		GO:0004364	Glutathione transferase activity
		GO:0016655	Oxidoreductase activity, acting on NADH or NADPH
		GO:0010241	*ent*-kaurene oxidase activity
		GO:0016838	Carbon–oxygen lyase activity, acting on phosphates
		GO:0008934	Inositol-1(or 4)-monophosphatase activity
		GO:0009854	Oxidative photosynthetic carbon pathway
		GO:0016859	*cis–trans* isomerase activity
		GO:0003924	GTPase activity
		GO:0030286	Dynein complex
	K101	GO:0005976	Polysaccharide metabolic process
		GO:0006528	Asparagine metabolic process
		GO:0016884	Carbon–nitrogen ligase activity
		GO:0008559	Xenobiotic-transporting ATPase activity
		GO:0042447	Hormone catabolic process
		GO:0000038	Very-long-chain fatty acid metabolic process
		GO:0004673	Protein histidine kinase activity
		GO:0010181	FMN binding
		GO:0016789	Carboxylic ester hydrolase activity
		GO:0004022	Alcohol dehydrogenase activity
		GO:0004322	Ferroxidase activity
		GO:0006468	Protein amino acid phosphorylation
		GO:0048765	Root hair cell differentiation
		GO:0046906	Tetrapyrrole binding
10 versus 20 dpa Processes up in 10 dpa	PimaS-7	GO:0016838	Carbon–oxygen lyase activity, acting on phosphates
		GO:0042389	ω-3 fatty acid desaturase activity
		GO:0016298	Lipase activity
		GO:0003700	Transcription factor activity
		GO:0016772	Transferase activity, transferring phosphorus groups
		GO:0005262	Calcium channel activity
		GO:0010143	Cutin biosynthetic process
		GO:0019104	DNA *N*-glycosylase activity
		GO:0009825	Multidimensional cell growth
		GO:0000270	Peptidoglycan metabolic process
	K101	GO:0004289	Subtilase activity
		GO:0019005	SCF ubiquitin ligase complex
		GO:0006605	Protein targeting

Table 1. (Contd.)

Developmental contrast	Process up-regulated in	GO annotation	Functional categories
		GO:0006461	Protein complex assembly
		GO:0010330	Cellulose synthase complex
		GO:0042562	Hormone binding
		GO:0009086	Methionine biosynthetic process
		GO:0016110	Tetraterpenoid catabolic process
		GO:0008061	Chitin binding
		GO:0017091	AU-specific RNA binding
		GO:0009873	Ethylene mediated signaling pathway
10 versus 20 dpa Processes up in 20 dpa	PimaS-7	GO:0042126	Nitrate metabolic process
		GO:0008194	UDP-glycosyltransferase activity
		GO:0003677	DNA binding
		GO:0009927	Histidine phosphotransfer kinase activity
		GO:0005982	Starch metabolic process
		GO:0015646	Permease activity
		GO:0016801	Hydrolase activity, acting on ether bonds
		GO:0015833	Peptide transport
	K101	GO:0050278	Sedoheptulose-bisphosphatase activity
		GO:0046914	Transition metal ion binding
		GO:0051260	Protein homooligomerization
		GO:0010313	Phytochrome binding
		GO:0016760	Cellulose synthase (UDP-forming) activity
		GO:0001718	Conversion of met-tRNAf to fmet-tRNA
		GO:0042586	Peptide deformylase activity

FDR, false discovery rate.

reactive oxygen species (ROS) (mainly H_2O_2) produced during oxidative stress (Li et al. 2007); (ii) "system development," including genes for receptor-like kinase (RLKs) and auxin-regulated proteins shown to be involved in cell differentiation and growth (Morris and Walker 2003); (iii) "protein amino acid adenylylation," containing four genes from a family of jasmonic acid–amino acid-conjugating enzymes, known to inhibit programmed cell death processes (Wasternack 2007); and (iv) "cell surface receptor linked signal transduction," containing protein kinase, leucine-rich repeat (LRR) protein and receptor-like protein kinases (RLKs) important for cell growth and development (Morris and Walker 2003; Li et al. 2005). However, along with these up-regulated categories in domesticated cotton, some differentially up-regulated functional categories early (2 dpa) in wild cotton were also observed, but none was found to be correlated with fiber development (Table 1).

At 10 dpa, fiber elongation and primary wall synthesis are prominent processes. The following biological processes were included in the over-represented categories in Pima S-7 (Table 1): (i) "oxidative phosphorylation," presumably associated with the high metabolic rates required during rapid cell growth (Attucci et al. 1991); (ii) "glutathione S-transferase activity" (GST), related to protection against oxidative stress (Liu et al. 2006); (iii) "*ent*-kaurene oxidase activity," which includes genes for gibberellin biosynthesis (Helliwell et al. 1999); (iv) "calcium channel activity," a process induced during oxidative stress (Rentel and Knight 2004) (v) "multidimensional cell growth," with several sterol-C-methyltransferase genes that play multiple roles in cell growth and development (Liu et al. 2006); (vi) "transcription factor activity," potentially reflecting diverse cellular processes such as lipid trafficking, transcription, regulation, and signal transduction involved in enhanced fiber elongation in cultivated cotton.

Over-represented functional categories in K101 at 10 dpa that may be related to fiber development included (i) "ferroxidase activity," perhaps regulating free iron so that ROS do not reach toxic levels (Orino et al. 2001); (ii) "tetrapyrole activity," related to cell-death and plastid-to-nucleus signal transduction (Tanaka and Tanaka 2007); (iii) "subtilase activity," representing a group of serine protease proteins, mainly known as toxic proteins to the cell perhaps related to stress conditions (Morinaga et al. 2007); (iv) "SCF ubiquitin-ligase complex," which facilitates ubiquitylation of proteins targeted for degradation (Ni et al. 2004); (v) "cellulose synthase complex," suggesting premature secondary cell wall thickening leading to fiber growth inhibition (John and Crow 1992; Gou et al. 2007) (vi) "AU-specific RNA binding," implicated in programmed cell-death in plants (Ohme-Takagi

et al. 1993); and (vii) "ethylene mediated signaling pathway," known as a key mediator of fiber cell growth (Drew et al. 2000).

By 20 dpa, *G. barbadense* fibers are transitioning to secondary cell wall synthesis, which is accompanied by a cessation in elongation (there is some genotypic and phenotypic variation in this developmental timing; Basra and Malik 1984). In cultivated *G. barbadense*, biological processes statistically over-represented at this stage (Table 1) included (i) "nitrogen metabolism," with 11 genes, including glutamine synthatase and nitrate reductase; (ii) "DNA binding," with 167 up-regulated genes encoding many transcription factors such as helix-loop-helix DNA-binding, BEL1-like homeodomain, nucleoid DNA-binding protein, cnd41-like protein, and MYB transcription factors; and (iii) "histidine phosphotransfer kinase activity" which plays an important role in plant stress signaling (Pitzschke and Hirt 2006). In contrast, the wild accession K101 did not show any differentially up-regulated category that has been linked to fiber development (Table 1).

Over-represented processes and up-regulated genes in cultivated *G. barbadense*

As noted above, gene expression differences between K101 and Pima S-7 were tabulated across fiber developmental stages (Fig. 1), yielding a total of 1208, 446, and 2551 genes that were significantly ($P<0.05$ and FDR <0.05) differentially expressed at 2, 10, and 20 dpa, respectively. These differentially expressed genes were grouped at the biological process level (GO annotation) and subsequently, at single-gene level for each stage of fiber development (Tables 2 and 3, respectively). The expression pattern of some of these genes in Pima S-7 relative to K101 is summarized in Fig. 2.

The functional categories differentially up-regulated early (2 dpa) in Pima S-7 included (i) "transcription regulator activity," with 61 genes encoding auxin-regulated proteins, the transcription factor Myb1, zinc finger proteins and DNA binding factors; (ii) "response to jasmonic acid stimulus," contributing to gibberellin biosynthesis (Wasternack 2007); (iii) "MAP kinase activity," involved in signal transduction through phosphorylation-induced transcription in cotton (Ji et al. 2003); and (iv) "peroxidase activity," comprising peroxidase and CATA1_GOSHI Catalase isozyme-1 genes that catalyze oxidation–reduction reactions (Ni et al. 1990; Arpat et al. 2004). Other important functional categories included "response to hormone stimulus," "response to stress" and "voltage-gated calcium channel activity" (Table 2). These functional categories harbored genes related to fiber development, such as peroxidase, putative bHLH (helix–loop–helix) transcription factor, asparagine synthase, and carbonic anhydrase (Table 3, Fig. 2, A–C), suggesting their involvement in cell proliferation, protection against oxidative stress molecules and enhanced DNA binding (Brears et al. 1993; Friedrichsen et al. 2002; Hoang and Chapman 2002). An over-represented biological process in Pima S-7 is "body growth" (Table 2), within which a noticeable up-regulated gene in Pima S-7 is a "NAM-like protein" (*No Apical Meristem*) (Table 3; Fig. 2D), previously shown to regulate cell division activity, orientation, or cell expansion patterns (Aida and Tasaka 2006).

Comparison between the two *G. barbadense* accessions at 10 dpa showed "glycolate oxidase activity" and "peroxisomes" to be over-represented processes, suggesting their putative role in modulating ROS and H_2O_2 levels (Table 2) (Moller 2001). At the gene level, protein kinase (Fig. 2E) and RLK were over-represented in Pima S-7, suggesting that these genes are involved in enhanced cell growth in domesticated cotton (Table 3) (Shiu and Bleecker 2001; Chico et al. 2002; Pitzschke and Hirt 2006).

At 20 dpa, none of the up-regulated biological processes were inferred to be connected to fiber development except "voltage-gated calcium channel activity" (Table 2). At the gene level, a putative protein kinase and a peroxidase were up-regulated in Pima S-7; the first of these is important for protein modification through phosphorylation, whereas the second is involved in cellular redox modulation as well as cell wall extensibility (Table 3) (Chico et al. 2002; Arpat et al. 2004). An oleosin gene, known to be responsible for oil body biogenesis (Wahlroos et al. 2003) was also observed to be up-regulated in Pima S-7. The WRI1(WRINKLED1) transcription factor, known for its role in the control of metabolism, particularly in glycolysis (Cernac et al. 2006), was up-regulated in Pima S-7 (Figs. 2F and 5). The foregoing observations are suggestive of biological processes and genes that may have become up-regulated by the process of domestication, and by extension, suggests a role for these processes and genes in enhanced fiber growth.

Over-represented processes and up-regulated genes in wild *G. barbadense*

Among the up-regulated biological processes in K101 at 2 dpa, none were obviously associated with fiber growth (Table 2), consistent with its agronomically inferior fiber. At 10 dpa, some biological processes enriched for genes involved in "water channel activity" and "beta-glucuronidase activity," previously described as participating in cotton fiber elongation (Liu et al. 2006; Taliercio and Boykin 2007), were also observed to be up-regulated in the wild form (Table 2). A pore-forming, toxin-like protein, Hfr-2, involved in cell membrane disruption through pore formation (Puthoff et al. 2005), was up-regulated at 2 and at 10 dpa (Table 2, Fig. 2G).

At 20 dpa, none of the biological processes were obviously connected to fiber-specific developmental programs, but at the gene level, a pectinesterase-like protein was significantly up-regulated in K101 (Table 3, Fig. 2H). There is evidence that

Table 2. List of developmental stage-specific biological processes over-represented between K101 and Pima S-7 ($P < 0.05$ and FDR < 0.05)

Developmental stage	GO annotation	Biological function	FDR
2 dpa(PimaS-7) versus 2 dpa(K101)	GO:0030528[1]	Transcription regulator activity	1.61E-08
	GO:0035264[1]	Body growth	1.61E-08
	GO:0050896[1]	Response to hormone stimulus	1.45E-07
	GO:0009753[1]	Response to jasmonic acid stimulus	3.04E-04
	GO:0004066[1]	Asparagine synthase (glutamine-hydrolyzing) activity	1.48E-05
	GO:0006800[1]	Oxygen and reactive oxygen species metabolic process	0.001496
	GO:0019239[1]	Deaminase activity	0.002212
	GO:0006979[1]	Response to oxidative stress	6.97E-04
	GO:0005245[1]	Voltage-gated calcium channel activity	0.003294
	GO:0004707[1]	MAP kinase activity	0.004166
	GO:0004601[1]	Peroxidase activity	0.004166
	GO:0006575[2]	Amino acid derivative metabolic process	2.32E-04
	GO:0009813[2]	Flavonoid biosynthetic process	0.001985
	GO:0009096[2]	Aromatic amino acid family biosynthetic process	0.003656
	GO:0009502[2]	Photosynthetic electron transport chain	0.014448
	GO:0005507[2]	Copper ion binding	0.017381
	GO:0004375[2]	Glycine dehydrogenase (decarboxylating) activity	0.017482
	GO:0019842[2]	Vitamin binding	0.020633
	GO:0000075[2]	Cell cycle checkpoint	0.024942
10 dpa(Pima S-7) versus 10 dpa(K101)	GO:0008891[1]	Glycolate oxidase activity	2.39E-08
	GO:0009853[1]	Photorespiration	3.35E-08
	GO:0016614[1]	Oxidoreductase activity, acting on CH–OH group	3.72E-05
	GO:0031403[1]	Lithium ion binding	7.29E-05
	GO:0008934[1]	Inositol-1(or 4)-Monophosphatase activity	7.29E-05
	GO:0006564[1]	L-serine biosynthetic process	7.29E-05
	GO:0004514[1]	Nicotinate-diphosphorylase/carboxylating activity	0.002839
	GO:0008441[1]	3′(2′),5′-bisphosphate nucleotidase activity	0.00858
	GO:0009435[1]	NAD biosynthetic process	0.010159
	GO:0005777[1]	Peroxisome	0.022418
	GO:0005245[1]	Voltage-gated calcium channel activity	0.028523
	GO:0046982[1]	Protein heterodimerization activity	0.047864
	GO:0004437[1]	Inositol or phosphatidylinositol phosphatase activity	0.047864
	GO:0015250[2]	Water channel activity	3.55E-08
	GO:0009699[2]	Phenylpropanoid biosynthetic process	0.011644
	GO:0004566[2]	β-glucuronidase activity	0.01423
	GO:0003962[2]	Cystathionine gamma-synthase activity	0.022319
	GO:0004553[2]	Hydrolase activity	0.047461
	GO:0009886[2]	Post-embryonic morphogenesis	0.048676
20 dpa(Pima S-7) versus 20 dpa(K101)	GO:0042128[1]	Nitrate assimilation	0.007404
	GO:0008265[1]	Mo-molybdopterin cofactor sulfurase activity	0.012996
	GO:0008934[1]	Inositol-1(or 4)-monophosphatase activity	0.023998
	GO:0017057[1]	6-phosphogluconolactonase activity	0.02442
	GO:0004239[1]	Methionyl aminopeptidase activity	0.0354
	GO:0046556[1]	α-*N*-arabinofuranosidase activity	0.0354
	GO:0046423[1]	Allene-oxide cyclase activity	0.0354
	GO:0009044[1]	Xylan 1,4-β-xylosidase activity	0.038306
	GO:0004316[1]	3-oxoacyl-[acyl-carrier-protein] reductase activity	0.046045
	GO:0005245[1]	Voltage-gated calcium channel activity	0.046045
	GO:0016984[2]	Ribulose-bisphosphate carboxylase activity	0.007285
	GO:0030130[2]	Clathrin coat of trans-Golgi network vesicle	0.029553
	GO:0031976[2]	Plastid thylakoid	0.044023

[1]Processes up-regulated in Pima S-7,
[2]Processes up-regulated in K101. Blast2GO was used to infer a putative function for each gene.
Statistical analyses revealed classes of genes overrepresented relative to the total microarray probe set. FDR, false discovery rate.

Table 3. Some developmental stage-specific fiber genes up-regulated between K101 and Pima S-7 ($P < 0.05$ and FDR < 0.05)

Developmental contrast	Contig ID	Functional category	Fold change
2 dpa(Pima S-7) versus 2 dpa(K101)			
	Cotton16_16772_01[1]	Putative bHLH transcription factor	59.1
	Cotton16_09277_01[1]	Asparagine synthetase 2	35.2
	Cotton16_32620_01[1]	Ornithine decarboxylase	17.6
	Cotton16_24297_01[1]	Carbonic anhydrase	15.0
	Cotton16_13324_01[1]	TGF-β receptor, type I/II extracellular region	14.3
	Cotton16_32282_01[1]	NAM-like protein	13.6
	Cotton16_00336_03[2]	Pore-forming toxin-like protein Hfr-2	35.69
	Cotton16_35453_01[2]	Protease inhibitor/LTP family protein	7.91
	Cotton16_25592_01[2]	Basic blue copper protein	7.10
	Cotton16_08771_01[2]	Cytochrome P450 [Panax ginseng]	6.88
	Cotton16_17728_01[2]	Flavonoid 3′-hydroxylase	6.09
10 dpa(Pima S-7) versus 10 dpa(K101)			
	Cotton16_12044_01[1]	RNA-directed DNA polymerase	30.17
	Cotton16_51625_01[1]	Glycolate oxidase, putative	19.39
	Cotton16_53104_01[1]	Somatic embryogenesis receptor-like kinase 3	7.12
	Cotton16_30826_01[1]	Cyclin-like F-box	6.40
	Cotton16_55933_01[1]	Putative protein kinase	6.29
	Cotton16_00336_03[2]	Pore-forming toxin-like protein Hfr-2	13.65
	Cotton16_14894_01[2]	Nodulin-like protein	5.63
	Cotton16_45442_01[2]	Cytochrome P450	5.61
20 dpa(Pima S-7) versus 20 dpa(K101)			
	Cotton16_55933_01[1]	Putative protein kinase	10.52
	Cotton16_46395_01[1]	15.8 kDa oleosin	9.15
	Cotton16_00001_243[1]	Peroxidase	8.79
	Cotton16_51673_01[1]	WRI1; DNA binding/transcription factor	8.21
	Cotton16_00073_01[2]	Pectinesterase/methylesterase family protein	9.90
	Cotton16_00560_02[2]	Kinesin-like protein	9.10
	Cotton16_34905_01[2]	Phosphoglycerate dehydrogenase-like protein	8.67
	Cotton16_00043_04[2]	Putative phosphatase	8.58

Genes shown exhibit more than 5-fold expression change between K101 and Pima S-7, and are arranged in decreasing order of fold change.
[1]Up-regulated in Pima S-7;
[2]Up-regulated in K101.
FDR, false discovery rate.

exogenous pectinesterase induces secondary cell wall thickening and inhibits cell growth (Bosch et al. 2005), suggesting its involvement in the inception of early maturation and reduced growth in K101.

Up-regulated translational regulators in Pima S-7

We identified genes that were differentially expressed between the two accessions across *all three* studied fiber developmental stages. A total of 66 and 41 such up-regulated genes were identified in Pima S-7 and K101, respectively. Ten of these from Pima S-7 are shown in Table 3. In Pima S-7, many of these 66 genes encode protein and DNA binding factors, ATP binding, receptor-like protein kinase activity (RLK) as well as calcium ion transport activity (Table 4) and almost 40% of the total are transcription factors and translational regulators. Members of the LRR subfamily and RLK were the major subsets of genes highly over-expressed in Pima S-7, up to $\sim$ 62-fold in comparison to the wild accession K101 (Table 4, Fig. 3A). Additionally, "ring finger protein family" (SNURF), "zinc finger proteins" and "two-pore calcium channel" were also over-represented throughout fiber development (Fig. 3, B and C). These proteins are involved in signal transduction, basal transcription, and calcium ion flux (Furuichi et al. 2001). An over-represented gene implicated in protein trafficking, $'H^+$-transporting two-sector ATPase, exhibited up to a $\sim$ 45-fold difference in gene expression (Kim et al. 1999) (Fig. 3D). In contrast to the aforementioned genes in Pima S-7, none of the 40 genes differentially over-expressed in the wild accession at all three time points are obviously involved in a biological process known to be important to fiber development or elongation.

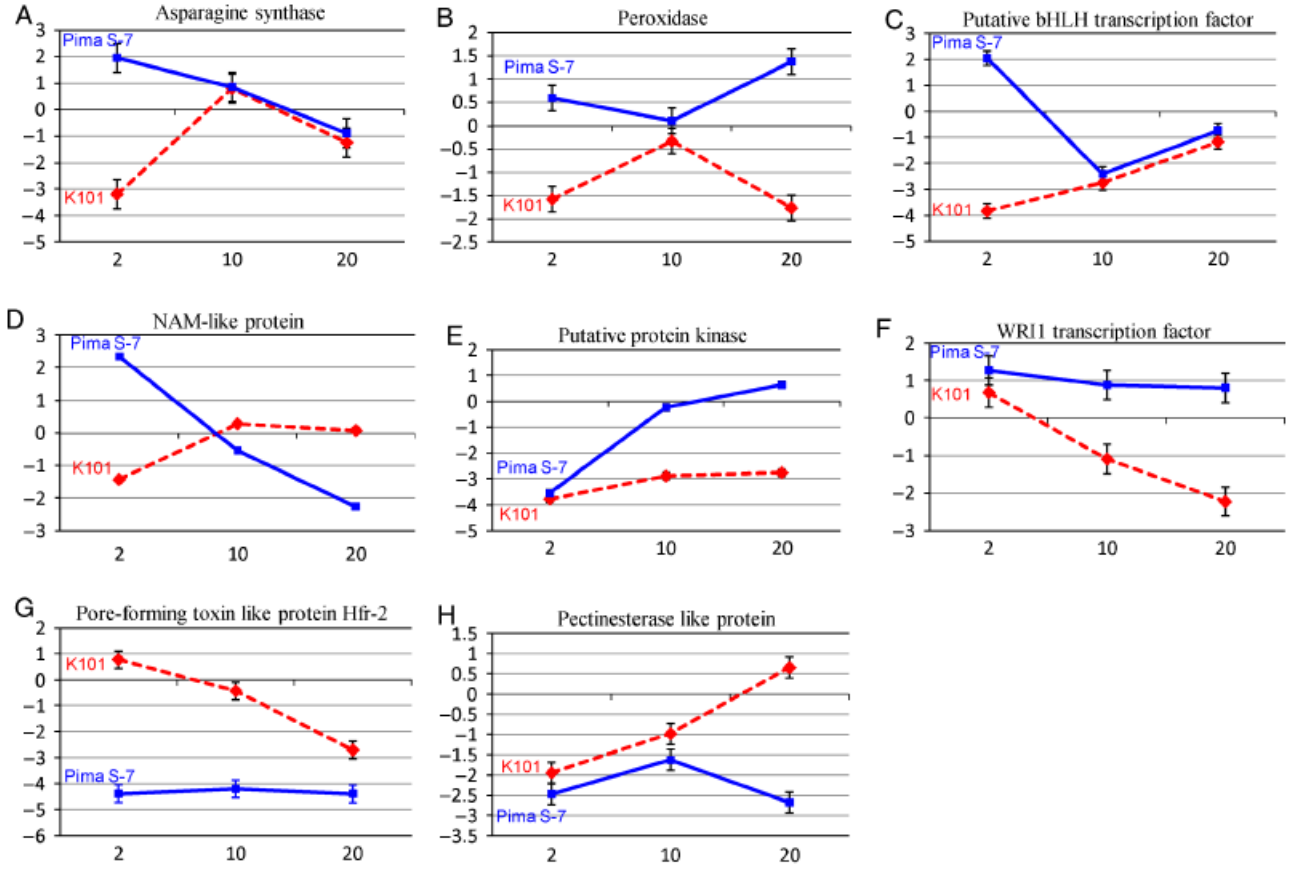


Fig. 2. Differential gene expression in wild and domesticated *Gossypium barbadense* at three stages (2, 10 and 20 dpa) of fiber cell development. Expression patterns are shown for (A) asparagine synthase (Cotton16_09277_01; AAC49613); (B) peroxidase (Cotton16_00001_243; AAA99868); (C) putative bHLH transcription factor (Cotton16_16772_01; ABW97699); (D) NAM-like protein (Cotton16_32282_01; AAF05865); (E) putative protein kinase (Cotton16_55933_01); (F) WRI 1 transcription factor (Cotton16_51673_01; NP_191000); (G) pore-forming toxin like protein Hfr-2 (Cotton16_00336_03; AAW48295); (H) pectinesterase-like protein (Cotton16_00073_01; AAM97070).

DISCUSSION

Crop domestication entails evolutionary responses in which plant populations adapt to human selection. Most plant species exhibit marked changes in a variety of phenotypes in response to this selection, most noticeably in traits consciously under selection (e.g., fruit size, yield, evenness of maturation) (Burger et al. 2008). As Darwin recognized (Darwin 1859), study of the phenotypic variation between wild and cultivated plants presents an opportunity to generate insight into general principles of evolution, using morphologically variable antecedent and descendant morphologies in a comparative framework. Here, we apply this logic to a comparative evolutionary analysis of the ovular epidermal seed trichomes that in domesticated forms are termed "cotton," but which vary dramatically in morphology among accessions and species (Fig. 1) (Applequist et al. 2001). Using wild and domesticated accessions of *G. barbadense*, we provide an initial perspective on the molecular-genetic correlates and responses (using gene expression differences as a proxy for genetic changes) of human artificial selection during the transition from a wild perennial plant bearing short, coarse, weak fibers to a modern annualized crop plant bearing much longer, finer, and stronger fibers. Our analysis reveals extraordinary complexity in the underlying genetic architecture, notwithstanding the fact that the trait in question is a mere single cell.

In general, cotton fiber development involves stages of cellular initiation, primary wall synthesis, secondary wall deposition, maturation, and death (Kim and Triplett 2001). Thousands of differentially expressed genes contribute to each stage of fiber development, although only a small number are known to be specifically involved in this complex developmental process. Many studies have identified and characterized genes regulating fiber development in cotton. These fiber-specific or fiber-enriched genes are involved in diverse pro-

Table 4. Ten highly up-regulated genes throughout fiber development in Pima S-7 ($P < 0.05$ and FDR < 0.05)

Contig ID	Blast hits	Fold change	Biological process
Cotton16_47501_01	NBS-LRR type disease resistance protein Rps1-k-1	62.61	Protein binding
Cotton16_36187_01	H^+-transporting two-sector ATPase	44.58	ATP binding
Cotton16_25302_01	Two-pore calcium channel	26.49	Calcium ion transport
Cotton16_55347_01	Retrotransposon protein,	16.53	RNA-directed DNA polymerase activity
Cotton16_56087_01	RING finger family protein	5.14	Protein binding
Cotton16_11435_01	GDP dissociation inhibitor	3.82	Rho GDP-dissociation inhibitor activity
Cotton16_39201_01	DNA binding/transcription factor	3.82	DNA binding
Cotton16_11436_01	Oxidoreductase	3.54	Oxidoreductase activity
Cotton16_44830_01	Mitogen-activated protein kinase 1	3.15	Protein serine/threonine kinase activity
Cotton16_03794_01	Zinc finger (Ran-binding) family protein	2.78	DNA binding

Genes are ordered in descending order of fold change (domesticated to wild).
FDR, false discovery rate.

cesses, including cell elongation and cell wall biogenesis (Wilkins and Jernstedt 1999), cytoskeleton formation (Whittaker and Triplett 1999), lipid trafficking (Lee et al. 2006), cell membrane and DNA metabolism (Wu et al. 2007), cell turgor pressure and plasmodesmatal dynamics (Ruan et al. 2001), and responses to phytohormones (Yang Samuel et al. 2006; Taliercio and Boykin 2007). Expression profiling of cotton fibers across a developmental time-course has demonstrated that the fiber transcriptome is exceptionally complex, with over half and perhaps three-quarters or more of the total genome transcribed at one or more stages of fiber development (Hovav et al. 2007).

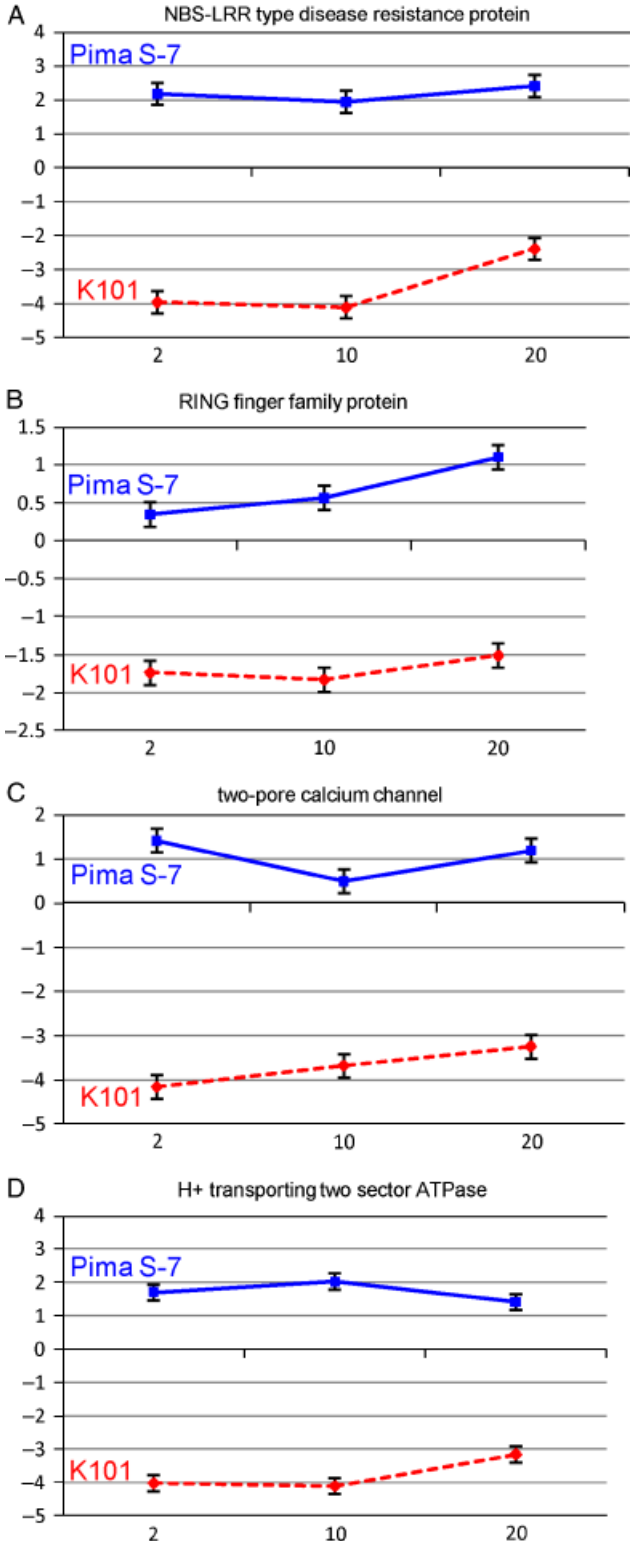
This remarkable complexity of the cotton fiber transcriptome is well-reflected in the present study, which interrogated tens of thousands of genes from early to late elongation and into the transition into secondary wall synthesis. Hundreds of genes were differentially expressed during development and between wild and cultivated accessions, reflecting a diversity of biochemical pathways and cellular processes. We note that the number of genes changing developmentally within accessions is greater than that between accessions at any stage ($\sim$ 10 times more in some cases), emphasizing the dynamic nature and developmental complexity of this unique plant cell type. Assessment of the nature of both classes of genes provides potential clues into development as well as evolutionary responses to selection.

An interpretive model for gene expression changes accompanying domestication

During fiber elongation various genes and biological processes were differentially expressed between domesticated and wild forms of *G. barbadense*. Here, we synthesize this information in the context of fiber development and previous studies of cotton fiber gene expression. An interpretive model is presented in Fig. 5, and is explained in the following paragraphs.

One of the key processes implicated in both cotton fiber development and evolution is regulation of H_2O_2 and other ROS. Rapid cell elongation involves cell-wall relaxation through nonenzymatic reactions mediated by H_2O_2 and other ROS that cleave polysaccharides (Fry 1998; Foreman et al. 2003; Liszkay et al. 2004). H_2O_2, therefore, is a necessary molecule for cell elongation, including *G. hirsutum* fibers where inhibition of H_2O_2 production or scavenging of existing H_2O_2 from the system prevents cell wall differentiation (Potikha et al. 1999). However, higher levels of H_2O_2 may arrest cell elongation through apparent stimulation of cell wall stiffening and can even promote programmed cell death or necrosis (Schopfer 1996; Rodriguez et al. 2002). Recent comparative proteomic analysis between regular and mutant cotton fibers also implicated an antioxidant protein, ascorbate peroxidase, in H_2O_2 homeostasis during cell development (Li et al. 2007). This cytosolic gene (*GhAPX1*) accumulates during cotton fiber elongation, and when in vitro cultured ovules are treated with exogenous H_2O_2 or ethylene, *GhAPX1* expression and total APX activity increases, as does fiber cell elongation. One possibility suggested by these observations is that a key difference between long (Pima S-7) and short (K101) fibers might be differential regulation and fine-tuning of cellular H_2O_2 and other ROS levels. This suggestion is supported in an evolutionary context by recent comparative expression profiling experiments conducted between wild and cultivated *diploid* cotton species (Hovav et al. 2008). In this study, gene expression analysis showed that many genes involved with stress responses were up-regulated early in wild accession, and those involved with modulating H_2O_2 and other ROS levels were up-regulated in domesticated accession, suggesting that the evolution of long spinable fibers in cotton was accompanied by novel expression of genes assist-

ing in the regulation of ROS. The present study thus parallels these earlier results but for a separate domestication event at the allopolyploid level.



As summarized in Fig. 5, no antioxidant genes were up-regulated in the wild accession at 2 dpa, whereas several key genes were in the domesticated accession Pima S-7. In wild cotton, one response to higher concentration of peroxides may be enhanced transcription of stress-related genes and production of ethylene. Ethylene may promote fiber elongation by activating genes important for cell wall synthesis, wall loosening, or cytoskeleton arrangement (Shi et al. 2006), but increased ethylene levels may lead to growth inhibition and cell death (Herbert et al. 2001). As shown in our model (Fig. 5), "ethylene responsive genes," up-regulated in Pima S-7 at 2 dpa, includes genes responsible for auxin biosynthesis, which is involved in ethylene-mediated cell growth (Rahman et al. 2001). This connection suggests that up-regulated ethylene responsive genes in Pima S-7 may fine-tune ethylene levels for maximizing fiber elongation in domesticated cotton.

An important class of ROS-regulating enzymes is the peroxidases, which eliminate toxic molecules such as superoxide and hydroxide radicals generated as a byproduct of aerobic respiration. As shown in Fig. 2B, peroxidase was up-regulated in Pima S-7 throughout the developmental stages studied. Additional RT-PCR data on wild and cultivated accessions of *G. hirsutum* as well as *G. barbadense* demonstrates a similar, parallel effect in a second domesticated cotton species; that is, peroxidase is over-expressed in the domesticated accession TM1 relative to a wild accession (Yuc) that represents a model of the ancestral forms first domesticated by humans (Fig. 4C). This parallel transformation accompanying domestication of two different species, as well as the data of Hovav et al. (2008) showing a similar effect between wild and domesticated diploid cotton species, constitutes *prima facie* evidence for a convergent, repeated metabolic shift during cotton evolution and subsequent domestication. One intriguing question that arises concerns the extent to which this metabolic convergence, at the level of antioxidant scavenging and associated physiology, reflects parallel or convergent genes and types of mutations.

ROS may also act as signal transducers. Relevant genes up-regulated in Pima S-7 fibers relative to K101 include LRR, RING finger, zinc-finger family proteins along with mitogen-activated protein kinase (MAPK), receptor like kinases (RLKs) and voltage-dependent plasma membrane Ca^{2+} channels. The LRR receptor kinases represent the largest group of receptor kinases, thought to mediate protein–protein

Fig. 3. Genes up-regulated across all three time points (2, 10, and 20 dpa) of fiber development in wild and cultivated *Gossypium barbadense*. Expression patterns are shown for (A) NBS-LRR (leucine-rich repeat) type disease resistance protein (Cotton16_47501_01; Q2YE88); (B) RING finger family protein (Cotton16_56087_01); (C) two-pore calcium channel (Cotton16_25302_01; BAB55460); and (D) H+ transporting two sector ATPase (Cotton16_36187_01).

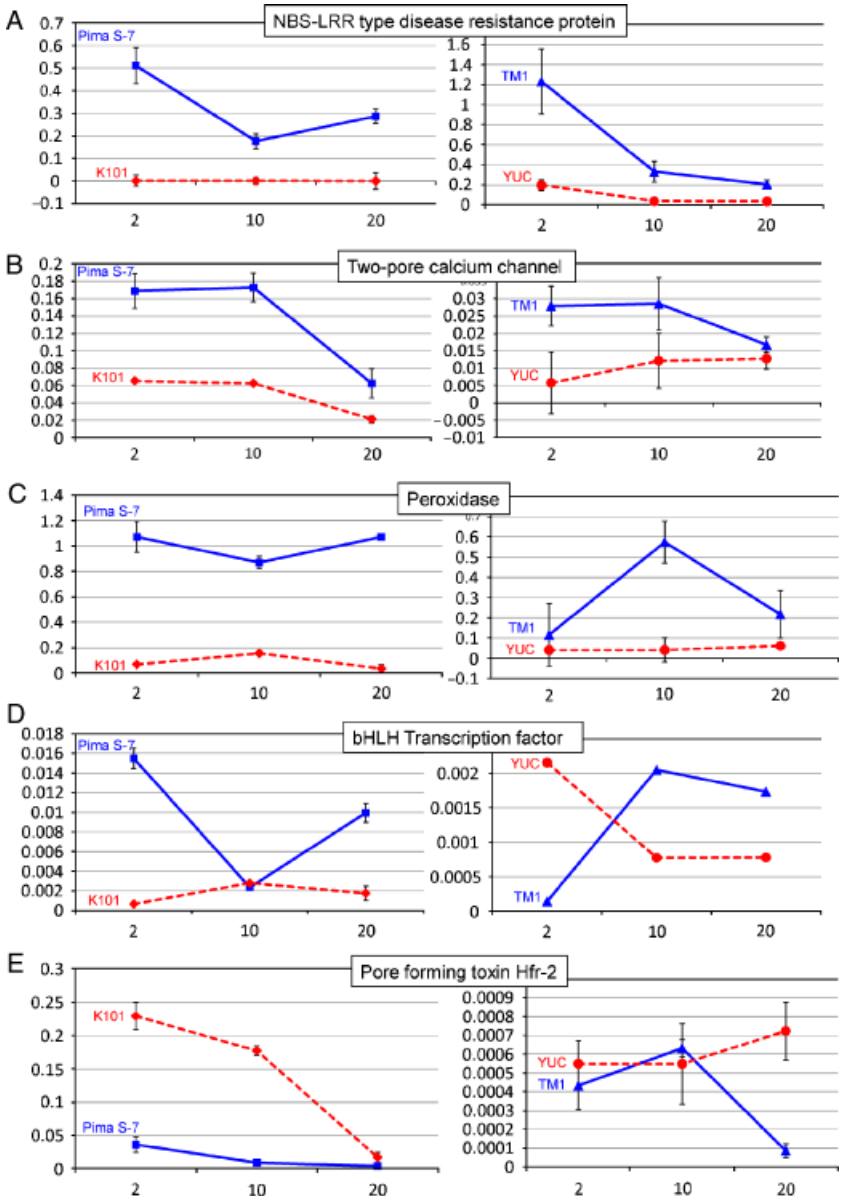


Fig. 4. Quantitative, real-time PCR analyses of five candidate genes controlling H_2O_2 levels and potentially involved in enhanced fiber elongation. Investigated were cultivated and wild forms of the two domesticated cotton species, *Gossypium barbadense* (cultivated = Pima S-7; wild = K101), and *Gossypium hirsutum* (cultivated = TM1; wild = *yucatanense*, YUC). Each point on the graph represents the mean of three biological replications. The *x*-axis shows three stages of fiber development and the *y*-axis indicates relative gene expression.

interactions (Kobe and Deisenhofer 1994; Shiu and Bleecker 2001). They have been suggested to play a role in signal transduction for enhanced transcription associated with cotton fiber development (Li et al. 2005). The fact that cotton LRRs are expressed at much higher levels in the fibers of Pima S-7 (Fig. 3A) makes these a promising class of transcription factors for further investigation. As was the case with peroxidases, RT-PCR data on a second cultivated cotton species, *G. hirsutum*, (Fig. 4A), raises the intriguing possibility that human selection for enhanced fiber length resulted in parallel or convergent adjustments in LRR expression levels as well.

In plants, calcium channels are known to play essential roles in initiating signal transduction processes during plant growth and development (Thuleau et al. 1994; Pottosin and Schonknecht 2007). As shown in Fig. 5, a consistently over-represented regulatory gene family in Pima S-7 is ''two-pore calcium channel'' during fiber development (Fig. 3C). Higher concentrations of Ca^{+2} in fiber initials than in surrounding ovular cells, and certain genes enriched in fiber initials and elongating fibers have been identified through microarray and fluorescent dye staining (Taliercio and Boykin 2007). In present study, quantitative RT-PCR analyses confirmed higher levels of a calcium channel gene in both domesticated *G. barbadense* and *G. hirsutum*, relative to their wild antecedents (Fig. 4B), lending support to the hypothesis of a role for calcium channel activity in fiber elongation.

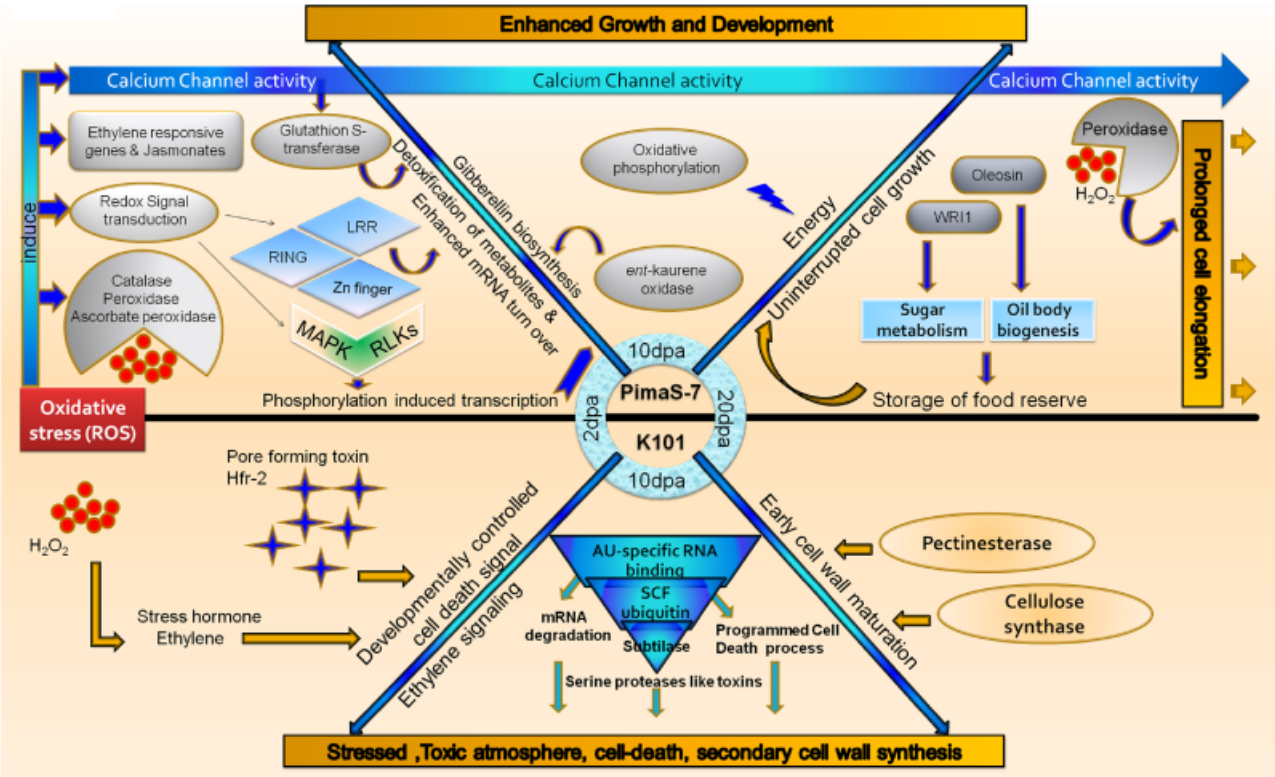


Fig. 5. A hypothetical model based on microarray results showing candidate genes and biological processes statistically over-represented in wild and cultivated *Gossypium barbadense* at three developmental stages of fiber (2, 10, and 20 dpa) studied (see text for details). The central black line in the model separates up-regulated processes/genes in Pima S-7 and K101 at three stages of fiber development. The four large arrows demarcate developmental stages. Various up-regulated biological functions/genes involved early (2 dpa) in fiber development are shown, including genes up-regulated in Pima S-7 that encode proteins involved in oxidative stress, redox signaling, phophorylation-induced transcription, and hormone signaling. At 10 dpa in Pima S-7, oxidative phosphorylation and *ent*-kaurene oxidase activity are shown as prominent processes for energy production and gibberellin biosynthesis. At 20 dpa in Pima S-7, peroxidase is up-regulated, as are WRI1 and oleosin for sugar metabolism and oil body biogenesis, respectively. In wild cotton (K101, lower half) at 2 dpa, oxidative stress (ROS) and stress hormone signaling suggest earlier onset of programmed cell death-like processes. By 10 dpa, additional processes are shown related to mRNA degradation and programmed cell death signals. At 20 dpa, cellulose synthase and pectinesterase are shown as up-regulated, leading to early cell wall maturation and cessation of elongation.

A "putative bHLH (helix-loop-helix) transcription factor" (Friedrichsen et al. 2002) was ranked as the most up-regulated gene in Pima S-7, with an $\sim$60-fold change in expression relative to K101 (Table 3, Fig. 2C). These observations are relevant to our RT-PCR results (Fig. 4D), which show that in cultivated *G. hirsutum*, bHLH transcription factor expression at 2 dpa was less than in a wild accession, but at 10 and 20 dpa it was significantly higher in cultivated *G. hirsutum*. Another up-regulated gene in Pima S-7 is "asparagine synthase" (Fig. 2A) earlier known to be a major constituent of one of the most abundant cell wall proteins E6 asparagine-rich polypeptide expressed during cotton fiber elongation (John and Crow 1992). Up-regulation of this gene in Pima S-7 suggests a role in enhanced fiber development connected to nitrogen metabolism.

Processes over-represented only in the wild accession K101 include "subtilase activity," which comprises a group of serine proteases. These proteins have a wide range of functional activity, but some serine proteases, like subtilase, may lead to stress responses (Morinaga et al. 2007). In the same fashion, the "AU-specific RNA binding" process and the "SCF ubiquitin-ligase complex," shown earlier to promote mRNA degradation and cell-death (Ohme-Takagi et al. 1993) and promotion of PCD processes in plants (Gray et al. 1999; Devoto et al. 2002), are also up-regulated in K101. Down-regulating these proteins during fiber development in the cultivated accession Pima S-7 may be a key for prolonged fiber elongation. As shown in Fig. 5, one gene up-regulated at 2 dpa is the "pore-forming toxin-like protein Hfr-2," which RT-PCR results confirmed as up-regulated throughout development (Fig. 2G). Although the action of this gene in cotton has yet to be explored, it might play a negative role in fiber development in K101. In support of this hypothesis,

RT-PCR data on wild and cultivated accessions of *G. hirsutum* yields similar results, with higher expression of Hfr-2 in the wild accession, particularly later in development (Fig. 4E). It is conceivable that this represents an additional convergent or parallel response to domestication in *G. barbadense* and *G. hirsutum*.

A number of fiber-specific or fiber-enhanced genes have been characterized that are involved in secondary cell wall synthesis in cotton (Wilkins and Jernstedt 1999; Arpat et al. 2004). Two main categories are the cellulose synthase (*Ces*A1) and pectin-methylestrase (PME) gene families (Wilkins and Jernstedt 1999; Bosch et al. 2005). RNA expression levels of *Ces*A1 provide a good marker of secondary wall cell synthesis in cotton fibers (Arpat et al. 2004). Similarly, in *Nicotiana*, exogenous PME induces cell wall thickening and inhibits cell growth (Bosch 2005). In our study (Fig. 5), *Ces*A1 and PME are over-represented in K101 at 10 and 20 dpa relative to Pima S-7 during fiber development, suggesting a possible role in early cell wall maturation and potentially shortening the fiber elongation period in wild cotton relative to its domesticated derivative (Fig. 2H).

Conclusion

The present study implicates a diverse array of metabolic pathways and networks as being involved in the evolution of elongated epidermal seed trichomes, providing the foundation for later human domestication of an important crop plant. At least in part it appears that avoidance or delay of stress-like processes may underlie the increased elongation in Pima S-7 fiber development compared with K101 fiber, in conjunction with an increased ability to modulate cellular redox balance in the growing cell. Other processes involved in signal transduction and hormone signaling may be important regulators of prolonged fiber growth in Pima S-7, as might the down-regulation of cell wall maturation genes (cellulose synthase and pectinesterase) that could lead to growth inhibition in wild cotton. The work described here provides clues into the processes and genes that may have been selected by humans, starting with initial domestication of a wild perennial thousands of years ago, through the development of modern elite lines such as Pima S-7. This selection has radically altered metabolic pathways and flux, resulting in one of the world's most important fiber plants. Future functional analyses are required to test the many hypotheses generated here, and to explore their individual and aggregate effects on cellular physiology and ultimately, morphology. One striking aspect of the transcriptomic differences between wild and domesticated *G. barbadense* described here is that they appear to have parallels in a second domesticated cotton species, *G. hirsutum*. An exciting prospect will be to ultimately determine the underlying the genetic nature of these apparently overlapping, parallel, metabolic transformations.

Acknowledgments

The authors thank the National Science Foundation Plant Genome Program for financial support and the Department of Biotechnology, Government of India for a research fellowship to B. C. The authors also thank Lex Flagel for his help in data analysis and Alan Gingle for database management.

REFERENCES

Aida, M., and Tasaka, M. 2006. Morphogenesis and patterning at the organ boundaries in the higher plant shoot apex. *Plant Mol. Biol.* 60: 915–928.

Applequist, W. L., Cronn, R. C., and Wendel, J. F. 2001. Comparative development of fiber in wild and cultivated cotton. *Evol. Dev.* 3: 3–17.

Arpat, A. B., et al. 2004. Functional genomics of cell elongation in developing cotton fibers. *Plant Mol. Biol.* 54: 911–929.

Attucci, S., Carde, J. P., Raymond, P., Saint-Ges, V., Spiteri, A., and Pradet, A. 1991. Oxidative phosphorylation by mitochondria extracted from dry Sunflower seeds. *Plant Physiol.* 95: 390–398.

Basra, A., and Malik, C. 1984. Development of the cotton fiber. *Int. Rev. Cytol.* 89: 65–113.

Bosch, M., Cheung, A. Y., and Hepler, P. K. 2005. Pectin methylesterase, a regulator of pollen tube growth. *Plant Physiol.* 138: 1334–1346.

Brears, T., Liu, C., Knight, T. J., and Coruzzi, G. M. 1993. Ectopic overexpression of Asparagine synthetase in transgenic tobacco. *Plant Physiol.* 103: 1285–1290.

Brubaker, C. L., Koontz, J. A., and Wendel, J. F. 1993. Bidirectional cytoplasmic and nuclear introgression in the new world cottons, *Gossypium barbadense* and *G. hirsutum* (Malvaceae). *Am. J. Bot.* 80: 1203–1208.

Burger, J. C., Chapman, M. A., and Burke, J. M. 2008. Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95: 113–122.

Cernac, A., Andre, C., Hoffmann-Benning, S., and Benning, C. 2006. WRI1 is required for seed germination and seedling establishment. *Plant Physiol.* 141: 745–757.

Chico, J. M., Raices, M., Tellez-Inon, M. T., and Ulloa, R. M. 2002. A calcium-dependent protein kinase is systemically induced upon wounding in tomato plants. *Plant Physiol.* 128: 256–270.

Clark, R. M., Wagler, T. N., Quijada, P., and Doebley, J. F. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38: 594–597.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.

Cubas, P., Coen, E., and Martinez-Zapater, J. M. 2001. Ancient asymmetries in the evolution of flowers. *Curr. Biol.* 11: 1050–1052.

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. W. Clowes and Sons and Charing Cross Publisher, London.

Devoto, A., et al. 2002. COI1 links jasmonate signaling and fertility to the SCF ubiquitin-ligase complex in *Arabidopsis*. *Plant J.* 32: 457–466.

Dillehay, T. D., Rossen, J., Andres, T. C., and Williams, D. E. 2007. Preceramic adoption of peanut, squash, and cotton in northern Peru. *Science* 316: 1890–1893.

Doebley, J. 2004. The genetics of maize evolution. *Ann. Rev. Genet.* 38: 37–59.

Doebley, J. 2006. Unfallen grains: how ancient farmers turned weeds into crops. *Science* 312: 1318–1319.

Doebley, J., Stec, A., and Hubbard, L. 1997. The evolution of apical dominance in maize. *Nature* 386: 485–488.

Drew, M. C., Hea, C.-J., and Morgan, P. W. 2000. Programmed cell death and aerenchyma formation in roots. *Trends Pl. Sci.* 5: 123–127.

Ehrenreich, I. M., and Purugganan, M. D. 2006. The molecular genetic basis of plant adaptation. *Am. J. Bot.* 93: 953–962.

Foreman, J., et al. 2003. Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* 422: 442–446.

Friedrichsen, D. M., et al. 2002. Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* 62: 1445–1456.

Fry, S. C. 1998. Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem. J.* 332: 507–515.

Fryxell, P. A. 1979. *The Natural History of the Cotton Tribe*. Texas A&M University Press, College Station, TX.

Furuichi, T., Cunningham, K. W., and Muto, S. 2001. A putative two-pore channel *AtTPC1* mediates Ca^{2+} flux in *Arabidopsis* leaf cells. *Plant Cell Physiol.* 42: 900–905.

Gou, J.-Y., Wang, L.-J., Chen, S.-P., Hu, W.-L., and Chen, X.-Y. 2007. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res.* 17: 422–434.

Gray, W. M., et al. 1999. Identification of an SCF ubiquitin-ligase complex required for auxin response in *Arabidopsis thaliana*. *Genes Dev.* 13: 1678–1691.

Helliwell, C. A., Poole, A., Peacock, J. W., and Dennis, E. S. 1999. *Arabidopsis* ent-kaurene oxidase catalyzes three steps of gibberellin biosynthesis. *Plant Physiol.* 119: 507–510.

Herbert, R. J., et al. 2001. Ethylene induces cell death at particular phases of the cell cycle in the tobacco TBY-2 cell line. *J. Exp. Bot.* 52: 1615–1623.

Hoang, C. V., and Chapman, K. D. 2002. Regulation of carbonic anhydrase gene expression in cotyledons of cotton (*Gossypium hirsutum* L.) seedlings during post-germinative growth. *Plant Mol. Biol.* 49: 449–458.

Hovav, R., et al. 2008. The evolution of spinable cotton fiber entailed natural selection for prolonged development and a novel metabolism. *PloS Genet.* 4: e25.

Hovav, R., Udall, J. A., Hovav, E., Rapp, R. A., Flagel, L., and Wendel, J. F. 2007. A majority of genes are expressed in the single-celled seed trichome of cotton. *Planta* 227: 319–329.

Ji, S.-J., et al. 2003. Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res.* 31: 2534–2543.

John, M. E., and Crow, L. J. 1992. Gene expression in cotton (*Gossypium hirsutum* L.) fiber: cloning of the mRNAs. *Proc. Natl. Acad. Sci. USA* 89: 5769–5773.

Kellogg, E. A. 2004. Evolution of developmental traits. *Curr. Opin. Plant Biol.* 7: 92–98.

Kim, H. J., and Triplett, B. A. 2001. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127: 1361–1366.

Kim, W., Wan, C.-Y., and Wilkins, T. A. 1999. Functional complementation of yeast *vma1*Δ cells by a plant subunit A homolog rescues the mutant phenotype and partially restores vacuolar H^+-ATPase activity. *Plant J.* 17: 501–510.

Kobe, B., and Deisenhofer, J. 1994. The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* 19: 415–421.

Komatsuda, T., et al. 2007. Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc. Natl. Acad. Sci. USA* 104: 1424–1429.

Konishi, S., et al. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312: 1392–1396.

Kramer, E. M., and Irish, V. F. 1999. Evolution of genetic mechanisms controlling petal development. *Nature* 399: 144–148.

Lee, J. J., et al. 2006. Developmental and gene expression analyses of a cotton naked seed mutant. *Planta* 223: 418–432.

Li, C. B., Zhou, A. L., and Sang, T. 2006. Rice domestication by reducing shattering. *Science* 311: 1936–1939.

Li, H.-B., Qin, Y. M., Yu, P., Wen-Qiang, S., Wen-Qian, M., and Yu-Xian, Z. 2007. A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development. *New Phytol.* 75: 462–471.

Li, Y.-L., Sun, J., and Xia, G.-X. 2005. Cloning and characterization of a gene for an LRR receptor-like protein kinase associated with cotton fiber development Mol. *Genet. Genom.* 273: 217–224.

Liszkay, A., van der Zalm, E., and Schopfer, P. 2004. Production of reactive oxygen intermediates O_2^-, H_2O_2, and OH by maize roots and their role in wall loosening and elongation growth. *Plant Physiol.* 136: 3114–3123.

Liu, D., Zhang, X., Tu, L., Zhu, L., and Guo, X. 2006. Isolation by suppression-subtractive hybridization of genes preferentially expressed during early and late fiber development stages in cotton. *Mol. Biol.* 40: 741–749.

Moller, M. 2001. Plant mitochondria and oxidative stress: electron transport, NADPH turnover, and metabolism of reactive oxygen species. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 52: 561–591.

Morinaga, N., et al. 2007. Two distinct cytotoxic activities of subtilase cytotoxin produced by shiga-toxigenic *Escherichia coli*. *Infect. Immunity* 75: 488–496.

Morris, E. R., and Walker, J. C. 2003. Receptor-like protein kinases: the keys to response. *Curr. Opin. Plant Biol.* 6: 339–342.

Ni, W., Turley, R. B., and Trelease, R. N. 1990. Characterization of a cDNA encoding cottonseed catalase. *Biochem. Biophys. Acta* 1049: 219–222.

Ni, W., et al. 2004. Regulation of flower development in *Arabidopsis* by SCF complexes. *Plant Physiol.* 134: 1574–1585.

Ohme-Takagi, M., Taylor, C. B., Newman, T. C., and Green, P. J. 1993. The effect of sequences with high Au content on messenger-RNA stability in tobacco. *Proc. Natl. Acad. Sci. USA* 90: 11811–11815.

Orino, K., Lehman, L., Tsuji, Y., Ayaki, H., Torti, S. V., and Torti, F. M. 2001. Ferritin and the response to oxidative stress. *Biochem. J.* 357: 241–247.

Paterson, A. H. 2002. What has QTL mapping taught us about plant domestication? *New Phytol.* 154: 591–608.

Paterson, A. H., et al. 1995. Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* 269: 1714–1718.

Percy, R. G., and Wendel, J. F. 1990. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor. Appl. Genet.* 79: 529–542.

Pitzschke, A., and Hirt, H. 2006. Mitogen-activated protein kinases and reactive oxygen species signaling in plants. *Plant Physiol.* 141: 351–356.

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

Pottosin, I. I., and Schonknecht, G. 2007. Vacuolar calcium channels. *J. Exp. Bot.* 58: 1559–1569.

Puthoff, D. P., Sardesai, N., Subramanyam, S., Nemacheck, J. A., and Williams, C. E. 2005. Hfr-2, a wheat cytolytic toxin-like gene, is upregulated by virulent Hessian fly larval feeding. *Mol. Plant Pathol.* 6: 411–423.

Rahman, A., Amakawa, T., Goto, N., and Tsurumi, S. 2001. Auxin is a positive regulator for ethylene-mediated response in the growth of *Arabidopsis* roots. *Plant Cell Physiol.* 42: 301–307.

Rentel, M. C., and Knight, M. R. 2004. Oxidative stress-induced calcium signaling in *Arabidopsis*. *Plant Physiol.* 135: 1471–1479.

Rodriguez, A. A., Grunberg, K. A., and Taleisnik, E. L. 2002. Reactive oxygen species in the elongation zone of maize leaves are necessary for leaf extension. *Plant Physiol* 129: 1627–1632.

Ruan, Y.-L., Llewellyn, D. J., and Furbank, R. T. 2001. The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K^+ transporters and expansin. *Plant Cell* 13: 47–60.

Schopfer, P. 1996. Hydrogen peroxide-mediated cell-wall stiffening in vitro in maize coleoptiles. *Planta* 199: 43–49.

Seelanan, T., Schnabel, A., and Wendel, J. F. 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Syst. Bot.* 22: 259–290.

Shi, Y. H., et al. 2006. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18: 651–664.

Shiu, S.-H., and Bleecker, A. B. 2001. Plant receptor-like kinase gene family: diversity, function, and signaling. *Science* 113: 22.

Simons, K. J., et al. 2006. Molecular characterization of the major wheat domestication gene Q. *Genetics* 172: 547–555.

Sliwinski, M. K., Bosch, J. A., Yoon, H.-S., Balthazar, M. V., and Baum, D. A. 2007. The role of two LEAFY paralogs from *Idahoa scapigera* (Brassicaceae) in the evolution of a derived plant architecture. *Plant J.* 51: 211–219.

Storey, J. D., and Tibshirani, R. 2003. SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In G. Parmigiani, E. S. Garrett, R. A. Irizarry, and S. L. Zeger (eds) *The Analysis of Gene Expression Data: Methods and Software*. Springer, New York, pp. 272–290.

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. 2006. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18: 283–294.

Taliercio, E. W., and Boykin, D. 2007. Analysis of gene expression in cotton fiber initials. *BMC Plant Biol*. 7–22.

Tanaka, R., and Tanaka, A. 2007. Tetrapyrrole biosynthesis in higher plants. *Ann. Rev. Plant Biol*. 58: 321–346.

Thuleau, P., Ward, J. M., Ranjeva, R., and Schroeder, J. I. 1994. Voltage-dependent calcium-permeable channels in the plasma-membrane of a higher-plant cell. *EMBO J*. 13: 2970–2975.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Udall, J. A., et al. 2006. A global assembly of cotton ESTs. *Genome Res*. 16: 441–450.

Udall, J. A., et al. 2007. Spotted cotton oligonucleotide microarrays for gene expression analysis. *BMC Genomics* 8: 81.

Velleman, P. F., and Hoaglin, D. C. 1981. *The ABC of Exploratory Data Analysis*. Duxbury Press, Belmont, CA.

Vollbrecht, E., Springer, P. S., Goh, L., Buckler, E. S., and Martienssen, R. 2005. Architecture of floral branch systems in maize and related grasses. *Nature* 436: 1119–1126.

Wahlroos, T., Soukka, J., Denesyuk, A., Wahlroos, R., Korpela, T., and Kilby, N. J. 2003. Oleosin expression and trafficking during oil body biogenesis in tobacco leaf cells. *Genesis* 35: 125–132.

Wan, C., and Wilkins, T. 1994. A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Annal. Biochem*. 223: 7–12.

Wang, H., et al. 2005. The origin of the naked grains of maize. *Nature* 436: 714–719.

Wasternack, C. 2007. Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Annal. Bot*. 100: 681–697.

Wendel, J. F. 1995. Cotton. In N. Simmonds and J. Smartt (eds) *Evolution of Crop Plants*. Longman, London, pp. 358–366.

Wendel, J. F., and Albert, V. A. 1992. Phylogenetics of the cotton genus *Gossypium* - character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot*. 17: 115–143.

Wendel, J. F., and Cronn, R. C. 2003. Polyploidy and the evolutionary history of cotton. *Adv. Agron*. 78: 139–186.

Westengen, O. T., Huamán, Z., and Heun, M. 2005. Genetic diversity and geographic pattern in early South American cotton domestication. *Theor. Appl. Genet*. 110: 392–402.

Whittaker, D. J., and Triplett, B. A. 1999. Gene-specific changes in alpha-tubulin transcript accumulation in developing cotton fibers. *Plant Physiol*. 121: 181–188.

Wilkins, T., and Jernstedt, J. 1999. Molecular genetics of developing cotton fibers. In A. S. Basra (ed) *Cotton Fibres: Developmental Biology, Quality Improvement, and Textile Processing*. The Haworth Press, Inc, NY, pp. 231–269.

Wu, Y., Llewellyn, D., White, R., Ruggiero, K., Al-Ghazi, Y., and Dennis, E. 2007. Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibre initial cells on the surface of cotton ovules. *Planta* 226: 1475–1490.

Yang Samuel, S., et al. 2006. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J*. 47: 761–775.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1. Primer used in this study for Real-Time PCR. F = Forward primer, R = Reverse primer.

Table S2. List of common biological processes from co-expressed genes in K101 and Pima S-7 during fiber development ($P < 0.05$ and FDR < 0.01, > 2 fold expression change).

Please note: Wiley Periodicals, Inc. are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.