Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*)

Ying Bao[a], Guanjing Hu[b], Lex E. Flagel[c], Armel Salmon[d], Magdalena Bezanilla[e], Andrew H. Paterson[f], Zining Wang[f], and Jonathan F. Wendel[b,1]

[a]School of Life Science, Qufu Normal University, Qufu, Shandong 273165, China; [b]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011; [c]Department of Biology, Duke University, Durham, NC 27708; [d]Unité Mixte de Recherche, Centre National de la Recherche Scientifique, Université de Rennes 1, 35 042 Rennes, France; [e]Biology Department, University of Massachusetts, Amherst, MA 01003-9297; and [f]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602

Cotton is remarkable among our major crops in that four species were independently domesticated, two allopolyploids and two diploids. In each case thousands of years of human selection transformed sparsely flowering, perennial shrubs into highly productive crops with seeds bearing the vastly elongated and abundant single-celled hairs that comprise modern cotton fiber. The genetic underpinnings of these transformations are largely unknown, but comparative gene expression profiling experiments have demonstrated up-regulation of profilin accompanying domestication in all three species for which wild forms are known. Profilins are actin monomer binding proteins that are important in cytoskeletal dynamics and in cotton fiber elongation. We show that *Gossypium* diploids contain six profilin genes (*GPRF*1–*GPRF*6), located on four different chromosomes (eight chromosomes in the allopolyploid). All but one profilin (*GPRF*6) are expressed during cotton fiber development, and both homeologs of *GPRF*1–*GPRF*5 are expressed in fibers of the allopolyploids. Remarkably, quantitative RT-PCR and RNAseq data demonstrate that *GPRF*1–*GPRF*5 are all up-regulated, in parallel, in the three independently domesticated cottons in comparison with their wild counterparts. This result was additionally supported by iTRAQ proteomic data. In the allopolyploids, there This usage of novel should be fine, since it refers to a novel evolutionary process, not a novel discovery has been novel recruitment of the sixth profilin gene (*GPRF*6) as a result of domestication. This parallel up-regulation of an entire gene family in multiple species in response to strong directional selection is without precedent and suggests unwitting selection on one or more upstream transcription factors or other proteins that coordinately exercise control over profilin expression.

polyploidy | crop evolution | fiber | directional selection

With its highly exaggerated, unicellular seed hairs (trichomes), cotton provides the world's most important source of renewable natural fiber. The origin of cultivated cotton is notable among our major crop plants in that four *Gossypium* species were independently domesticated, in different geographic regions, by separate prehistoric cultures (1). This parallel domestication process involved two allopolyploid (AD genome) species from the New World, *Gossypium hirsutum* (the source of "Upland" cotton) and *Gossypium barbadense* (the source of "Pima" or "Egyptian" cotton), and two diploid (A genome) species from the Old World, *Gossypium arboreum* and *Gossypium herbaceum*. As a consequence of thousands of years of human-mediated selection and agronomic improvement, each of the cultivated species underwent a series of phenotypic modifications, including transformation from perennial shrubs and small trees to more compact annual plants, loss of photoperiod sensitivity, reduction in seed dormancy, and most importantly, morphological transitions in fibers that dramatically enhanced fiber yield and quality. At present, the genetic bases of these possibly contemporaneous transformations are unknown.

In contrast to the wild progenitors, which have relatively short and coarse, tan-colored fibers, modern cultivars possess much longer, finer, stronger, and whiter fiber (Fig. 1). This similarity among the end-products of strong human directional selection raises a general question about the level and details of the processes governing the evolutionary transformations from wild to domesticated forms. It is of interest to understand the nature of this evolutionary parallelism at various levels of biological organization, from morphology to genotype. With respect to the former, analyses of fiber growth curves reveal that cotton domestication at both the diploid and allopolyploid levels is associated with a prolonged period of fiber elongation (2). As to the genomic underpinnings of fiber evolution, a number of recent studies are generating insights and a rich data set based on comparative transcriptome profiling of fiber development in wild vs. domesticated cotton (3–6), and from ongoing comparative proteomic analyses (7). Rapp et al. (2010), for example, reported thousands of genes that are differentially expressed in developing fiber between wild and domesticated *G. hirsutum*, many associated with key developmental processes presumed to play important roles in primary and secondary wall synthesis and in modulation of reactive oxygen species (3, 5, 6). Thus, the evolutionary process associated with strong directional selection practiced by aboriginal domesticators appears to be extraordinarily complex, at least insofar as perturbation of the transcriptional network, especially in the early stages of fiber development (4).

Because fiber length per se is an agronomically important property, much attention has been focused on the molecular mechanisms of fiber elongation (8–10). In plants, a highly dynamic actin cytoskeleton has emerged as a key determinant of cell morphology. Numerous studies have demonstrated that dynamic actin is essential for cell expansion in tip-growing cells, such as root hairs, pollen tubes, and the protonemata of mosses (11–13). Furthermore, actin is also required for expansion of trichomes (14, 15), a category that includes cotton "fiber."

The dynamic rearrangement of actin filaments involves maintaining a proper balance between filamentous and monomeric actin, which is controlled by a number of actin-binding proteins. Profilin is one of the most abundant actin monomer binding
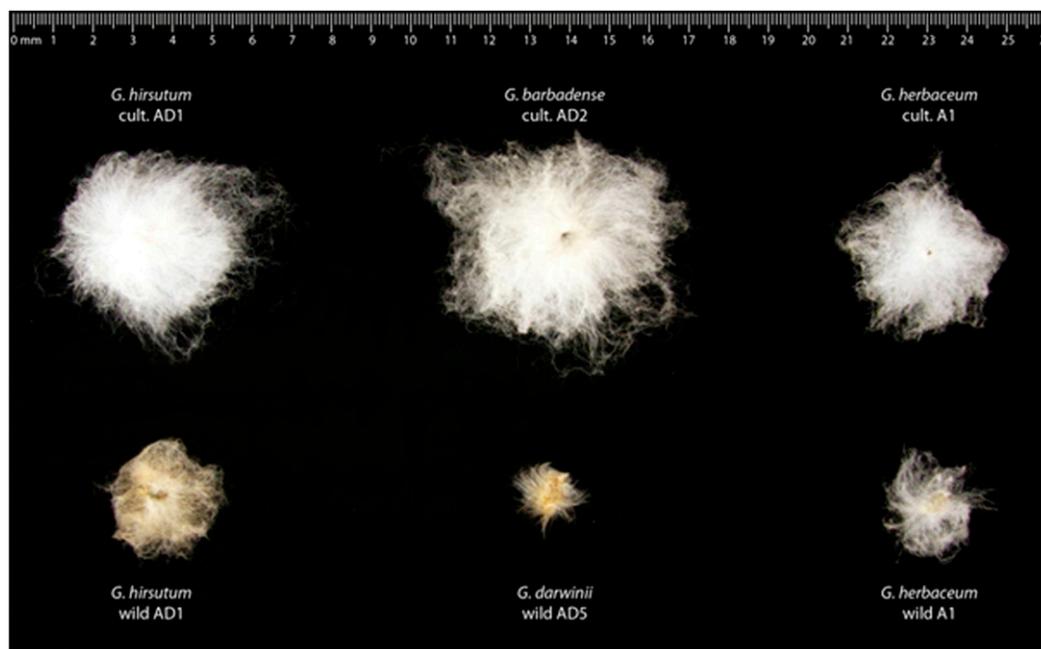
Fig. 1. Morphological differences between cultivated cottons (from *Left* to *Right*: *G. hirsutum*, *G. barbadense*, and *G. herbaceum*) and their wild counterparts (from *Left* to *Right*: *G. hirsutum*, *G. darwinii*, and *G. herbaceum*).

proteins and has been extensively characterized in vitro and in vivo (11, 16, 17). The role of profilin in regulating actin dynamics may be to help maintain a pool of actin monomers competent to polymerize at the fast-growing end of the actin filament (18–21). This is achieved, in part, by the interaction of profilin with the formin class of actin nucleators to promote actin polymerization and elongation (22–24). However, profilin also binds to phosphoinositides and other proteins (25, 26), suggesting that it plays a broad role in modulating the actin cytoskeleton.

Profilin has been shown to affect tip growth in moss, root hairs, pollen tubes, and cotton fiber (15, 27–29). Profilin is expressed during early cotton fiber development (14, 15), and on the basis of the observation that overexpression in transgenic tobacco cells produced elongated cells with thicker and longer microfilament cables, Wang et al. (15) suggested that profilin plays a role in the rapid elongation of cotton fibers by promoting actin polymerization.

In previous investigations of cotton domestication based on microarray data, profilin genes ranked among the most highly differentially expressed genes (4, 14, 30), suggesting that they, or one or more upstream regulators, were targets of human selection. Moreover, this observation was repeated for three different, independently domesticated cotton species. Here, we study this evolutionary parallelism in more detail, describing the cotton profilin gene family and its expression in fiber collected from all three of the domesticated species for which wild forms are known. A total of six profilin gene members (12 homeologs) were identified in *Gossypium*, and all but one of these are expressed during cotton fiber development. Remarkably, all five of the genes expressed in fiber are up-regulated in all three independently domesticated cottons in comparison with their wild counterparts, a result congruent with proteomic data. This parallel up-regulation of a gene family, instead of a single gene, is without precedent and suggests commonalities in the upstream targets of human-mediated directional selection among three species of domesticated cotton.

Results

Characterization of the Cotton Profilin Gene Family.
To understand the gene family structure of profilin in *Gossypium*, we first

analyzed profilin genes in other plants. Profilin is a protein of 130–134 aa encoded by a small gene family. Plant profilins fall into two classes, one thought to be constitutively expressed and the other predominantly expressed in anthers or pollen (31–34). The size of the profilin gene family varies among plants. Only one profilin gene is annotated in algae (*Chlamydomonas reinhardtii* and *Volvox carteri*), whereas 2–14 gene members have been detected in land plants. No length variation among genes is revealed in *Physcomitrella* (132 aa for four genes), *Selaginella* (132 aa for two genes), or monocots (131 aa for three genes in *Oryza* and five genes in *Zea*), but most eudicot genomes encode profilin proteins of varying lengths. In *Arabidopsis*, for example, the five profilin genes encode protein products with either 131 or 134 aa, the length difference reflecting a three-codon indel in the first exon. At the same position, a two-codon indel is apparent in other eudicots (such as *Ambrosia*, *Carica*, *Glycine*, *Gossypium*, *Lotus*, *Manihot*, *Medicago*, *Ricinus*, and *Vitis*).

PCR amplification followed by cloning and sequencing in *Gossypium* led to the identification of six profilin gene family members (*GPRF1*–*GPRF6*) in all diploid cotton species surveyed. For each of these six genes in the allopolyploids, both copies from the two resident genomes were recovered, as expected; thus, a pair of homeologs exists for each of *GPRF1*–*GPRF6*, denoted by subscripts representing their genome of origin (A_T deriving from the A-genome progenitor and D_T deriving from the D-genome progenitor).

All six profilin genes share a similar structure, including three exons and two introns (Fig. S1). Gene lengths vary widely; although from 569 to 1623 bp, this variability is due mostly to intron size variation. Both introns of *GPRF1* and *GPRF6* are much shorter (65–96 bp) than introns of the other four profilins (382–633 bp); the longest intron (633 bp), in *GPRF3*, is almost 10-fold longer than the shortest one (65 bp in *GPRF6*). Interestingly, this level of variation in profilin intron sizes has not been detected in other plants. Alignment of protein sequences (Fig. S1) revealed that a 6-bp indel in the first exon is responsible for the longer protein products of *GPRF1* and *GPRF6* (134 aa) than of the other four genes (132 aa). As described above, this indel was also observed in other eudicots, suggesting that the co-occurrence

of long and short profilin types originated early in eudicot evolution. Amino acids required for interaction with actin (35), poly-L-prolines (36) and PI(4,5)P2 (37) are conserved in the six *Gossypium* profilin genes.

GPRF1 and *GPRF6* are more similar to each other than they are to the remaining four genes, structurally and with respect to composition, although they differ significantly at the nucleotide level (e.g., $K_a = 0.075$; $K_s = 0.401$ for the A-genome paralogs) (Table S1). Intergenic comparisons within the A- and D-genome diploids show that all six genes are divergent from one another at the nucleotide level, although they retain 75–95% amino acid identity (Table S1). Nonsynonymous differences (K_a) among the six profilin genes vary 8- (A genome) to 10-fold (D genome) within a species. *GPRF2* and *GPRF5* have higher amino acid identity to one another than either does to *GPRF3* and *GPRF4*, with the latter gene being the most divergent of this group of four profilin genes containing long introns. Paralleling this distinction is the observation that the *GPRF4* introns are substantially shorter than those of *GPRF2*, *GPRF3*, and *GPRF5*. Synonymous substitution rates (K_s) between A and D orthologs fall within the range typically observed in cotton (38), ranging from rather slowly evolving genes (*GPRF2* and *GPRF5*; $K_s = 0.015$) to faster-evolving genes (*GPRF3*, $K_s = 0.099$). Sequence divergence is similar between the parental diploid species and their respective genomes in the allopolyploids; that is, orthologs from diploids (A vs. D) have K_s values similar to those of homeologs (A_T vs. D_T) in the allopolyploids, consistent with previous observations in *Gossypium* (38).

To investigate the history of gene duplication events that led to the observed profilin gene family structure, we constructed a maximum likelihood tree of profilins from sequenced plant genomes and data from other plants (Fig. S2). Rooted with the green alga sequences (*Chlamydomonas* and *Volvox*), sequences partitioned into two major clades (I and II, Fig. S2), a pattern suggestive of a duplication event early in angiosperm evolution. Six *Gossypium* genes, four long intron, 132 aa (i.e., *GPRF2*–*GPRF5*), and two short intron, 134 aa (*GPRF1* and *GPRF6*), fell into these same two clades, as did sequences from other eudicots and monocots. These data indicate that the two major clades of profilins originated from a duplication event before the divergence of monocots and eudicots. For *Gossypium*, phylogenetic analysis yielded the topology expected, with no evidence of gene conversion (39).

We mapped all of the profilin genes on the *Gossypium raimondii* genome sequence and linked these to the tetraploid map (Table S2). The mapping results showed that we identified all genomic copies, and that the six profilin genes (*GPRF1*–*GPRF6*) are located on four different chromosomes (eight in the allopolyploid, i.e., both homeologs) (Table S2). Notably, tandem pairs of *Gossypium* profilin genes, each containing one gene from the two basal clades (I and II) are located on two different chromosomes (9 and 6), indicating a second round of gene duplication.

Profilin Expression in Cotton Fiber. Reverse transcriptase PCR (RT-PCR) analysis indicated that *GPRF6* was not transcribed in fiber from any *Gossypium* species studied. Additionally, *GPRF1* was not detected in fibers from the D-genome diploid, but was expressed both in the A genome and the allopolyploid species (Fig. S3). In all allopolyploids, both homeologs of *GPRF1* to *GPRF5* (10 sequences total) were identified by amplification, cloning, and Sanger sequencing and through Illumina "RNA-Seq" transcriptome sequencing.

Gene-specific (although not homeolog-specific) primers for *GPRF1*–*GPRF5* were used to assess relative expression levels of profilins between wild and domesticated cottons using real-time, quantitative RT-PCR. Cultivated accessions of both diploid and allopolyploid species exhibited up-regulation relative to their wild counterparts of not just one profilin gene but all five genes

(*GPRF1*–*GPRF5*) in each of the three species studied (Fig. 2). The magnitude of up-regulation varied among genes and species, ranging from 2- to 106-fold, 11- to 390-fold, and 1.5- to 5.2-fold in *G. barbadense* (Pima S-6 vs. PW45), *G. hirsutum* (TM1 vs. Tx2094), and *G. herbaceum* (Wagad vs. A1-73), respectively (Fig. 2 A–C). Thus, there was far greater profilin up-regulation in the allopolyploids than in the diploid. We estimated the proportional expression of each of the five profilin genes in each sample (Fig. 2D). In *G. barbadense* and its wild relative *Gossypium darwinii*, the highest expression was found for *GPRF3* (28–35% of total), whereas *GPRF2* and *GPRF5* were the most highly expressed profilins in cultivated and wild *G. hirsutum* (29% of total). In *G. herbaceum*, *GPRF5* is the most highly expressed gene (~25% of total).

Fiber transcriptome RNAseq datasets yield data (Table 1) congruent with the results from quantitative RT-PCR: (*i*) Expression of *GPRF6* was not detected in fibers; (*ii*) *GPRF1* was expressed at low levels, especially in the wild forms; and (*iii*) convergent up-regulation of profilin genes was observed in domesticated cottons, with the sole exception that *GPRF5* was more highly expressed in *G. barbadense* accession K101 than in Pima S-6. *GPRF1* was expressed at a low level compared with the other profilins, but still was up-regulated by domestication (3.9-fold and 420-fold) in *G. barbadense* and *G. hirsutum*, respectively (Fig. 3 and Table 1).

In addition to estimating expression using transcriptomic estimates, we also estimated profilin expression during fiber development using iTRAQ proteomic analysis, which allows simultaneous protein identification and comparative quantification of multiple samples. From the *G. hirsutum* fiber proteomic data, peptides corresponding to profilin gene family members were detected. Among those, gene member-specific peptides were identified for *GPRF1* and *GPRF4*, respectively, which allowed the expression analysis for these two profilin genes. For *GPRF2*, *GPRF3*, and *GPRF5*, only peptides corresponding to their conserved region were detected, which were used to measure the combined expression of the three proteins as *GPRFX*. As shown in Table 2, all numbers are greater than 1, indicating greater protein expression in fibers from domesticated than in wild *G. hirsutum*. Ratios varied among stages and genes, from near parity to a 4.6-fold increase. Accordingly, and also because of high variance in some cases among different peptides of the same protein, statistical significance was not obtained in every

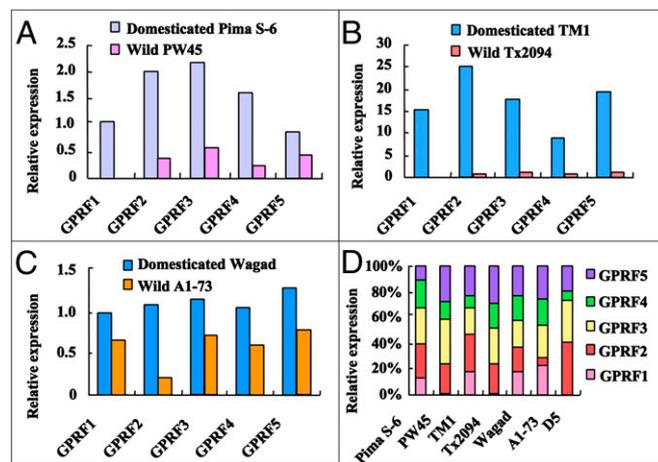**Fig. 2.** Quantitative RT-PCR estimates of expression for five profilin genes in cotton fiber. (*A*) Comparison between domesticated *G. barbadense* (Pima S-6) and its wild counterpart, *G. darwinii* (PW45). (*B*) Comparison between domesticated (TM1) and wild (Tx2094) *G. hirsutum*. (*C*) Comparison between domesticated (Wagad) and wild (A1-73) *G. herbaceum*. (*D*) Relative expression of each gene in each accession.

Table 1. Count reads and ratios for profilin genes of domesticated and wild cottons at 10-dpa fiber cDNA libraries

	Profilin gene	Domesticated count	Wild count	Total domesticated	Total wild	Ratio of domesticated to wild
Maxxa vs. Tx2094	GPRF1	448	1	8,343,842	7,823,250	420.0482*
	GPRF2	4,391	2,564	8,343,842	7,823,250	1.6057*
	GPRF3	4,630	2,340	8,343,842	7,823,250	1.8552*
	GPRF4	3,405	1,793	8,343,842	7,823,250	1.7806*
	GPRF5	1,446	1,065	8,343,842	7,823,250	1.2730*
Pima-6 vs. K101	GPRF1	439	129	4,863,645	5,515,575	3.8593*
	GPRF2	3,728	3,487	4,863,645	5,515,575	1.2124*
	GPRF3	2,952	3,032	4,863,645	5,515,575	1.1041*
	GPRF4	3,417	3,528	4,863,645	5,515,575	1.0984*
	GPRF5	700	1,505	4,863,645	5,515,575	0.5275*

Ratio of domesticated to wild = (domesticated count/total domesticated count)/(wild count/total wild count).
*P value <0.01 (Fisher's exact test).

comparison. Notably, and with the exception of GPRF1, for which only a single read was recorded in the RNAseq data in wild cotton (leading to a spuriously high domesticated-to-wild transcript ratio), these protein expression ratios closely mirror the transcript expression data (proteomic and transcript data are positively correlated; $r = 0.999$, $P < 0.01$).

Transcript Levels of Other Proteins Involved in Actin Cytoskeleton Remodeling. We examined the RNAseq data for expression of other proteins implicated in regulating actin or actin-mediated processes. Contigs detected are listed in Table S3, and total expression for several genes is shown for *G. hirsutum* in Fig. S4. The expression of actin was the highest in both 10-d postanthesis (dpa) and 20-dpa fibers of *G. hirsutum*, followed by profilin, and then actin-depolymerizing factor (ADF), whereas cyclase-associated protein (CAP) was expressed at relatively lower levels.

Discussion

Profilin Gene Family in *Gossypium*. Here we describe six profilin genes (GPRF1–GPRF6) from five species of *Gossypium*; four of these are newly recognized, but GPRF1 and GPRF3 were reported earlier from *G. hirsutum* (14, 15, 30, 40). All six cotton profilin genes in all species share high amino acid sequence identity (75–95%), gene structures (two introns and three exons), and conserved motifs, but fall into two distinct and ancient clades. In one clade (containing GPRF2–GPRF5), the first exon is missing two amino acids and has longer introns, whereas in the other clade (containing GPRF1 and GPRF6) the first exon is longer but the genes have shorter introns. Phylogenetic analysis further shows that these two classes of profilins are descendants of an ancient duplication that occurred early in angiosperm evolution, perhaps before the separation between eudicots and monocots (Fig. S2).
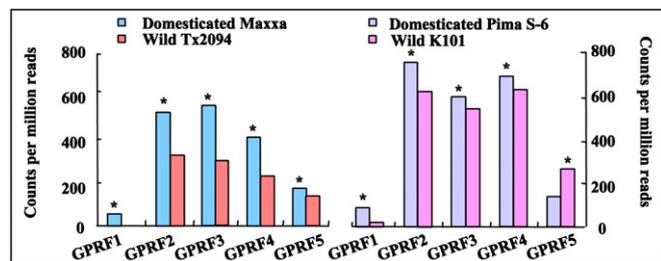

Fig. 3. Relative expression counts for five profilin genes based on ~16 million Illumina reads per accession. Maxxa and Tx2094 are cultivated and wild forms of *G. hirsutum*, whereas Pima S-6 and K101 are cultivated and primitive accessions of *G. barbadense*, respectively. The y axis is expressed in gene counts per million reads, and (*) represents significant expression divergence between domesticated and wild accessions (Fisher's exact test; $P < 0.01$).

Plant profilins generally are classified into two groups by their expression patterns. Kandasamy et al. (41) characterized the regulation of five *Arabidopsis* profilins in different organs and during microspore development, reporting that PRF1, PRF2, and PRF3 are expressed constitutively in all organs, whereas PRF4 and PRF5 are expressed specifically in pollen. As shown here, these two classes of *Arabidopsis* profilin genes fall into two clades, grouping with different cotton profilins. PRF1–PRF3 are grouped with GPRF2–GPRF5, whereas PRF4 and PRF5 are grouped with GPRF1 and GPRF6. This topological structure implies a possible functional correlation between these classes in *Arabidopsis* and *Gossypium*. Recently, Wang et al. (30) identified a profilin homolog, GhPFN2 (same as GPRF3 in the present study), from a cultivated variety of *G. hirsutum*. They found that GhPFN2 was expressed constitutively in multiple organs and preferentially in fiber cells. Expression was significantly induced during the period of rapid fiber elongation and secondary wall synthesis. Our expression analysis reveals that this class of four profilin genes, including GPRF2–GPRF5, are all expressed in fibers from three different cotton species.

The second class of cotton profilin genes (GPRF1 and GPRF6) differs from the other cotton profilins in expression. RT-PCR and RNAseq results show that GPRF6 is not transcribed in 10-dpa fibers. Similarly, with the exception of *G. herbaceum*, GPRF1 is either not expressed or is weakly expressed in developing fibers from wild cottons, although it is up-regulated in the domesticated allopolyploids (Figs. 2 and 3). Combined with the phylogenetic analyses, which indicate a homology between these two cotton genes and the pollen-specific genes of *Arabidopsis*, we speculate that little to no expression in fibers is ancestral for this group of profilins, and this is largely reflected in wild cottons. However, Ji et al. (14) recovered a profilin gene that was significantly up-regulated early in cultivated cotton (*G. hirsutum*) fiber development that corresponds to our GPRF1. This gene was later shown (15) to be predominantly expressed in rapidly elongating cotton fibers and that transcript abundance declined sharply with the onset of secondary wall synthesis. Our data indicate that GPRF1 has become associated with fiber cell elongation in cultivated but not wild cotton, suggesting novel gene recruitment associated with domestication.

Table 2. Protein ratios of profilins in fiber extracts from domesticated and wild *Gossypium hirsutum*

	5 dpa	10 dpa	20 dpa	25 dpa
GPRF1	1.1323	2.4878*	4.6008*	2.0378
GPRFX	1.1559	1.0140	1.7368	1.4000
GPRF4	1.2253	1.1013	1.3937**	1.1334

Dpa, days postanthesis.
*P value <0.05; **P value <0.01 (Student's t test).

EVOLUTION

Expression Alteration of Other Cytoskeletal Proteins. Actin binding proteins regulate the dynamics of the actin cytoskeleton by controlling the balance between monomeric, filamentous, and bundled actin. On the basis of RNAseq data derived from on-going studies of the fiber transcriptome in wild and domesticated cotton, several proteins involved in actin cytoskeleton dynamics demonstrated up-regulation in parallel with the profilins, including actin, ADF, and CAP (Fig. S4). Interestingly, profilin, ADF, and CAP are important for regulating the concentration of actin monomers. These data suggest that upon up-regulation of actin, control of the actin monomer pool is critical for the dramatic expansion in cotton fibers in domesticated cotton.

Parallel Up-Regulation of a Gene Family Under Domestication. As a specialized unicellular trichome with a greatly exaggerated length, cotton fiber represents a masterpiece of human domestication, made all the more remarkable by its parallel, independent origin in four cultivated species (three studied here). Little is known about the genetic, genomic, or metabolomic transformations that mediate these independent transformations, although insights are emerging from comparative expression profiling experiments (3–6). The present study sheds light on one aspect of this general question, implicating the up-regulation of the profilin gene family concomitant with strong directional selection under human domestication. Not only was the same protein family up-regulated by aboriginal domesticators in multiple species, but the effects have been widespread across the profilin gene family members, as opposed to affecting a single gene in each domesticate. Gene expression data revealed that expression of five genes was significantly enhanced in all cultivated cottons in comparison with their wild counterparts, including novel gene recruitment associated with domestication.

To our knowledge this observation of up-regulation of a gene family is without precedent in evolutionary biology, let alone parallel up-regulation in multiple species. Insights into the genetic basis of morphological change in nature are often facilitated using crop models, as Darwin famously noted in the introduction to *On the Origin of Species* when he wrote "*At the commencement of my observations it seemed to me probable that a careful study of domesticated animals and of cultivated plants would offer the best chance of making out this obscure problem.*" Since Darwin's time, numerous mutations have been identified that control traits selected by humans during the domestication process, including loss-of-function alleles, changes in coding sequence, or altered levels or domains of expression (42–46). Notwithstanding these striking discoveries, relatively little is understood about the downstream effects on domestication mutations or transcriptional and physiological networks, nor how these are propagated into the phenotypes being subjected to directional selection. Here we have illustrated one likely dimension of this process involved the altered regulation of a suite of related proteins, which simultaneously and in parallel in multiple species become up-regulated.

An exciting prospect for future work will be to isolate the causative lesions responsible for profilin gene family up-regulation, which likely comprise the hidden targets of human selection. In principle, these unknown targets of selection likely comprise one or more upstream transcription factors or other proteins that coordinately exercise control over profilin expression. It will be of considerable interest to reveal the degree of parallelism experienced in each species (47) and the effects of each species-specific mutation (or mutations) on the underlying transcriptional, proteomic, and metabolomic architecture of cotton fiber development. Insights into these and related questions will likely derive from a combination of forward genetic (e.g., introgression lines), population genetic (e.g., testing for selective sweeps), and genomic (e.g., expression profiling) approaches.

A final aspect of our results that merits highlighting is the extent to which the different profilin genes were up-regulated in each domesticated species. For example, in cultivated *G. hirsutum*, each of the five genes *GPRF*1–*GPRF*5 contributes substantially and relatively equitably to the total profilin transcriptome (18, 29, 21, 10, and 22%, respectively), whereas in the wild form of the same species, there is more variation among genes (1, 23, 28, 19, and 29%, respectively; Fig. 2*D*). The relative proportions and responses to selection in *G. barbadense* and *G. herbaceum* each present rather different patterns from this. When more is understood about the causative lesions in each species, it will be of considerable interest to study the complex rewiring of the transcriptional network that has responded similarly, but differently among these three species.

Materials and Methods

More detailed descriptions of all methods are provided in *SI Materials and Methods*. We used one domesticated and one wild accession for each of the three cotton species. The modern domesticated lines chosen were Pima S-6 (elite cultivar of *G. barbadense*), Texas Marker Stock 1 (TM1, the genetic and cytogenetic standard of *G. hirsutum*), and Wagad (an Indian cultivar of *G. herbaceum*). Choice of wild accessions was based on previous morphological and molecular evidence (48): for *G. barbadense*, because truly wild rather than feral forms are difficult to verify, its sister taxon from the Galapagos Islands *G. darwinii* (accession PW45) was chosen, as it previously was treated as conspecific with *G. barbadense* (49); for *G. hirsutum*, an unambiguously wild, sprawling shrub from the north coast of the Yucatan Peninsula, var. *yucatanense* accession Tx2094 (US Department of Agriculture GRIN accession PI 501501), was used; and for *G. herbaceum*, we chose a wild form from Botswana, *G. herbaceum* subsp. *africanum* (accession A1-73). We also included the best living model of the D-genome diploid progenitor, *G. raimondii*.

Using profilin genes from GenBank and *Gossypium* as query sequences, we searched our EST assemblies to design degenerate PCR primers (Table S4) to amplify profilin genes from *Gossypium*. PCR reactions were performed as described (50), and amplicons were cloned, sequenced (deposited in GenBank under accession numbers HM484221–HM484292), and mapped as described (*SI Materials and Methods*). Protein annotation and identification of conserved domains were facilitated using the Conserved Domains Database (51). We conducted phylogenetic analysis using species with available genome sequence, including 10 eudicot species, two monocots, *Selaginella moellendorffii* and *Physcomitrella patens*, and using two algae as the outgroup. Additionally, 66 plant profilins with annotations of sequence and functional information were also included in the analysis. Sequences were aligned using Jalview 2.5.1 (52). Maximum likelihood analysis was conducted using the default option as implemented in PhyML_3.0 (53). Confidence of the tree topology was assessed by a bootstrap set of 1,000 replicates. Estimation of nonsynonymous (K_a) and synonymous (K_s) substitution rates was performed within and between *Gossypium* species using DnaSP version 5 (54).

Fibers were harvested at 10 dpa because microarray data from *G. hirsutum* (4) showed that the expression level of profilin transcripts peaks at this stage. Total RNA was extracted as described (55). Profilin cDNA sequences were submitted to GenBank (HM543080–HM543138). To estimate transcript accumulation levels, we used quantitative real time RT-PCR analyses and Illumina RNAseq data. For the latter, we used fiber transcriptome data from 10- and 20-dpa fibers from wild and domesticated representatives of both domesticated polyploids (National Center for Biotechnology Information Sequence Read Archive Study SRP001603). We estimated expression levels for other actin binding proteins involved in actin cytoskeleton dynamics, including actin, ADF, and CAP using published cotton and *Arabidopsis* genes as queries and RNAseq data.

Total proteins were extracted from developing cotton fibers (56) using a liquid nitrogen/glass beads shearing method (57). Isolated proteins from wild and domesticated *G. hirsutum* were subjected to a comparative proteomic analysis using isobaric tags for relative and absolute quantification (iTRAQ) followed by strong-cation exchange fractionation and tandem mass spectrometry (10). The resulting mass spectrometry data from three replicates were processed and statistically analyzed using the ProteinPilot 4.0 software suite (AB SCIEX).

1. Wendel J, Cronn R (2003) Polyploidy and the evolutionary history of cotton. *Adv Agron* 78:139–186.
2. Applequist WL, Cronn R, Wendel JF (2001) Comparative development of fiber in wild and cultivated cotton. *Evol Dev* 3:3–17.
3. 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 Genomics* 10:378.
4. Rapp RA, et al. (2010) Gene expression in developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biol* 8:139.
5. Chaudhary B, et al. (2008) Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. *Evol Dev* 10:567–582.
6. Hovav R, et al. (2008) The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. *PLoS Genet* 4:e25.
7. Hu G, et al. (2011) Genomically biased accumulation of seed storage proteins in allopolyploid cotton. *Genetics* 189:1103–1115.
8. Shi YH, 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.
9. Li XB, Fan XP, Wang XL, Cai L, Yang WC (2005) The cotton ACTIN1 gene is functionally expressed in fibers and participates in fiber elongation. *Plant Cell* 17:859–875.
10. Mei W, Qin Y, Song W, Li J, Zhu Y (2009) Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. *J Genet Genomics* 36:141–150.
11. Baluska F, et al. (2000) Root hair formation: F-actin-dependent tip growth is initiated by local assembly of profilin-supported F-actin meshworks accumulated within expansin-enriched bulges. *Dev Biol* 227:618–632.
12. Augustine RC, Vidali L, Kleinman KP, Bezanilla M (2008) Actin depolymerizing factor is essential for viability in plants, and its phosphoregulation is important for tip growth. *Plant J* 54:863–875.
13. Vidali L, et al. (2009) Rapid formin-mediated actin-filament elongation is essential for polarized plant cell growth. *Proc Natl Acad Sci USA* 106:13341–13346.
14. Ji SJ, 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.
15. Wang HY, Yu Y, Chen ZL, Xia GX (2005) Functional characterization of *Gossypium hirsutum* profilin 1 gene (*GhPFN1*) in tobacco suspension cells. Characterization of in vivo functions of a cotton profilin gene. *Planta* 222:594–603.
16. Böttcher RT, et al. (2009) Profilin 1 is required for abscission during late cytokinesis of chondrocytes. *EMBO J* 28:1157–1169.
17. Tanaka M, Shibata H (1985) Poly(L-proline)-binding proteins from chick embryos are a profilin and a profilactin. *Eur J Biochem* 151:291–297.
18. Carlsson L, Nyström LE, Sundkvist I, Markey F, Lindberg U (1977) Actin polymerizability is influenced by profilin, a low molecular weight protein in non-muscle cells. *J Mol Biol* 115:465–483.
19. Pantaloni D, Carlier MF (1993) How profilin promotes actin filament assembly in the presence of thymosin beta 4. *Cell* 75:1007–1014.
20. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112:453–465.
21. Yarmola EG, Dranishnikov DA, Bubb MR (2008) Effect of profilin on actin critical concentration: a theoretical analysis. *Biophys J* 95:5544–5573.
22. Sagot I, Rodal AA, Moseley J, Goode BL, Pellman D (2002) An actin nucleation mechanism mediated by Bni1 and profilin. *Nat Cell Biol* 4:626–631.
23. Romero S, et al. (2004) Formin is a processive motor that requires profilin to accelerate actin assembly and associated ATP hydrolysis. *Cell* 119:419–429.
24. Pruyne D, et al. (2002) Role of formins in actin assembly: Nucleation and barbed-end association. *Science* 297:612–615.
25. Goldschmidt-Clermont PJ, Machesky LM, Baldassare JJ, Pollard TD (1990) The actin-binding protein profilin binds to PIP2 and inhibits its hydrolysis by phospholipase C. *Science* 247:1575–1578.
26. Bae YH, et al. (2010) Profilin1 regulates PI(3,4)P2 and lamellipodin accumulation at the leading edge thus influencing motility of MDA-MB-231 cells. *Proc Natl Acad Sci USA* 107:21547–21552.
27. Vidali L, Augustine RC, Kleinman KP, Bezanilla M (2007) Profilin is essential for tip growth in the moss *Physcomitrella patens*. *Plant Cell* 19:3705–3722.
28. Pang CY, et al. (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.
29. Zhao PM, et al. (2010) Proteomic identification of differentially expressed proteins in the Ligon lintless mutant of upland cotton (*Gossypium hirsutum* L.). *J Proteome Res* 9:1076–1087.
30. Wang J, et al. (2010) Overexpression of a profilin (GhPFN2) promotes the progression of developmental phases in cotton fibers. *Plant Cell Physiol* 51:1276–1290.
31. Huang S, McDowell JM, Weise MJ, Meagher RB (1996) The *Arabidopsis* profilin gene family. Evidence for an ancient split between constitutive and pollen-specific profilin genes. *Plant Physiol* 111:115–126.
32. Morales S, Jiménez-López JC, Castro AJ, Rodríguez-García MI, Alché JD (2008) Olive pollen profilin (Ole e 2 allergen) co-localizes with highly active areas of the actin cytoskeleton and is released to the culture medium during in vitro pollen germination. *J Microsc* 231:332–341.
33. Liu Q, Guo Z (2009) Molecular cloning and characterization of a profilin gene *BnPFN* from *Brassica nigra* that expressing in a pollen-specific manner. *Mol Biol Rep* 36:135–139.
34. Staiger CJ, et al. (1993) The profilin multigene family of maize: Differential expression of three isoforms. *Plant J* 4:631–641.
35. Schutt CE, Myslik JC, Rozycki MD, Goonesekere NC, Lindberg U (1993) The structure of crystalline profilin-beta-actin. *Nature* 365:810–816.
36. Mahoney NM, Janmey PA, Almo SC (1997) Structure of the profilin-poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. *Nat Struct Biol* 4:953–960.
37. Lassing I, Lindberg U (1985) Specific interaction between phosphatidylinositol 4,5-bisphosphate and profilactin. *Nature* 314:472–474.
38. Senchina DS, et al. (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol Biol Evol* 20:633–643.
39. Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol* 186:123–134.
40. Argiriou A, Kalivas A, Michailidis G, Tsaftaris A (2011) Characterization of PROFILIN genes from allotetraploid (*Gossypium hirsutum*) cotton and its diploid progenitors and expression analysis in cotton genotypes differing in fiber characteristics. *Mol Biol Rep*, 10.1007/s11033-011-1125-3.
41. Kandasamy MK, McKinney EC, Meagher RB (2002) Plant profilin isovariants are distinctly regulated in vegetative and reproductive tissues. *Cell Motil Cytoskeleton* 52:22–32.
42. Burger JC, Chapman MA, Burke JM (2008) Molecular insights into the evolution of crop plants. *Am J Bot* 95:113–122.
43. Burke JM, Burger JC, Chapman MA (2007) Crop evolution: From genetics to genomics. *Curr Opin Genet Dev* 17:525–532.
44. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321.
45. Tan L, et al. (2008) Control of a key transition from prostrate to erect growth in rice domestication. *Nat Genet* 40:1360–1364.
46. Vollbrecht E, Springer PS, Goh L, Buckler ES, 4th, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. *Nature* 436:1119–1126.
47. McGrath PT, et al. (2011) Parallel evolution of domesticated *Caenorhabditis* species targets pheromone receptor genes. *Nature* 477:321–325 10.1038/nature10378.
48. Wendel J, Brubaker C, Seelanan T (2010) *The Origin and Evolution of Gossypium. Physiology of Cotton*, eds Stewart J, Oosterhuis D, Heitholt J, Mauney J (Springer, The Netherlands), pp 1–18.
49. Wendel J, Percy R (1990) Allozyme diversity and introgression in the Galapagos endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem Syst Ecol* 18:517–528.
50. Bao Y, Ge S (2004) Origin and phylogeny of *Oryza* species with the CD genome based on multiple gene sequence data. *Plant Syst Evol* 249:55–66.
51. Marchler-Bauer A, et al. (2005) CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res* 33(Database issue):D192–D196.
52. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment. analysis workbench. *Bioinformatics* 25:1189–1191.
53. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.
54. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
55. Wan CY, Wilkins TA (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal Biochem* 223:7–12.
56. 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.
57. Hovav R, et al. (2008) A majority of cotton genes are expressed in single-celled fiber. *Planta* 227:319–329.

EVOLUTION