ANNUAL REVIEWS

Annual Review of Plant Biology

Recent Advances and Future Perspectives in Cotton Research

Gai Huang,[1,2] Jin-Quan Huang,[3] Xiao-Ya Chen,[3] and Yu-Xian Zhu[1]

[1]Institute for Advanced Studies, Wuhan University, Wuhan 430072, China; email: zhuyx@whu.edu.cn

[2]State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China

[3]State Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Annu. Rev. Plant Biol. 2021. 72:437–62

First published as a Review in Advance on January 11, 2021

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-080720-113241

Keywords

cotton genome, evolution, fiber elongation, genetics, gossypol, epigenetics

Abstract

Cotton is not only the world's most important natural fiber crop, but it is also an ideal system in which to study genome evolution, polyploidization, and cell elongation. With the assembly of five different cotton genomes, a cotton-specific whole-genome duplication with an allopolyploidization process that combined the A- and D-genomes became evident. All existing A-genomes seemed to originate from the A_0-genome as a common ancestor, and several transposable element bursts contributed to A-genome size expansion and speciation. The ethylene production pathway is shown to regulate fiber elongation. A tip-biased diffuse growth mode and several regulatory mechanisms, including plant hormones, transcription factors, and epigenetic modifications, are involved in fiber development. Finally, we describe the involvement of the gossypol biosynthetic pathway in the manipulation of herbivorous insects, the role of *GoPGF* in gland formation, and host-induced gene silencing for pest and disease control. These new genes, modules, and pathways will accelerate the genetic improvement of cotton.

Contents

1. INTRODUCTION

Cotton belongs to the genus *Gossypium* in the family Malvaceae and produces the single most important textile fiber, accounting for ~35% of the world's total annual fiber demands. It is also used as a model system to study plant polyploidization, cell elongation, and cell wall biogenesis (36, 137). The *Gossypium* genus comprises 45 diploid species ($2n = 2x = 26$) and seven tetraploid species ($2n = 4x = 52$) with extraordinary morphological variations, including different plant architectures ranging from wild perennial small trees and shrubs to cultivated herbaceous annuals, with variable leaf shapes and different fiber characteristics (**Figure 1**).

Cotton has long attracted attention from agricultural scientists, taxonomists, and evolutionary biologists, as it exhibits extraordinary genomic diversity with global radiation, which has led to the evolution of eight diploid cotton groups (A-, B-, C-, D-, E-, F-, G-, and K-genomes) plus an AD-genome clade. The *Gossypium* genus can be divided into three major lineages primarily based on morphological and geographical evidence: the New World clade (D- and AD-genomes), the African-Asian clade (A-, B-, E-, and F-genomes), and the Australian clade (C-, G-, and K-genomes) (**Figure 1*k***). Most wild cotton species have very short fibers that adhere tightly to the seed; only four cotton species, A_1, A_2, $(AD)_1$, and $(AD)_2$, have been domesticated to produce textile fiber (**Figure 1*l***). Upland cotton, *Gossypium hirsutum* [$(AD)_1$], presently dominates the world's cotton commerce by producing ~95% of the natural lint fiber used by the textile industry (34, 49). There are four overlapping stages during fiber development: initiation, elongation, secondary cell wall (SCW) biosynthesis, and maturation, which are defined on the basis of the number of days postanthesis (DPA). Cotton fibers can be further classified into two types: adherent fuzz fibers, which initiate at 5 to 10 DPA and grow to a final length of less than ~5 mm, and spinnable lint fibers, which initiate before flowering and grow to a final length of ~3 cm.

***Gossypium*:** a cotton genus of the family Malvaceae

Polyploidization: the evolutionary process that combines two or more sets of chromosomes in one nucleus

Domestication: the process of selecting various traits from crops to fit human requirements

Secondary cell wall (SCW): refers to the thickened layers formed internally and subsequently to the deposition of primary cell wall

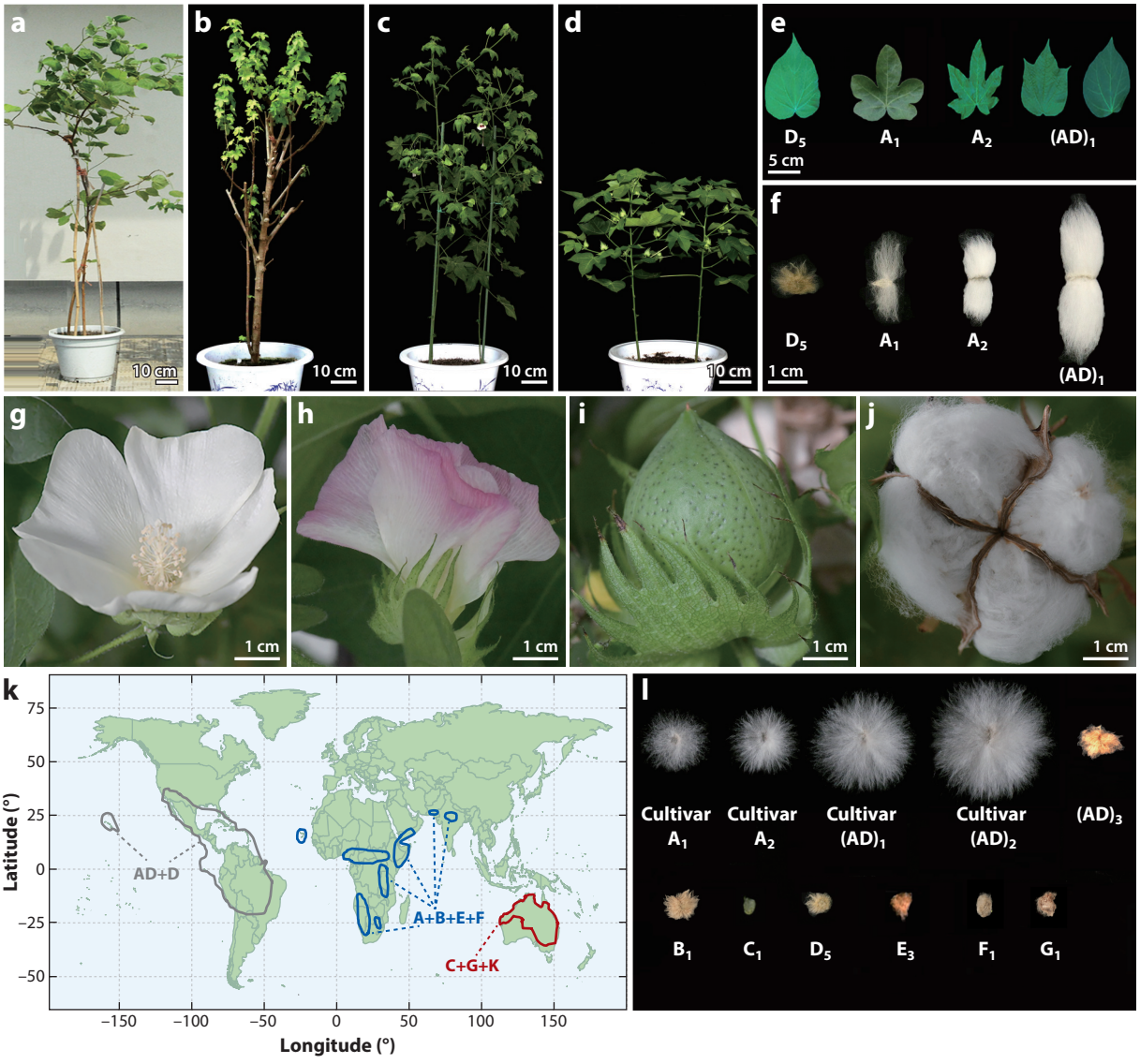


Figure 1

An overview of cotton phenotypes, geographical distributions, and seed trichome variation within the *Gossypium* genus. (*a–d*) Cotton plant architecture, showing (*a*) *Gossypium raimondii*, (*b*) *Gossypium herbaceum*, (*c*) *Gossypium arboreum*, and (*d*) *Gossypium hirsutum*. (*e*) Cotton leaf shape. (*f*) Cotton fiber. (*g–j*) Cotton flower and boll in *G. hirsutum*, including (*g*) white blooming flower, (*h*) reddened flower at 1 day postanthesis (DPA), (*i*) developing boll at 15 DPA, and (*j*) opened boll with mature fibers. (*k*) Geographical distributions of A- through G-, K-, and AD-genomes. Gray, blue, and red indicate the AD- and D-genome clade; the A-, B-, E-, and F-genome clade; and the C-, G-, and K-genome clade, respectively. (*l*) Phenotypes of cottonseed trichomes among cotton species.

Since the publication of the first draft of the cotton (*Gossypium raimondii*) genome in 2012 (111), 24 genome assemblies from 10 different cotton species have been reported (5, 11, 14, 34, 35, 49, 50, 58, 70, 101, 114, 136, 140, 147). In this review, we present and discuss recent advances in cotton genome and cotton biology. We also describe functional genomics research into cotton fiber

development, secondary metabolism, and biotic and abiotic stress responses. Finally, we address current challenges and suggest future directions for cotton research.

Whole-genome duplication: a form of polyploidization process that in an organism doubles the chromosomes and results in large numbers of retained duplicates

A_t- or D_t-subgenome: corresponding chromosome sets for the allotetraploid (AD)-genome resembling diploid A- or D-genomes, respectively

2. COTTON GENOMES

2.1. Cotton Genome Sequencing and Assembly

There are 45 diploid cotton species throughout the world. *G. raimondii*, also known as the D_5-genome that originated in Mexico, possesses one of the smallest nuclear genomes and is regarded as a potential donor to all polyploid cottons. The African native *Gossypium herbaceum* and the Asiatic native *Gossypium arboreum* were collectively called the A-genome and named specifically as the A_1- and A_2-genomes, respectively. Both *G. herbaceum* and *G. arboreum* have similar genomic characteristics and diverged in a short evolutionary time, ~0.7 million years ago. These diploid cottons have a relatively long history of cultivation, with ~1.5-cm fiber cells that are barely fit for human clothing. It is widely accepted in the cotton community that *G. herbaceum*- or *G. arboreum*-like and *G. raimondii*-like cotton were reunited by transoceanic dispersal that ultimately led to the rise of at least seven allotetraploid AD-genome species ($2n = 4x = 52$) [$(AD)_1$- to $(AD)_7$-genomes] (35). Two of them, *G. hirsutum* [$(AD)_1$-genome] and *Gossypium barbadense* [$(AD)_2$-genome], were domesticated for human use ~8,000 years ago and became the major cotton cultivars, since they produce ~3.0-cm-long fiber cells suitable for the modern textile industry (121).

The *G. raimondii* genome was the first to be sequenced and assembled, perhaps because it not only served as the D-genome donor, but it also is one of the smallest diploid cotton genomes (~800 Mb). To date, three *G. raimondii* assemblies from accessions $D_{5\text{-}3}$ (111), Ulbrich (70), and $D_{5\text{-}4}$ (101) have been released. Substantial chromosome rearrangements and a hexaploidization event ~115–146 million years ago (Mya) that is common to eudicots (123), as well as cotton-specific whole-genome duplications at 13–20 Mya, were indicated to have occurred during *G. raimondii* evolution (70, 111). $D_{5\text{-}4}$, the most up-to-date *G. raimondii* genome, successfully improved the genome quality and completeness by filling up gaps, resulting in a final genome size of ~761 Mb (101). It is well suited for various comparative, genetic, and genomic analyses.

Cultivar *Shixiya1* of *G. arboreum* (A_2-genome), which represents one of the phylogenetically closest relatives to the A_t-subgenome of tetraploid cottons, was first sequenced and assembled in 2014 (50) and further updated in 2018 (14). A third and final assembly of the A_2-genome has recently been published (35). In this latest version, ambiguous sequences and misassembled, unknown repeat sequences were substantially reduced, resulting in a more accurate genome size (1,637 Mb) and transposable element (TE) content (80.1%). The current assembly represents a chromosome-scale reference for *G. arboreum* with high completeness and accuracy (35). A comprehensive and high-resolution gene map of *G. arboreum*, which revealed thousands of tissue-specific expressed genes, alternative uses of transcription start sequences, and polyadenylation sites and alternative splicing hotspots in cotton, was published and deposited in public databases (105, 110). An A_1-genome, *G. herbaceum* var. *africanum* (A1-0076), was reported in the same article (35), as part of a large-scale comparative genomic study. The resulting assembly captured 1,556 Mb of genome sequence and a genome localization rate of 95.7% (35).

To date, there have been seven successful assemblies for the most widely cultivated cotton species, *G. hirsutum* line TM-1 [$(AD)_1$-genome], and one for the cultivar ZM24 genome (11, 34, 35, 49, 114, 136, 147). Comparative genomic studies revealed that there are three large-scale (>4 Mb) inversions located at chromosome 8 of the A_t-subgenome between TM-1 and ZM24, which led to isolated haplotypes for the two populations and suppressed meiotic recombination in this region (136). The final genome size reported in the latest assembly is 2,290 Mb without

gap-filling sequences, 99.2% of which were anchored and oriented on 26 chromosomes (35), which may supplant prior assemblies as a chromosome-grade reference genome for *G. hirsutum*.

The genomes of three accessions, including Xinhai21 (58), 3-79 (11, 114, 140), and Hai7124 (34), all from the much-less-cultivated tetraploid cotton *G. barbadense* [$(AD)_2$-genome], have been published, suggesting that scientists have great interest in this species, probably because of its longer and stronger fiber. The current *G. barbadense* reference genome (114) is 2,267 Mb in size, and its anchoring rate reaches 97.7%. Although *G. barbadense* and *G. hirsutum* diverged over a relatively short period of time (~0.4–0.6 Mya), these two genomes underwent distinguishable genomic divergence, with an average of 5.89 single-nucleotide polymorphisms (SNPs) per kilobase and one 170.2-Mb inversion. Meanwhile, in-depth comparative transcriptome analysis showed that the transcriptional activation of genes associated mainly with membrane transport, glycan biosynthesis, and carbon metabolism was associated with the longer-fiber phenotype in Hai7124 (34). Continuing efforts to update diploid and allotetraploid genomes will provide valuable resources for the cotton community to accelerate breeding programs aimed specifically at better fibers.

2.2. Comparative Analysis of Cotton Genomes

Cotton is an excellent system in which to study genome size evolution, as huge variations were found among *Gossypium* species, even within diploid genomes (109). All diploid cotton species have 13 chromosomes with genome sizes ranging from ~738 Mb in the D_5-genome to ~2,858 Mb in the Australian K-genome. Thus, *Gossypium* has undergone a greater-than-threefold increase in genome sizes in the approximately 13 to 15 million years since it began to diverge from a common ancestor, mainly due to the accumulation of long terminal repeat (LTR) retrotransposon activities (50). Among sequenced Malvales genomes (2, 21, 35, 95, 102, 142), *Gossypium* is phylogenetically closest to the *Gossypioides/Kokia* lineage, and the smallest cotton D-genomes have similar genome sizes to *Bombax ceiba* (895 Mb) and *Durio zibethinus* (715 Mb) (**Figure 2**). An analysis of 22 American D-genome accessions suggested that diploid D-genome cotton (subgenus *Houzingenia*) originated ~6.6 Mya, with subsequent diversification events during the mid-Pleistocene at ~0.5–2.0 Mya. The D-genome species range from ~750 Mb to ~900 Mb, a narrow range that may have resulted from a process called genome downsizing bias, which counteracts genome size growth by TEs (23). Compared with the two diploid A-genomes (A_1-genome, 1,556 Mb; A_2-genome, 1,637 Mb), the A_t-subgenome of *G. hirsutum* (1,449 Mb) is significantly reduced, whereas its D_t-subgenome is expanded from 738 Mb in the putative D-genome donor *G. raimondii* to the current subgenome of 822 Mb (49). This genomic downsizing in the A_t-subgenome and amplification in the D_t-subgenome are closely related to substantially more active TE insertions in the D_t-subgenome than in the A_t-subgenome. Both A_1- and A_2-genomes experienced a further twofold expansion in genome size relative to that of the D_5-genome, and, as a result, they consist of ~80% TEs, especially *Gypsy*-type LTRs (35, 50). A Gaussian probability density function analysis (35), which overcame the major pitfall of most previous similar studies that relied on the presence of both ends of full-length LTRs (34, 50), was used to estimate the insertion time of full and fragmented *Gypsy*-type LTRs in cotton genomes. The earliest LTR peak was found at ~5.7 Mya, which corresponds to the expected speciation time for the A- and D-genomes. The second peak, at ~2.0 Mya, in both D_t- and A_t-subgenomes, suggests that allotetraploid cotton may have formed at this time. The third peak, at ~0.89 Mya, occurred only in the two A-genomes, whereas the fourth peak, at ~0.61 Mya, and fifth peak are unique to the A_1- and A_2-genomes, respectively (35). These TE bursts ultimately shaped the distinct genomic architecture of cotton and contributed significantly to its genome size expansion, speciation, and evolution.

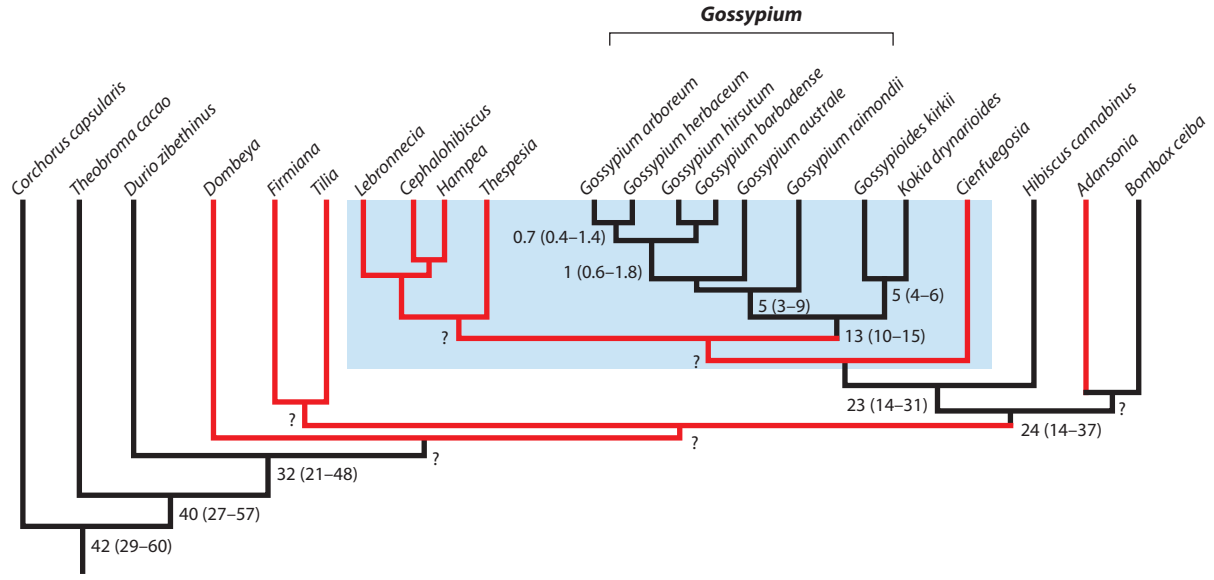


Figure 2

Inferred phylogenetic relationships of major species belonging to the Malvaceae family. Red lines indicate predicted possible relationships with no sequenced genomic information available for that genus. The shading represents the cotton tribe. The number with the 95% highest posterior density intervals in each clade represents the estimated divergence time in million years ago (indicated in parentheses).

Although the A_1- and A_2-genomes diverged over a relatively brief period of time, the two A-genomes accumulated substantial genomic and genetic differences. For example, the A_2-genome underwent a reciprocal translocation between chromosomes 1 and 2, and the A_1-genome experienced two large-scale (>40 Mb) inversions in chromosomes 10 and 12 after their speciation (22, 35). The genetic divergence between the two A-genomes reached nearly 1.0, suggesting that the A_1- and A_2-genomes can be clearly distinguished as two cotton species. By comparing two A-genomes with the A_t-subgenome of *G. hirsutum*, a large number of structural variants (SVs), including 61,053 in the A_1-genome and 61,383 in the A_2-genome, were characterized, with 35,997 (41.64%) of these SVs being shared in both A-genomes (35). These huge genetic differences and chromosomal SVs provide compelling evidence to suggest that the two A-genomes originated independently with no ancestor–progeny relationship between them, which may explain why interspecific hybridization between the two A-genomes is often unsuccessful.

3. CURRENT UNDERSTANDING OF COTTON EVOLUTION

3.1. Origin of *Gossypium* and Evolution of Allotetraploid Cotton

Structural variants (SVs): sequence alterations covering ≥50 base pairs that result in duplications, insertions, deletions, inversions, and translocations

Despite its different geographical origins and morphological and cytogenetic diversities, *Gossypium* constitutes a single monophyletic group that originated from a common ancestor ~5–10 Mya (121). Thus far, there are two views with regard to cotton evolution: One suggests that the B-genome is the primitive group that produced all other cotton species (12, 125); the second suggests that the D-genome is the common ancestor of all cotton taxa (13). The incongruence of phylogenetic relationships in this genus has highlighted the necessity of using multiple and independent studies, especially genome analyses, to fully elucidate the origin of any given group of higher plants.

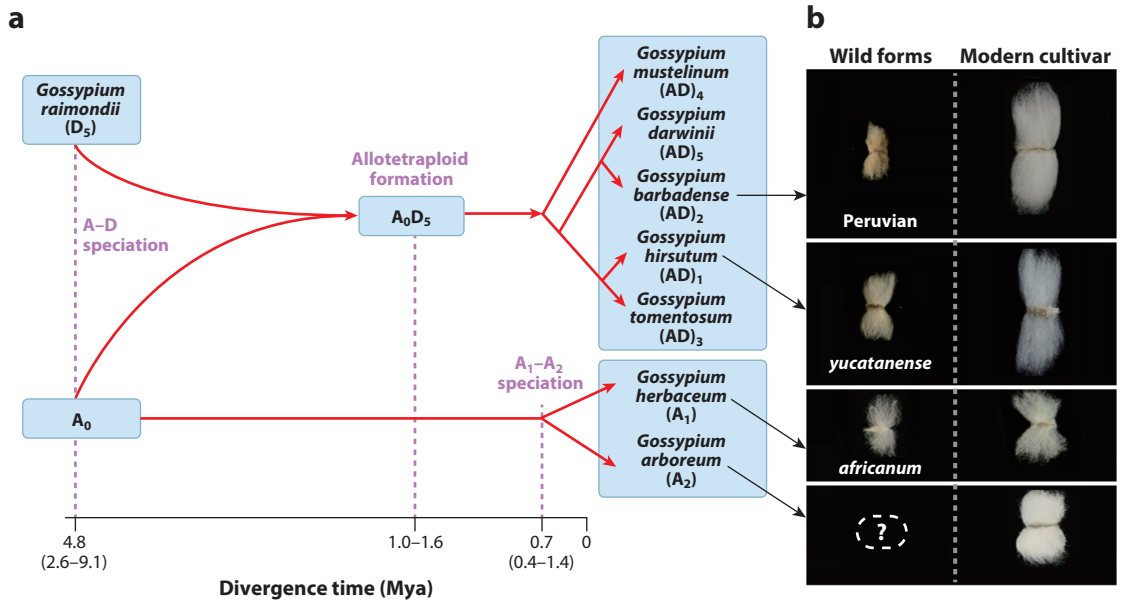


Figure 3

The evolution and domestication of cottons. (*a*) The origin and evolution of five allotetraploid cottons and two diploid A-genomes. (*b*) Cottonseed with associated fiber from the wild forms (*left*) and domesticated cultivars (*right*). The black arrows indicate the species of origin for each image.The Peruvian *Gossypium barbadense*, *Gossypium hirsutum* variety *yucatanense*, and *Gossypium herbaceum* subspecies *africanum* represent the wild cottons. Abbreviation: Mya, million years ago.

To date, seven allopolyploid cotton species, including $(AD)_1$, $(AD)_2$, *Gossypium tomentosum* $[(AD)_3]$, *Gossypium mustelinum* $[(AD)_4]$, *Gossypium darwinii* $[(AD)_5]$, *Gossypium ekmanianum* $[(AD)_6]$ and *Gossypium stephensii* $[(AD)_7]$, have been reported (18). Phylogenetic studies indicated that, among the first five species, $(AD)_4$ may serve as the basal clade, with $(AD)_1$ and $(AD)_3$ forming the second clade, whereas $(AD)_2$ and $(AD)_5$ form the third clade (**Figure 3*a***). This means the D_t-subgenome and A_t-subgenome in all of the five species originated from common diploid D- and A-genome species at 1.0 to 1.6 Mya and then gradually branched into five species within 0.20 to 0.63 Mya (11). The two subgenomes in each of the five polyploid species are highly conserved at the chromosomal, gene content, and nucleotide levels, with more substantial diversifications found in evolutionary rate, heterogeneities, and the expression patterns of gene families and homologs.

All tetraploid cottons are known to be directly descended from an allopolyploidization event involving hybridization between the A- and D-genome ancestors, followed by genome doubling (11, 49). Abundant studies support a *Gossypium* species resembling *G. raimondii* (D_5-genome) as the D-genome donor (34, 49, 114, 147). However, controversy persists as to which is the actual A-genome donor in tetraploid cottons. An early phylogenetic analysis suggested that a hybridization between Asiatic cotton (A_2-genome) and the American diploid cotton (D_5-genome) produced the ancestor for all tetraploid cottons (88). Based on a second study, African cotton (A_1-genome) was suggested to be an appropriate partner in the tetraploid cotton, as its genome is more primitive than the A_2-genome (22). Chromosomal-scale analysis of A_1-, A_2-, and $(AD)_1$-genomes revealed that neither the A_1- nor A_2-genome is the actual A-genome donor (35). Instead, a common and possibly extinct A_0-genome, which may serve as an ancestor to the A_1- and A_2-genome clade, is the closest relative to the A_t-subgenome rather than either the A_1- or A_2-genome (35). SNP analysis based on cotton populations showed that the A_t-subgenome was much closer to the outgroup

A_0-genome: the putative common ancestor of all A-genomes, including the extant diploid A-genomes and A_t-subgenome of allotetraploid cottons

D_5-genome (with 30.54% ancestral alleles) than the A_1- (20.52%) or A_2-genome (20.04%). Estimation of divergence time indicated that allotetraploid cottons were formed at ~1.0–1.6 Mya, which preceded the speciation of A_1- and A_2-genomes at ~0.7 Mya. Apparently, hybridization of this A_0-genome with a D_5 species produced the current allotetraploid cotton, whereas it subsequently diverged into the present-day A_1- and A_2-genomes (35) (**Figure 3*a***). This genome-based analysis will likely shut down the A_1- versus A_2-genome argument, especially if archaeological data are someday uncovered that confirm the existence of the currently extinct A_0-genome.

K_S: the number of synonymous substitutions per synonymous site

3.2. Evolution Following Polyploidization

Asymmetric genome evolution seems to have occurred in both *G. hirsutum* and *G. barbadense*, as there is an overall acceleration in evolutionary rate in the D_t-subgenome relative to that of the A_t-subgenome (49). The average K_s values for collinearity-supported gene pairs were 0.463 for the D_t-subgenome versus *Theobroma cacao* (a close relative to cotton) and only 0.422 for the A_t-subgenome in the same assay. Analysis of intergenic collinear regions also showed that the single-nucleotide variation rate for the D_t-subgenome versus the D-genome was greater than that for the A_t-subgenome versus the A-genome (49). Large variants in the genome, including two reciprocal translocations between chromosomes 2 and 3 and chromosomes 4 and 5, chromosomal rearrangements, and large pericentric inversions were specifically found in the A_t-subgenome, suggesting that the A-genome in diploid cotton was reorganized sometime after the A_0-genome was incorporated into the allotetraploid cotton genome (35, 114).

Polyploidization in cotton induced a wide spectrum of gene expression changes and novel interactions. A transcriptomic analysis of 35 vegetative and reproductive tissues has demonstrated that 20 to 40% of homoeologous gene pairs showed A_t- or D_t-subgenome-biased expression in *G. hirsutum* (147). Comparative fiber transcriptomes among wild strains, domesticated strains, and their F_1 hybrids uncovered genome-wide and novel *cis*- and *trans*-regulatory patterns (3). A total of 1,655 fiber-expressed genes with *cis*- and *trans*-regulatory variations were found to form through divergence and domestication. Of these, 513 genes exhibited *cis*-only divergence, 301 genes exhibited *trans*-only divergence, and the remaining 841 genes exhibited both *cis*- and *trans*-divergence. A_t-biased expression is more often associated with *trans*-only regulatory mechanisms, whereas D_t-biased expression is more closely related to *cis*-only regulatory changes (3). In addition, up to 80% of the long noncoding RNAs (lncRNAs) were reported to exhibit allelic expression in the allotetraploid cotton, leading to the hypothesis that hybridization and polyploidization enabled the neofunctionalization of lncRNA transcription (149).

3.3. Cotton Domestication

Cotton was one of the earliest domesticated economical crops, and it has experienced at least four parallel domestications that resulted in the production of diploid and allotetraploid cotton species (30, 78). Perhaps the most obvious phenotypic change in domesticated cotton is its significantly elongated and much-strengthened fiber cells (**Figure 3*b***). *G. hirsutum* was initially domesticated from a wild race named *yucatanense* in the Yucatan peninsula. In addition to *yucatanense*, there are six other recognized races (*punctatum*, *palmeri*, *latifolium*, *marie-galante*, *morrilli*, and *richmondi*) in the same group that may be considered as semiwild forms of cotton (17). Studies of a variation map from 352 wild and domesticated Upland cotton accessions revealed a total of 93 domestication sweeps that occupied 178 Mb of the genome (74 Mb in the A_t-subgenome, 104 Mb in the D_t-subgenome). Also, 1,228 genetic loci from the D_t-subgenome and 549 from the A_t-subgenome were found under strong selective pressure (113). Another experiment showed that 4,754

genetically selected loci associated with fiber-related traits were located in the D_t-subgenome, while only 2,587 were found in the A_t-subgenome (62), suggesting that the coexisting subgenomes of *G. hirsutum* may have undergone asymmetric domestication selections.

A survey of genetic diversity for *G. barbadense* germplasms indicated that its early domestication might have taken place in coastal regions of northwestern Peru and southwestern Ecuador around Guayaquil Gulf. Thus, northwestern Peruvian and southwestern Ecuadorian accessions have a high diversity of primitive types and are basal to other *G. barbadense* accessions (122). A number of interspecific reciprocal introgression events have been found between *G. hirsutum* and *G. barbadense*, and the introgression pattern is significantly biased toward the gene flow from *G. hirsutum* into *G. barbadense* (16). Comprehensive epigenomic analysis of wild and domesticated cottons identified 519 differentially methylated genes, some of which, such as flowering time and seed dormancy, contribute significantly to domestication traits (86). These identified genetic loci reveal the molecular basis underlying phenotypic changes and provide targets for cotton breeding.

Introgression: the introduction of genetic materials from one species into another species

Both diploid cottons *G. herbaceum* and *G. arboreum* have short, coarse, and weak fibers with very low commercial value today, although they were historically domesticated for human clothing. Morphological evidence and genomic data support the assumption that the wild progenitor of *G. herbaceum* was descended from the southern African subspecies *africanum* (35) (**Figure 3*b***). Although the ancestor of *G. arboreum* is not clearly known yet, we have reasons to suggest that it might have been domesticated first in either Madagascar or the Indus Valley before being dispersed to China from India and/or Pakistan and subsequently to other areas (35, 120). The first variation map of diploid cultivated cotton was constructed by the resequencing of 230 *G. arboreum* and 13 *G. herbaceum* accessions mostly collected from China (14). Phylogenetic analysis based on SNPs is consistent with the hypothesis that the Chinese *G. arboreum* geographical race originated in South China and was subsequently introduced into the Yangtze River and Yellow River regions (14). Pairwise comparison among accessions from the South China, Yangtze River and Yellow River regions identified a number of genetically divergent regions that overlap with identified quantitative trait loci encoding traits related to maturity, yields, boll weight, and disease resistance, which indicated that Chinese *G. arboreum* accessions have been under strong human and/or geographical selections (14). These identified domesticated loci represent an important high-resolution genetic resource that will facilitate the improvement of complex cotton traits and enable important characteristics of diploid cottons to be transferred into the tetraploids through interspecific hybridization.

4. REGULATION OF COTTON FIBER DEVELOPMENT

4.1. Modes of Fiber Cell Initiation and Elongation

Cotton fibers grow in a highly polarized manner governed by the actin cytoskeleton and microtubule organization. Quite a number of cytoskeleton-related genes, such as the actin-related *GhACT_LI1* (96), *GhACTIN1* (53), *GhADF1* (107), *GhFIM2* (143), and *GhPFN2* (108); the tubulin-related *GhTUA9* (51); and the kinesin-encoding *GhKCH1* (71), are critical for fiber elongation and cell wall formation in cotton. RNA interference (RNAi) of *GhACTIN1* expression disrupts the actin cytoskeleton network with significantly reduced fiber elongation growth without affecting fiber initiation (53). Downregulation of the gene encoding an actin depolymerizing factor, *GhADF1*, increases both fiber length and strength (107). Overexpression of *GhFIM2*, which encodes an actin-bundling protein, accelerates fiber growth with increased actin bundle formation (143). Conversely, overexpression of a gene that encodes fiber-preferential actin-binding protein, *GhPFN2*, results in the termination of cell elongation and a short-fiber phenotype, possibly caused by the formation of thicker and more abundant F-actin bundles during the early elongation stages

(108). Thus, the actin cytoskeleton and actin depolymerization may play an important role during the fast and robust fiber elongation growth period.

Linear cell-growth mode: in this mode, microtubules are oriented transversely with diffuse growth in the shank of the fiber cell while the cell's apex has characteristics of tip growth

The mode of cotton fiber cell elongation has been a topic of heated discussions among scientists working in this field in the past several decades. Fiber cells were once regarded as tip-growing cells that depended exclusively on apical cell wall synthesis (89) or a diffuse growth mode due to the presence of transversely arranged microtubules perpendicular to the growth axis of an elongating fiber cell (71, 100). A linear cell-growth mode was proposed based on the finding that a large number of genes involved in vesicle coating and trafficking were preferentially expressed at various stages of fiber growth (76). Live-cell imaging of elongating fiber cytoskeletons showed that these cells may elongate via a unique tip-biased diffuse growth mode (139). Observations of actin networks in cotton fibers, as imaged by the F-actin reporter ABD2-GFP, in multiple transgenic cotton lines, discovered several common properties with *Arabidopsis* hypocotyls or root epidermal cells that are known to elongate via a diffuse growth mode. Also, microtubules in the fiber shanks, as monitored through an mCherry-tagged microtubule plus-end tracking protein EB1b cassette, were also deposited transversely to the growth axis, with a clear microtubule-depleted zone in the fiber apex. Four-dimensional images show endosomal vesicles, stained and visualized using the FM4-64 lipophilic dye as a membrane marker, evenly distributed along the elongating fiber cells and moving bidirectionally along the fiber shank to and from the fiber tip. Taken together, they concluded that cotton fibers probably elongate via a tip-biased diffuse growth mode, similar to the previously proposed linear cell-growth mode (139).

4.2. Fiber Cell Wall Structure: Hypotheses and Recent Evidence

Fiber cell wall formation occurs in three phases: (*a*) a cell elongation period at ~5–25 DPA, characterized solely by the synthesis of the primary cell wall (PCW), which may contain ~23% cellulose fibrils, ~22% proteins, and various amounts of polysaccharides including xyloglucan and pectins; (*b*) the transition stage; and (*c*) the SCW deposition period at ~20–45 DPA (76, 85). The extensibility of the PCW is mediated in part by xyloglucan hydrolases, as overexpression of *GhXTH1* in transgenic cotton results in 15–20% longer fibers when compared with those of wild types (47). During the PCW synthesis stage, several cell wall–loosening proteins, such as GhRDL1 and GhEXPA1, have been characterized to play key roles in cell wall reconstruction (130).

The SCW in cotton fibers is composed of 95% cellulose with almost no lignin, which is a notable difference from other cell types in higher plants, and is synthesized and deposited inside the PCW (76). Cellulose synthesis, which is catalyzed by the cellulose synthase complexes (CSCs), is the predominant event during the SCW deposition stage. The catalytic subunits of plant CSCs for *Arabidopsis* PCW synthesis are encoded by at least three cellulose synthase (CesA) genes (*CesA1*, *CesA3*, and *CesA2* or *CesA6*) with three different *CesA* genes (*CesA4*, *CesA7*, and *CesA8*) being responsible for SCW synthesis (1). Analysis of the *G. hirsutum* genome revealed that there are 32 *CesA* genes in Upland cotton in contrast to 10 in *Arabidopsis*. Among them, *CesA4*, *CesA7*, and *CesA8* were predominately expressed in fiber cells during SCW biosynthesis at levels 1.5- to 40-fold higher in cultivated species than in wild cottons, suggesting their potential role in enhancing lint yield and quality (49, 147). Although CSCs are likely crucial for producing cellulose microfibrils, the number of CesA monomers in a given CSC and the organization of CesAs in the plant kingdom, including cotton, have been debated for many years. The architecture of a poplar tree PttCesA8 complex with homotrimers giving rise to 18 cellulose chains in the context of rosette-like CSCs, stabilized by cytosolic conserved regions and transmembrane helical exchanges, has recently been revealed by cryogenic electron microscopy (73). Researchers will need to further

explore the precise architecture of CSCs to better understand the molecular mechanisms of cellulose biosynthesis.

Other genes that encode proteins essential for cellulose biosynthesis have also been heavily scrutinized. For example, sucrose synthase (SuSy) catalyzes a reversible reaction with a preference to convert sucrose into fructose and uridine diphosphate (UDP)-glucose, both of which are substrates for cellulose biosynthesis (79). Several SuSy genes, including *GhSusA1* (44) and *SusC* (4), are involved in SCW synthesis or fiber cell elongation. Sucrose efflux, which provides the major osmotically active solute to generate the turgor pressure required to drive fiber cell elongation (90), is regulated by *GhSWEET12*, which encodes a sucrose transporter. In plants, RNAi that reduces expression of *GhSWEET12* or its main regulator, myeloblastosis viral oncogene homolog (MYB) transcription factor (TF) *GhMYB212*, significantly decreases sucrose accumulation and results in shorter fibers and a lower lint index.

***lintless-fuzzless* (*fl*):** an Upland cotton mutant obtained from the variety Xuzhou 142 that possesses no fibers

4.3. Epigenetic Modifications and Small RNAs During Fiber Development

Using deep transcriptome sequencing, scientists characterized 35,802 lncRNAs and 2,262 circular RNAs, among which 645 lncRNAs were preferentially expressed in the *lintless-fuzzless* (*fl*) mutant and 651 in fiber-attached lines (32). Virus-induced gene silencing experiments showed that silencing lncRNA XLOC_545639 and XLOC_039050 in the *fl* mutant increased the number of sites of fiber initiation on the ovules, whereas silencing XLOC_079089 in wild-type Xu142 resulted in a short-fiber phenotype. Fiber cells contained higher rates of DNA methylation, mediated predominantly by an active H3K9me2-dependent pathway rather than the classical RNA-directed DNA methylation pathway, when compared to ovule tissue (115). Reduced expression of *GhHDA5*, a histone deacetylase that is preferentially expressed from −1 to 0 DPA, results in very few fiber initials and a much lower lint yield. These RNAi cotton lines showed alterations in H_2O_2 homeostasis and elevated autophagic cell death, suggesting that *GhHDA5* may modulate the expression of stress- and development-related genes involved in fiber growth and development (46).

A *trans*-acting small interfering RNA (tasiRNA) gene, *TAS3*, that represses the auxin response factor *ARF4* expression during the rapid fiber elongation stage is triggered by miR390, whereas *TAS4* is triggered by miR828 and miR858 to generate 21-nt tasiRNAs responsible for fiber initiation (25). Suppressing microRNA (miRNA)156/157 function results in a significant reduction in the mature fiber length, which further illustrates an essential role of these miRNAs in fiber elongation (57). Applications of the CRISPR/Cas9 genome-editing system and similar techniques (9) may help us understand the roles of small RNA and epigenetic regulation in cotton.

4.4. Transcription Factors in Fiber Cell Development

In *Arabidopsis*, the MYB-basic helix-loop-helix (bHLH)-WD40 (MBW) complex including GLABROUS1 (GL1), GL3, ENHANCER OF GLABRA3 (EGLf3), and TRANSPARENT TESTA GLABRA1 (TTG1) regulates expression of the homeodomain leucine zipper (HD-ZIP) IV TF GL2, which promotes trichome growth, whereas TRIPTYCHON (TRY) and CAPRICE (CPC) counteract the complex-formation process by competing with GL1 for GL3 and EGL3 binding (77, 119). Cotton fiber is a distinct type of trichome derived from seed epidermis that is composed of unbranched and extensively elongated single cells. Over 400 *MYB* genes, including those encoding the GL1-type R2R3 MYBs (80), are expressed preferentially in at least one stage of fiber development. As a homolog of *Arabidopsis GL1*, the R2R3 MYB TF *GaMYB2*, which is specifically expressed in the early stages of fiber development, is able to rescue trichome formation in the *Arabidopsis gl1* mutant (117). When the homeodomain-containing protein (HOX) *GaHOX1*,

a homolog of *Arabidopsis GL2*, was introduced into *Arabidopsis thaliana* plants under the control of an authentic *GL2* promoter, the trichome-deficient phenotype of *gl2-2* was complemented successfully (24). Other homologs of the *Arabidopsis* MBW components, including two WD40-repeat proteins (TTG1 and TTG3), GL3 homologs (GhDEL65/61 and GhMYC1), CPC, and TRY, were also functionally characterized in cotton (56, 83). Although these MYB genes may be functionally equivalent to *Arabidopsis* GL1, silencing of *GhMYB109* led to only a mild reduction in fiber length (72). GhTTG2, a homolog of *Arabidopsis* AtTTG2, promotes proanthocyanidin biosynthesis, which results in the production of a brown color in cotton fibers (132), suggesting that the GL1-type MYBs have diverse roles in cotton fiber development.

Several lines of evidence suggest that MIXTA-like MYB TFs, such as the ovule-specific *GhMYB25* and *GhMYB25-like* genes, have been identified to act as the master regulators of cotton fiber initiation. RNAi of the *GhMYB25-like* gene results in glabrous or fiberless cotton seeds, phenocopying the *fl* mutant (103). Map-based cloning of the so-called naked (fuzzless) seed mutant *N1* indicated that the *MYBMIXTA-like TF 3* (*MML3*)/*GhMYB25-like* gene (located on chromosome A12) is associated with fuzz development. Further studies showed that this locus contains two *MML* genes, *MML3* and *MML4*, arranged tandemly to control fuzz initiation (104). Analysis of cotton mutants with impaired fiber development fixed a fiber determinant locus to the homologous region on chromosome D12 (124), where the target gene *GhMML4* (*GhMML4_D12*) functions together with its partner, a diverged WD40 repeat protein (GhWDR), to regulate spinnable lint production (99).

Thirteen HD-ZIP IV genes in *G. arboreum* and 26 in *G. hirsutum* were identified by a genome-wide analysis (8), and three of these, which encode the TFs HOX1 to HOX3, were isolated from both cotton species (24). Although HOX1 is most similar to *Arabidopsis* GL2 (66% identical), experimental data from transgenic cotton plants indicated that GhHOX3 may act as a core regulator of fiber elongation. Silencing of *GhHOX3* expression in *G. hirsutum* by RNAi or as a result of cosuppression drastically reduces fiber length, with fuzz-like short fibers produced on the seeds (82). GhHOX3 interacts with GhHD1 as well as with the cotton DELLA protein GhSLR1, which is a repressor of the phytohormone gibberellic acid (GA) that is well known for promoting plant cell growth. The GhHOX3-GhHD1 complex exhibits higher transcriptional activity towards *GhRDL1* and *GhEXPA1* than either alone, whereas GhSLR1 competitively binds to GhHOX3 to impede the transcriptional activation required for transmitting the GA signal for fiber cell elongation (**Figure 4**).

TCP proteins form another plant-specific TF family and contain a conserved noncanonical bHLH domain for DNA binding. A class I TCP gene of *G. barbadense*, *GbTCP*, is preferentially expressed in fiber cells and is proposed to regulate jasmonate (JA) levels and, subsequently, the level of reactive oxygen species, as well as calcium signaling (28). PRE1, an atypical bHLH factor, promoted hypocotyl cell elongation in *Arabidopsis*. In tetraploid cotton, *GhPRE1A* from its A_t-subgenome is expressed specifically in fiber cells and is presumed to regulate fiber elongation, whereas its D_t-subgenome homolog is inactivated by a TATA-box fragment deletion in its promoter region (148).

The NAM, ATAF1 and ATAF2, and CUC2 (NAC)-type TFs in *Arabidopsis* act as the primary regulators of SCW formation (94). One *G. hirsutum* NAC gene, *GhFSN1*, is specifically expressed in fiber cells during the SCW biosynthesis stage. Overexpression of *GhFSN1* significantly increases cell wall thickness at the expense of fiber lengths, with both *GhMYBL1* and *GhKNL1* TFs known to control *GhFSN1* expression directly (141). Recently, miR319-regulated CIN-type *TCP* genes, such as *GhTCP4*, were shown to play a role in modulating cotton fiber cell elongation and promoting SCW biosynthesis as well (6). GhTCP4 interacted antagonistically with GhHOX3 to constrain fiber elongation growth. At the early stages of cotton fiber development, miR319 is

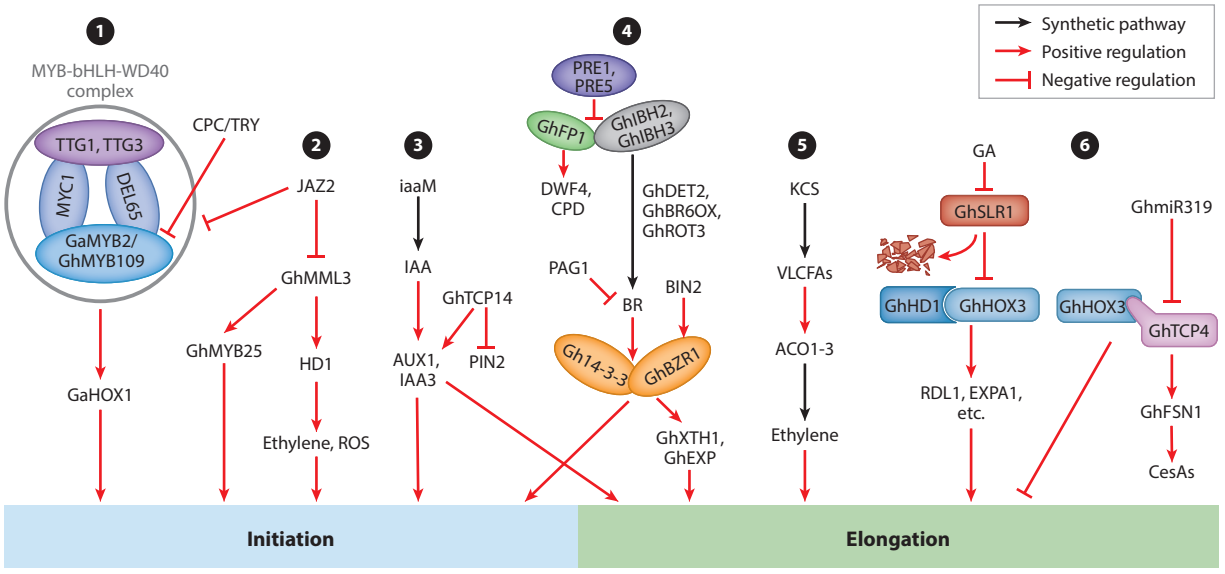


Figure 4

Regulatory networks for cotton fiber initiation and elongation. A total of six major pathways are presented on the schematic model, including (❶) MYB-bHLH-WD40-GaHOX1, (❷) MIXTA-like MYB transcription factors (GhMML3 and GhMYB25), (❸) plant hormones IAA, (❹) BR-mediated regulatory networks, (❺) the signaling between VLCFAs and ethylene production that is essential for maintaining a linear cell-growth mode, and (❻) GA-GhHOX3. The items with colored ovals or rectangles show the protein-protein interactions. Abbreviations: bHLH, basic helix-loop-helix; BR, brassinosteroid; CesA, cellulose synthase; GA, gibberellic acid; IAA, indole-3-acetic acid; ROS, reactive oxygen species; VLCFA, very-long-chain fatty acid.

abundant, and its target TCPs are maintained at low levels, whereas GhHOX3 actively promotes fiber cell elongation. When miR319 expression declines during later stages of fiber growth, the levels of TCPs increase to promote cellulose biosynthesis and SCW formation. Overexpression of an miR319-resistant form reduces fiber cell elongation and accelerates SCW deposition, resulting in short fibers with thicker cell walls (6). Thus, the dynamics of *GhHOX3* expression and the miR319-targeted *TCP* expression patterns modulate the cotton fiber transition from cell elongation to cell wall thickening (**Figure 4**). Additionally, two LIM-domain TFs, GhXLIM6 (54) and WLIM1a (27), affect SCW biosynthesis, although their specific molecular mechanisms remain largely elusive. Experimental analyses combined with newly acquired genomic data will further our understanding of fiber cell development and cellulose biosynthesis for both fiber yield and quality improvement.

4.5. Regulatory Networks That Govern Fiber Development

The plant hormones GA, indole-3-acetic acid (IAA), JA, ethylene, and brassinosteroids (BRs) are known to play pivotal roles in fiber cell development (84, 119, 128). Exogenous application of GA or elevation of its endogenous level by introducing *GhGA20ox1*, a GA biosynthesis gene, promotes significant cotton fiber initiation and elongation (129). Manipulation of endogenous IAA levels by overexpression of *GhiaaM*, a gene responsible for IAA biosynthesis, under an epidermal-specific promoter, greatly enhances fiber initiation and yield (144). Addition of exogenous IAA stimulates *GhTCP14* expression predominantly during the fiber cell initiation and elongation stages (116). GhTCP14 binds directly to the promoters of genes encoding the auxin uptake carrier (AUX1), the

auxin response protein (IAA3), and the auxin efflux carrier (PIN2). Thus, GhTCP14 may play a role in IAA-mediated cotton fiber cell differentiation and elongation. By contrast, overexpression of a JA signaling pathway repressor, *GhJAZ2*, which interacts with GhMYB25-like/GhMML3 and GhGL1, results in fewer fuzz fibers and shorter lint fibers with smaller cotton seeds (31).

Very-long-chain fatty acids (VLCFAs): fatty acids with a chain length of >18 carbons

Gossypol: a polyphenol aldehyde, isolated only from the cotton plant, that harbors a naphthalene core with two aldehydic groups and six hydroxyl groups

Ethylene biosynthesis is the most upregulated biochemical pathway during early cotton fiber development, and the exogenous application of C_2H_4 in the semi–in vivo ovule culture system promotes significant fiber, but not ovule, cell growth (84). Genes responsible for ethylene production, mainly 1-aminocyclopropane-1-carboxylic acid oxidases 1 through 3 (ACO1–ACO3), are expressed at significantly higher levels during this growth stage, and the amount of ethylene released from cultured ovules is correlated with ACO expression and with the rate of fiber growth. Ethylene may promote cotton fiber elongation by increasing the expression of SuSy, tubulin, and expansin genes (84). Differences in ethylene production and its regulatory mechanisms in three cotton species—*G. raimondii* with nonspinnable fibers, *G. arboreum* with shorter fibers, and *G. hirsutum* with long fibers—have been further elucidated from a genomic viewpoint by sequencing and assembling these cotton genomes (49, 50). Elongating fibers contain high levels of ACO activities, likely promoted by elevated levels of saturated very-long-chain fatty acids (VLCFAs), as external C24:0 added to the ovule culture medium results in significant fiber cell elongation, preceded by a rapid and significant increase in *ACO* expression with a prompt and robust ethylene production. This C24:0 effect is blocked in the presence of an ethylene biosynthesis inhibitor (75). Overexpression of *KCS6*, a key gene in VLCFA biosynthesis in Upland cotton, increases the final length of the fiber significantly (~6.0–12%) (35), indicating that VLCFAs may act upstream of the ethylene pathway (76, 127). Comparative proteomics between the *fl* mutant and wild type revealed that most of the wild-type ovule preferential proteins were involved in the biosynthesis of pectic precursors important for PCW during cotton fiber and *Arabidopsis* root hair elongation (69). Collectively, these data support the notion that ethylene and C24:0 may promote cotton fiber and *Arabidopsis* root hair growth by activating the pectic network, especially via enhanced UDP-L-rhamnose and UDP-D-galacturonic acid biosynthesis.

BRs regulate cell elongation, as a BR-deficient cotton mutant, *pag1*, displays a dwarf plant phenotype with reduced fiber length (112, 138). Elevation of BR levels by overexpressing the BR biosynthetic gene steroid 5α-reductase, *GhDET2* (60), or its regulator, the bHLH TF GhFP1 (59), substantially increases the final fiber lengths in transgenic plants. GhFP1 was confirmed to bind with the E-box sequences located at the promoters of two genes related to BR synthesis, *GhDWF4* and *GhCPD* (59). Also, external BR application induces the expression of *brassinazole-resistant 1* (*GhBZR1*), whereas phosphorylation of BZR1 by GhBIN2 kinase facilitates its binding to an acidic regulatory protein Gh14-3-3. Overexpression of Gh14-3-3 results in a significant increase in mature fiber length, and lines in which its expression is suppressed show a retarded fiber initiation process with significantly shorter final fibers (150). The molecular mechanisms regarding fiber cell growth mediated by either BRs or 14-3-3 proteins need to be scrutinized further.

5. SPECIALIZED METABOLITES AND STRESS RESPONSES

5.1. Genetic Control of Pigmented Gland and Nectary Formation

Bioactive and specialized metabolites are often stored in defined plant tissues or structures to avoid self-toxicity. For example, the cotton gland forms a cavity surrounded by thick-walled cells in which specialized metabolites accumulate and are stored. There are glandless cotton cultivars, such as the recessive mutant (gl_2gl_3) of *G. hirsutum* and the dominant mutant (Gl_2^e) of *G. barbadense*, which do not accumulate gossypol and related sesquiterpene aldehydes in aerial organs, including

seed (66, 92). Genetic mapping of Gl_2^e by cross populations identified a pigment gland formation gene (*GoPGF*) that encodes a bHLH TF. Repressing *GoPGF* expression by virus-induced gene silencing leads to emerging leaves that are glandless or have very few glands (61). Sequence analysis of recessive glandless *G. hirsutum* mutants showed that a single T or A is inserted into *GhPGF*, located in the A_t-subgenome or D_t-subgenome, respectively, which results in the premature termination of the corresponding protein, whereas the molecular mechanism that dominantly impairs PGF function remains obscure (61). RNA-seq analysis of embryos of the near-isogenic glanded versus glandless plants identified 33 genes expressed immediately prior to gland formation, among which are three homologous genes, including the cotton gland formation 1 (bHLH TF CGF1), NAC TF CGF2, and synonym of *GoPGF* (CGF3). Knockout of *GhCGF3* resulted in a glandless phenotype, which supports the hypothesis that bHLH TFs serve as master regulators of lysigenous gland formation (42). Phylogenetic analysis showed that PGF is closely related to bHLH14 and distantly related to three *Arabidopsis* MYCs (MYC2, MYC3, and MYC4) that have important roles in JA signaling (74) (**Supplemental Figure 1**). Comparative transcriptomes of three glandless versus four glanded cultivars identified an MYB TF, CGP1, as a regulator of sesquiterpene phytoalexin biosynthesis (20). CGP1 contributes significantly to gland pigmentation, probably by interacting with GoPGF to form a heterodimer in the nucleus, which suggests the presence of a molecular linkage between gland development and the biosynthesis of its metabolic contents. Both the signal involved in gland formation and the mechanism underlying cell death in the gland chamber require further investigations.

Aromatization: conversion of a nonaromatic compound to an aromatic compound

Supplemental Material >

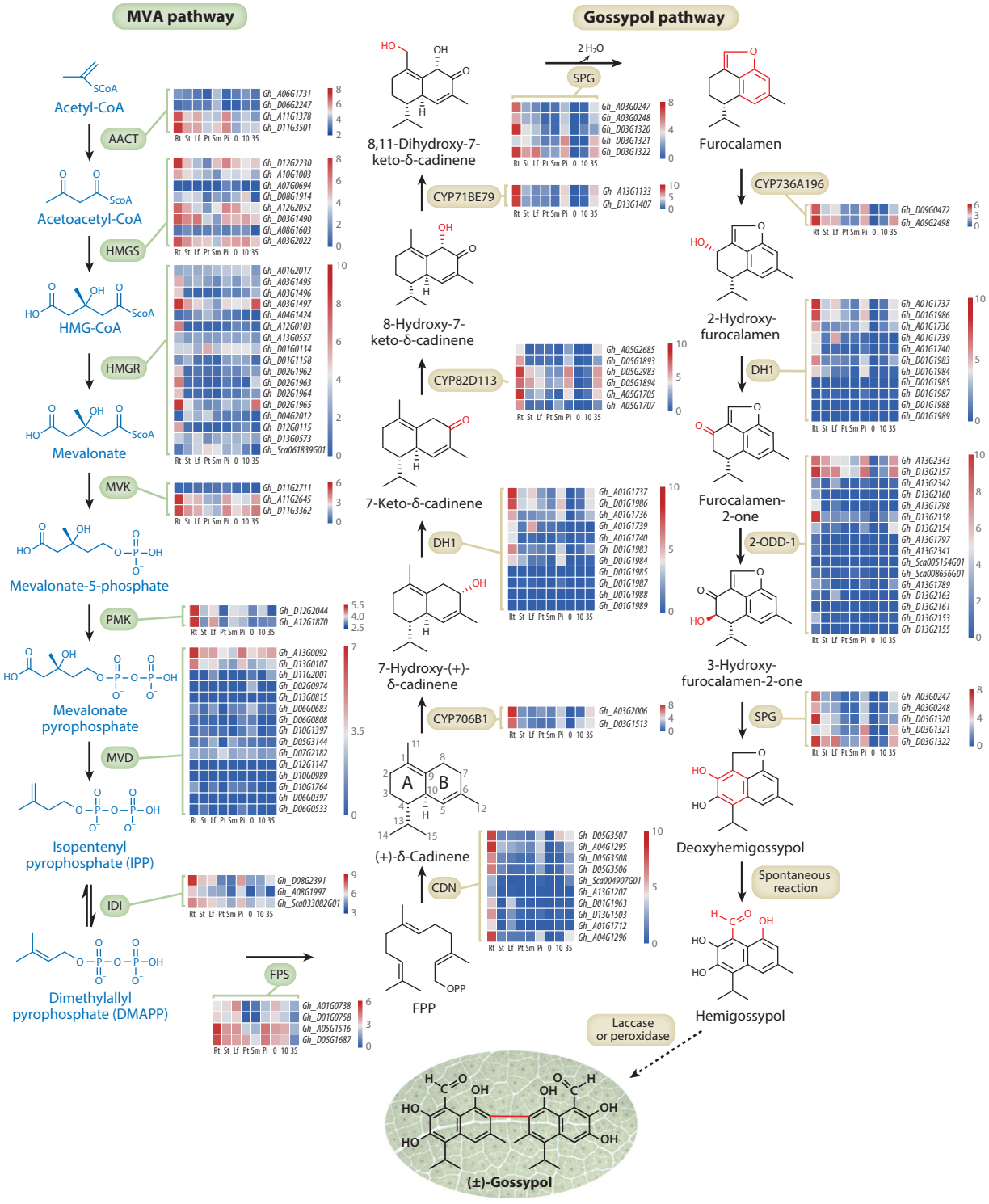
Another type of glandular structure in cotton, named extrafloral nectary, contains an array of metabolites enriched in saccharides and amino acids that function in plant defense against herbivores (29). Quantitative trait locus analysis identified a Phox-Bem1 domain–containing gene, *GaNEC1*, that participates in regulating nectary formation (33). Whether they use packed poisons for deterring predators or sweet juice for engaging ants and other insects, glands of various types are important in mediating plant interactions with other organisms, serving as a driving force for coevolution and biodiversity.

5.2. The Gossypol Biosynthetic Pathway

Cotton plants produce a group of lineage-specific sesquiterpenoids, such as gossypol and hemigossypolone, which have antifungal, antibacterial, or insecticidal activities toward diverse herbivores, including the devastating lepidopterans cotton bollworm and beet armyworm (97, 126). Gossypol is the major, if not only, sesquiterpene phytoalexin present in cotton seeds, while hemigossypolone predominates over gossypol in leaves.

Progress in cotton genomics has facilitated the elucidation of the gossypol biosynthetic pathway from farnesyl diphosphate (FPP) to hemigossypol (98) (**Figure 5**). FPP is synthesized by FPP synthase (FPS) from the cytoplasmic mevalonate and is then catalyzed by (+)-δ-cadinene synthase (CDN) to produce the bicyclic (+)-δ-cadinene with high efficiency (>98%) in cotton (10, 41). Silencing the *CDN* gene in cotton ultimately generated cotton lines with ultralow seed gossypol content (91). Six enzymes, including four P450 monooxygenases (CYP706B1, CYP82D113, CYP71BE79, and CYP736A196), one 2-oxoglutarate/Fe (II)-dependent dioxygenase, and one short-chain alcohol dehydrogenase, have been identified as being responsible for the oxidative modifications of the (+)-δ-cadinene skeleton (**Figure 5**).

Hemigossypol harbors a naphthalene core with two fused benzene rings. A specialized Zn^{2+}-dependent glyoxalase (GLX) variant, SPG, aromatizes both rings of the naphthalene core efficiently without cofactors, completing the biosynthetic pathway for hemigossypol (37, 38) (**Figure 5**). *G. hirsutum* has two GLXI genes, located on chromosome 13 of each subgenome,



(*Caption appears on following page*)

Figure 5 (*Figure appears on preceding page*)

Gossypol pathway genes and their expression profiles, showing enzymes of the MVA and gossypol biosynthetic pathways and their close homologs. Heat maps with the color scale representing $\log_2$-transformed FPKM (fragments per kilobase of gene per million reads) values were created by computing the expression value for each gene (37, 147). The blue chemical structures in the left panel indicate the chemical compounds involved in the MVA pathway. The newly formed groups catalyzed by the gossypol pathway enzymes are shown in red. The dashed arrow indicates an unidentified reaction, and the numbers 0, 10, and 35 under each heat map indicate the DPA of the ovule. Abbreviations: 2-ODD, 2-oxoglutarate/Fe (II)-dependent dioxygenase; AACT, acetoacetyl-CoA thiolase; CDN, (+)-δ-cadinene synthase; DH, short-chain alcohol dehydrogenase; DPA, days postanthesis; FPP, farnesyl diphosphate; FPS, farnesyl diphosphate synthase; HMGR, hydroxymethylglutaryl-CoA reductase; HMGS, hydroxymethylglutaryl-CoA synthase; IDI, isopentenyl diphosphate isomerase; Lf, leaf; MVA, mevalonate; MVD, mevalonate 5-diphosphate decarboxylase; MVK, mevalonate kinase; Pi, pistil; PMK, mevalonate 5-phosphate kinase; Pt, petal; Rt, root; Sm, stamen; SPG, specialized glyoxalase I; St, stem.

and six SPG genes (including a psuedogene), which form two homoeologous pairs in a cluster on chromosome 3. Compared with parental GLXs, SPG has lost the signal peptide for exclusive cytoplasmic localization and the GSH-binding domain to shift the pocket to accommodate cyclic substrates. While the aromatization mechanism seems distinct, the specialization process of SPG represents a textbook example of enzyme evolution: Upon gene duplication and subsequent local duplications, the newly derived copies undergo functional innovations, starting from the catalytic promiscuity of the parental enzyme, along with the removal of unnecessary domains (37).

Oxidative dimerization of hemigossypol into gossypol likely occurs in apoplastic space and requires peroxidase or laccase activities (15). Transgenic *Arabidopsis* plants expressing *GaLAC1*, a root-specific secretory laccase gene from *G. arboreum*, produce significant laccase activity that is able to transform polluting phenolic compounds in soil to non- or less-toxic forms for phytoremediation (106). Because of the restricted rotation of the internaphthyl bond, gossypol exists as a racemic mixture in cotton; the ratio of the two atropisomers, (+)-gossypol and (−)-gossypol, varies with plant organs and among different species and cultivars. For example, the ratio of (+)/(−)-gossypol is ∼3:2 in *G. hirsutum* cultivars grown in the U.S. but can be as high as 98:2 in the same cultivars when grown in Brazil. Dirigent (DIR) proteins steer regio- and enantioselective coupling of phenoxy radicals catalyzed by nonspecific enzymes. *GbDIR2* from *G. barbadense* and *GhDIR3* and *GhDIR4* from *G. hirsutum* were cloned and are known to confer atroposelective synthesis of (+)-gossypol (15).

5.3. Advances in Cotton Resistance to Biotic and Abiotic Stresses

A variety of insect herbivores, such as the Lepidopteran chewing caterpillar, the cotton bollworm, and the *Hemipteran* sucking insect, the cotton aphid, are frequently found in cotton fields. Predation by cotton bollworm induces JA, ethylene, and GA pathway genes and downregulates the salicylic acid (SA) pathway in *G. hirsutum* plants (39). JA is a regulator for mediating insect defense, together with miR156 and its target, SPL9 (65). The JA response is prompt in young plants but shows progressive delays with an age-dependent tendency that has an inverse correlation with the levels of SPL9 group proteins. Herbivorous insects have also evolved complex mechanisms, including the release of effector molecules into the host and the production of diversified detoxification enzymes, to overcome host resistance. Since the cloning of the first effector glucose oxidase from *Helicoverpa zea* (68), several effectors, such as HARP1 (7) and cytochrome P450 monooxygenases (81, 93), have been identified. Not only phytochemicals but also phytohormones (JA and SA) are able to activate insect P450 genes during insect invasion (52, 93). Both structural changes in P450 proteins and upregulations of P450 gene expression have been associated with insect tolerance to gossypol and insecticides (deltamethrin and fenvalerate) (45, 93). When a double-stranded RNA (dsRNA) construct targeting the gossypol-inducible P450 gene *CYP6AE14* (i.e., *dsCYP6AE14*)

Double-stranded RNA (dsRNA): synthetic small hairpin RNAs or short interfering RNAs that silence target gene expression

Supplemental Material >

was introduced into *Arabidopsis* plants, *CYP6AE14* expression was reduced significantly with a concomitant reduction in larval weight gain (64). Similarly, using *Helicoverpa armigera*–targeting *dsNDUFV2* (encoding a mitochondrial complex I subunit), *dsJHAMT* (juvenile hormone acid methyltransferase), *dsJHBP* (JH-binding protein), and *dsFAR* (fatty acyl-CoA reductase), which targets *Miridae* sap-sucking *Adelphocoris suturalis*, to transform cotton plants produced positive results for pest control (**Supplemental Figure 2**). These data suggest that RNAi-triggered signals are successfully delivered into target organisms during pest ingestion or pathogen infection and produce sequence-specific and selective effects for biological invasion control.

Cotton yield losses are caused mainly by the fungal pathogen *Verticillium dahliae* and cotton leaf curl virus disease (CLCuD). Phytohormones SA, JA, and ethylene (126, 133) and nucleotide-binding site (NBS)-leucine-rich repeat (LRR) protein genes (135) are involved in the plant defense responses against *V. dahliae*, which secretes a deacetylase that modifies chitin oligomers to avoid lysin motif (LysM) recognition and prevent pattern-triggered host immunity (19). Plant cells, however, produce chitinase to disrupt the fungal cell wall, whereas *V. dahliae* releases VdSSEP1, a secretory serine protease with hydrolytic activity toward class IV (such as Chi28) cotton chitinase, to protect its cell wall chitin from being digested. The cotton apoplastic protein CRR1 binds to and protects Chi28 from being attacked by VdSSEP1 to maintain the plant defenses (26).

The Ca^{2+}-dependent cysteine protease (Clp-1) and isotrichodermin C-15 hydroxylase of *V. dahliae* determine the virulence of this fungus. Cotton plants export endogenous miRNAs, including miR166 and miR159, via plant vascular systems into *V. dahliae* hyphae that proliferate extensively in vascular tissues, and these miRNAs recognize fungal *Clp-1* and *HiC-15* transcripts and trigger an RNAi-like process (146). Transgenic cotton plants expressing an RNAi construct against another *V. dahliae* virulence gene, *hygrophobins1*, show enhanced resistance to the fungus (145) (**Supplemental Figure 2**). An miRNA-like small RNA from *V. dahliae* was found to epigenetically repress a fungal virulence gene via chromatin remodeling during pathogenicity (118), which may help develop novel cross-kingdom RNAi strategies for plant protection (87, 131).

CLCuD, which is caused by whitefly-transmitted monopartite begomoviruses in association with the modulation of satellites, is widespread and has been a serious problem since the 1980s (63). The sole satellite-encoded protein βC1 was shown to inhibit the host plant ubiquitination process through an interaction with the SKP1/CUL1/F-box complex to impair plant hormone signaling, leading to viral accumulation and symptom aggregation in host plants (43). βC1 also disturbs the plant autophagy pathway via interfering with the cytosolic glyceraldehyde-3-phosphate dehydrogenase and autophagy-related protein 3 in host cells (40). As only *G. arboretum*, among all cultivated cotton species, is immune to current CLCuD viruses, elucidation of the responsible genetic and molecular mechanisms will likely help introduce CLCuD-resistant traits into widely grown allotetraploid cotton cultivars.

Among major commodity crops, the highest portion of cotton (~57%) is grown in water-deficient areas. Overexpression of abscisic acid (ABA) biosynthesis genes, including *ABI3* and *AtABI5*, in cotton plants elevates endogenous ABA levels, which results in enhanced drought tolerance with higher water use efficiency (67). A WRKY TF gene, *GhWRKY59*, confers drought tolerance in cotton plants (48), whereas a genome-wide association study indicates that the sodium/calcium exchanger GaCATX8 is likely responsible for drought tolerance in *G. arboreum* populations (14). Several other genes, including the bZIP TF GhABF2 (55) and aldehyde dehydrogenase *ALDH21* from *Syntrichia caninervis* (134), were also determined in transgenic cotton plants to confer drought stress resistance. This progress inspires further research with regard to adaptation traits obtained from different cotton cultivars or species, which will help cotton field management in arid and semiarid regions.

6. FINAL REMARKS: OPPORTUNITIES, EXCITEMENT, AND FUTURE GOALS

As more high-quality genome sequences are completed for cotton species and cultivars, especially the remaining diploid wild cotton and related plant genomes, along with efficient gene manipulation techniques and the use of other omics datasets, complex regulatory networks and metabolic pathways will be deciphered to provide cotton breeders with functionally new genes and modules for crop improvement. Comparative genomic analysis may help to accurately determine the divergence time among *Gossypium* species, the closest relative of *G. raimondii*, and the time of whole-genome duplications in this tribe. Whether all cultivated and wild tetraploid cotton species originated from the hybridization of the A_0-genome with a *G. raimondii*–like genome and then diversified into the present-day allotetraploid cotton species remains to be determined through further in-depth genomic analyses.

Two unique regulatory mechanisms—networks centered on MML-type TFs (104, 124) and a positive regulation by VLCFA-mediated ethylene biosynthesis (75, 84)—are in operation during cotton fiber development. How these MML and HD-ZIP IV TFs function in cotton fibers, what their major downstream targets are, what crosstalk or coupling mechanisms exist between ethylene and VLCFAs as well as the orchestration with other plant hormones such as BR and GA are questions that all await answers. Cell wall biosynthesis via the *CesA* genes plays an important role in fiber growth, yet direct evidence with regard to specific CesA functions and structures and the organization of various CesA complexes, in addition to the similarities and differences between the fiber cell wall and other cell wall types, is lacking. Genetic mutations of individual *CesA* genes in cotton and high-resolution protein structures of the CesA complex will likely resolve these challenging questions.

SUMMARY POINTS

1. Polyploidization, whole-genome duplication events, and high numbers of transposable element bursts forged the complexity and uniqueness of the cotton genomes.
2. A now-extinct A_0-genome may serve as the common ancestor for all existing diploid A-genomes as well as those present in allotetraploid cotton species.
3. Cotton fiber may elongate via the linear cell-growth mode that relies on a positive regulation of very-long-chain-fatty-acid-mediated ethylene biosynthesis, and the regulatory divergence of key enzymes in this pathway resulted in different fiber phenotypes among various cotton species.
4. Engineered production of RNA interference–inducing double-stranded RNA in cotton plants triggered specific fungal or pest gene silencing that conferred strong pathogen resistance.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Drs. Li Zhang, Ning-Jing Liu, Zhi-Wen Chen, Xin Li, Jun-Feng Cao, Chao-Chen Huang, and Zi-Shou Zheng for helpful discussions and assistance. We apologize to any authors

whose work may not have been referenced owing to length restrictions. This work was supported by grants from the Natural Science Foundation of China (31690090 and 31690091 to Y.-X.Z.), the Young Elite Scientists Sponsorship Program by the Chinese Association for Science and Technology (CAST) (2019QNRC001 to J.-Q.H.), and the National Postdoctoral Program for Innovative Talent (8206200045 to G.H.).

LITERATURE CITED

1. Anderson CT, Kieber JJ. 2020. Dynamic construction, perception, and remodeling of plant cell walls. *Annu. Rev. Plant Biol.* 71:39–69
2. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, et al. 2011. The genome of *Theobroma cacao*. *Nat. Genet.* 43:101–8
3. **Bao Y, Hu GJ, Grover CE, Conover J, Yuan DJ, Wendel JF. 2019. Unraveling *cis* and *trans* regulatory evolution during cotton domestication. *Nat. Commun.* 10:5399**
4. Brill E, van Thournout M, White RG, Llewellyn D, Campbell PM, et al. 2011. A novel isoform of sucrose synthase is targeted to the cell wall during secondary cell wall synthesis in cotton fiber. *Plant Physiol.* 157:40–54
5. Cai Y, Cai X, Wang Q, Wang P, Zhang Y, et al. 2020. Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis. *Plant Biotechnol. J.* 18:814–28
6. Cao JF, Zhao B, Huang CC, Chen ZW, Zhao T, et al. 2020. The miR319-targeted *GhTCP4* promotes the transition from cell elongation to wall thickening in cotton fiber. *Mol. Plant* 13:1063–77
7. Chen CY, Liu YQ, Song WM, Chen DY, Chen FY, et al. 2019. An effector from cotton bollworm oral secretion impairs host plant defense signaling. *PNAS* 116:14331–38
8. Chen E, Zhang X, Yang Z, Wang X, Yang Z, et al. 2017. Genome-wide analysis of the HD-ZIP IV transcription factor family in *Gossypium arboreum* and *GaHDG11* involved in osmotic tolerance in transgenic *Arabidopsis*. *Mol. Genet. Genom.* 292:593–609
9. Chen K, Wang Y, Zhang R, Zhang H, Gao C. 2019. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70:667–97
10. Chen XY, Chen Y, Heinstein P, Davisson VJ. 1995. Cloning, expression, and characterization of (+)-δ-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. *Arch. Biochem. Biophys.* 324:255–66
11. **Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, et al. 2020. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52:525–33**
12. Chen ZW, Grover CE, Li PB, Wang YM, Nie HS, et al. 2017. Molecular evolution of the plastid genome during diversification of the cotton genus. *Mol. Phylogenet. Evol.* 112:268–76
13. Cronn RC, Small RL, Haselkorn T, Wendel JF. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* 89:707–25
14. Du X, Huang G, He S, Yang Z, Sun G, et al. 2018. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* 50:796–802
15. Effenberger I, Harport M, Pfannstiel J, Klaiber I, Schaller A. 2017. Expression in *Pichia pastoris* and characterization of two novel dirigent proteins for atropselective formation of gossypol. *Appl. Microbiol. Biotechnol.* 101:2021–32
16. Fang L, Gong H, Hu Y, Liu C, Zhou B, et al. 2017. Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol.* 18:33
17. Fang L, Wang Q, Hu Y, Jia Y, Chen J, et al. 2017. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49:1089–98
18. Gallagher JP, Grover CE, Rex K, Moran M, Wendel JF. 2017. A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Syst. Bot.* 42:115–23
19. Gao F, Zhang BS, Zhao JH, Huang JF, Jia PS, et al. 2019. Deacetylation of chitin oligomers increases virulence in soil-borne fungal pathogens. *Nat. Plants* 5:1167–76
20. Gao W, Xu FC, Long L, Li Y, Zhang JL, et al. 2020. The gland localized *CGP1* controls gland pigmentation and gossypol accumulation in cotton. *Plant Biotechnol. J.* 18:1573–84

3. *Cis*- and *trans*-regulatory elements during cotton domestication.

11. Assembly of five allotetraploid cotton genomes, altered epigenetic landscapes, recombination suppression, and wild introgression.

21. Gao Y, Wang H, Liu C, Chu H, Dai D, et al. 2018. *De novo* genome assembly of the red silk cotton tree (*Bombax ceiba*). *Gigascience* 7:giy051
22. Gerstel DU. 1953. Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* 7:234–44
23. Grover CE, Arick MA, Thrash A, Conover JL, Sanders WS, et al. 2019. Insights into the evolution of the New World diploid cottons (*Gossypium*, subgenus *Houzingenia*) based on genome sequencing. *Genome Biol. Evol.* 11:53–71
24. Guan XY, Li QJ, Shan CM, Wang S, Mao YB, et al. 2008. The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the *Arabidopsis GLABRA2*. *Physiol. Plant.* 134:174–82
25. Guan XY, Pang MX, Nah G, Shi XL, Ye WX, et al. 2014. miR828 and miR858 regulate homoeologous *MYB2* gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat. Commun.* 5:3050
26. Han LB, Li YB, Wang FX, Wang WY, Liu J, et al. 2019. The cotton apoplastic protein CRR1 stabilizes chitinase 28 to facilitate defense against the fungal pathogen *Verticillium dahliae*. *Plant Cell* 31:520–36
27. Han LB, Li YB, Wang HY, Wu XM, Li CL, et al. 2013. The dual functions of WLIM1a in cell elongation and secondary wall formation in developing cotton fibers. *Plant Cell* 25:4421–38
28. Hao J, Tu L, Hu H, Tan J, Deng F, et al. 2012. GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *J. Exp. Bot.* 63:6267–81
29. Heil M. 2011. Nectar: generation, regulation, and ecological functions. *Trends Plant Sci.* 16:191–200
30. Hovav R, Chaudhary B, Udall JA, Flagel L, Wendel JF. 2008. Parallel domestication, convergent evolution and duplicated gene recruitment in allopolyploid cotton. *Genetics* 179:1725–33
31. Hu HY, He X, Tu LL, Zhu LF, Zhu ST, et al. 2016. GhJAZ2 negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor GhMYB25-like. *Plant J.* 88:921–35
32. Hu HY, Wang MJ, Ding YH, Zhu ST, Zhao GN, et al. 2018. Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 16:1002–12
33. Hu W, Qin W, Jin Y, Wang P, Yan Q, et al. 2020. Genetic and evolution analysis of extrafloral nectary in cotton. *Plant Biotechnol. J.* 18:2081–95
34. Hu Y, Chen J, Fang L, Zhang Z, Ma W, et al. 2019. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51:739–48
35. Huang G, Wu Z, Percy RG, Bai M, Li Y, et al. 2020. Genome sequence of *Gossypium herbaceum*, and genome update of *G. arboreum* and *G. hirsutum* provide insights into cotton A-genome evolution. *Nat. Genet.* 52:516–24
36. Huang G, Zhu YX. 2019. Plant polyploidy and evolution. *J. Integr. Plant Biol.* 61:4–6
37. Huang JQ, Fang X, Tian X, Chen P, Lin JL, et al. 2020. Aromatization of natural products by a specialized detoxification enzyme. *Nat. Chem. Biol.* 16:250–56
38. Huang JQ, Lin JL, Guo XX, Tian X, Tian Y, et al. 2020. RES transformation for biosynthesis and detoxification. *Sci. China Life Sci.* 63:1–6
39. Huang XZ, Chen JY, Xiao HJ, Xiao YT, Wu J, et al. 2015. Dynamic transcriptome analysis and volatile profiling of *Gossypium hirsutum* in response to the cotton bollworm *Helicoverpa armigera*. *Sci. Rep.* 5:11867
40. Ismayil A, Yang M, Haxim Y, Wang Y, Li J, et al. 2020. *Cotton leaf curl Multan virus* βC1 protein induces autophagy by disrupting the interaction of autophagy-related protein 3 with glyceraldehyde-3-phosphate dehydrogenases. *Plant Cell* 32:1124–35
41. Jacobowitz JR, Weng JK. 2020. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annu. Rev. Plant Biol.* 71:631–58
42. Janga MR, Pandeya D, Campbell LM, Konganti K, Villafuerte ST, et al. 2019. Genes regulating gland development in the cotton plant. *Plant Biotechnol. J.* 17:1142–53
43. Jia Q, Liu N, Xie K, Dai Y, Han S, et al. 2016. CLCuMuB βC1 subverts ubiquitination by interacting with NbSKP1s to enhance geminivirus infection in *Nicotiana benthamiana*. *PLOS Pathog.* 12:e1005668
44. Jiang YJ, Guo WZ, Zhu HY, Ruan YL, Zhang TZ. 2012. Overexpression of *GhSusA1* increases plant biomass and improves cotton fiber yield and quality. *Plant Biotechnol. J.* 10:301–12
45. Joußen N, Agnolet S, Lorenz S, Schöne SE, Ellinger R, et al. 2012. Resistance of Australian *Helicoverpa armigera* to fenvalerate is due to the chimeric P450 enzyme CYP337B3. *PNAS* 109:15206–11
46. Kumar V, Singh B, Singh SK, Rai KM, Singh SP, et al. 2018. Role of GhHDA5 in H3K9 deacetylation and fiber initiation in *Gossypium hirsutum*. *Plant J.* 95:1069–83

35. Resolved controversial concepts surrounding the A-genome donor of diploid A-genomes and other allotetraploid cotton species.

37. The distinct aromatization mechanism evolved from a detoxification pathway in cotton.

47. Lee J, Burns TH, Light G, Sun Y, Fokar M, et al. 2010. Xyloglucan endotransglycosylase/hydrolase genes in cotton and their role in fiber elongation. *Planta* 232:1191–205
48. Li F, Li M, Wang P, Cox KL Jr., Duan L, et al. 2017. Regulation of cotton (*Gossypium hirsutum*) drought responses by mitogen-activated protein (MAP) kinase cascade-mediated phosphorylation of GhWRKY59. *New Phytol.* 215:1462–75
49. Li FG, Fan GY, Lu CR, Xiao GH, Zou CS, et al. 2015. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33:524–30
50. Li FG, Fan GY, Wang K, Sun F, Yuan Y, et al. 2014. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* 46:567–72
51. Li L, Wang XL, Huang GQ, Li XB. 2007. Molecular characterization of cotton *GhTUA9* gene specifically expressed in fibre and involved in cell elongation. *J. Exp. Bot.* 58:3227–38
52. Li X, Schuler MA, Berenbaum MR. 2002. Jasmonate and salicylate induce expression of herbivore cytochrome P450 genes. *Nature* 419:712–15
53. 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* 17:859–75
54. Li Y, Wang NN, Wang Y, Liu D, Gao Y, et al. 2018. The cotton XLIM protein (GhXLIM6) is required for fiber development via maintaining dynamic F-actin cytoskeleton and modulating cellulose biosynthesis. *Plant J.* 96:1269–82
55. Liang CZ, Meng ZH, Meng ZG, Malik W, Yan R, et al. 2016. GhABF2, a bZIP transcription factor, confers drought and salinity tolerance in cotton (*Gossypium hirsutum* L.). *Sci. Rep.* 6:35040
56. Liu B, Zhu Y, Zhang T. 2015. The R3-MYB gene *GhCPC* negatively regulates cotton fiber elongation. *PLOS ONE* 10:e0116272
57. Liu NA, Tu LL, Tang WX, Gao WH, Lindsey K, Zhang XL. 2014. Small RNA and degradome profiling reveals a role for miRNAs and their targets in the developing fibers of *Gossypium barbadense*. *Plant J.* 80:331–44
58. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, et al. 2015. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Sci. Rep.* 5:14139
59. Liu ZH, Chen Y, Wang NN, Chen YH, Wei N, et al. 2020. A basic helix-loop-helix protein (GhFP1) promotes fibre elongation of cotton (*Gossypium hirsutum*) by modulating brassinosteroid biosynthesis and signalling. *New Phytol.* 225:2439–52
60. Luo M, Xiao Y, Li X, Lu X, Deng W, et al. 2007. GhDET2, a steroid 5α-reductase, plays an important role in cotton fiber cell initiation and elongation. *Plant J.* 51:419–30
61. Ma D, Hu Y, Yang CQ, Liu BL, Fang L, et al. 2016. Genetic basis for glandular trichome formation in cotton. *Nat. Commun.* 7:10456
62. Ma Z, He S, Wang X, Sun J, Zhang Y, et al. 2018. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50:803–13
63. Mansoor S, Briddon RW, Zafar Y, Stanley J. 2003. Geminivirus disease complexes: an emerging threat. *Trends Plant Sci.* 8:128–34
64. Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, et al. 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* 25:1307–13
65. Mao YB, Liu YQ, Chen DY, Chen FY, Fang X, et al. 2017. Jasmonate response decay and defense metabolite accumulation contributes to age-regulated dynamics of plant insect resistance. *Nat. Commun.* 8:13925
66. McMichael SC. 1960. Combined effects of glandless genes *gl2* and *gl3* on pigment glands in cotton plant. *Agron. J.* 52:385–86
67. Mittal A, Gampala SS, Ritchie GL, Payton P, Burke JJ, Rock CD. 2014. Related to ABA-insensitive3 (ABI3)/Viviparous1 and AtABI5 transcription factor coexpression in cotton enhances drought stress adaptation. *Plant Biotechnol. J.* 12:578–89
68. Musser RO, Hum-Musser SM, Eichenseer H, Peiffer M, Ervin G, et al. 2002. Herbivory: Caterpillar saliva beats plant defences. *Nature* 416:599–600
69. Pang C-Y, Wang H, Pang Y, Xu C, Jiao Y, et al. 2010. Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and *Arabidopsis* root hair elongation. *Mol. Cell. Proteom.* 9:2019–33

61. The basic helix-loop-helix-type transcription factor GoPGF as the positive regulator of glandular trichome formation in cotton.

70. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, et al. 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492:423–27
71. Preuss ML, Kovar DR, Lee YRJ, Staiger CJ, Delmer DP, Liu B. 2004. A plant-specific kinesin binds to actin microfilaments and interacts with cortical microtubules in cotton fibers. *Plant Physiol.* 136:3945–55
72. Pu L, Li Q, Fan XP, Yang WC, Xue YB. 2008. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics* 180:811–20
73. Purushotham P, Ho R, Zimmer J. 2020. Architecture of a catalytically active homotrimeric plant cellulose synthase complex. *Science* 369:1089–94
74. Qi TC, Wang JJ, Huang H, Liu B, Gao H, et al. 2015. Regulation of jasmonate-induced leaf senescence by antagonism between bHLH subgroup IIIe and IIId factors in *Arabidopsis*. *Plant Cell* 27:1634–49
75. Qin YM, Hu CY, Pang Y, Kastaniotis AJ, Hiltunen JK, Zhu YX. 2007. Saturated very-long-chain fatty acids promote cotton fiber and *Arabidopsis* cell elongation by activating ethylene biosynthesis. *Plant Cell* 19:3692–704
76. Qin YM, Zhu YX. 2011. How cotton fibers elongate: a tale of linear cell-growth mode. *Curr. Opin. Plant Biol.* 14:106–11
77. Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10:63–70
78. Renny-Byfield S, Page JT, Udall JA, Sanders WS, Peterson DG, et al. 2016. Independent domestication of two Old World cotton species. *Genome Biol. Evol.* 8:1940–47
79. Ruan YL, Llewellyn DJ, Furbank RT. 2003. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *Plant Cell* 15:952–64
80. Salih H, Gong WF, He SP, Sun GF, Sun JL, Du XM. 2016. Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. *BMC Genet.* 17:129
81. Schuler MA. 2012. Insect P450s: mounted for battle in their war against toxins. *Mol. Ecol.* 21:4157–59
82. Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, et al. 2014. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nat. Commun.* 5:5519
83. Shangguan XX, Yang CQ, Zhang XF, Wang LJ. 2016. Functional characterization of a basic helix-loop-helix (bHLH) transcription factor GhDEL65 from cotton (*Gossypium hirsutum*). *Physiol. Plant.* 158:200–12
84. Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, et al. 2006. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* 18:651–64
85. Singh B, Avci U, Eichler Inwood SE, Grimson MJ, Landgraf J, et al. 2009. A specialized outer layer of the primary cell wall joins elongating cotton fibers into tissue-like bundles. *Plant Physiol.* 150:684–99
86. Song Q, Zhang T, Stelly DM, Chen ZJ. 2017. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* 18:99
87. Song XW, Li Y, Cao XF, Qi YJ. 2019. MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.* 70:489–525
88. Stephens SG. 1944. Phenogenetic evidence for the amphidiploid origin of New World cottons. *Nature* 153:53–54
89. Stiff MR, Haigler CH. 2016. Cotton fiber tips have diverse morphologies and show evidence of apical cell wall synthesis. *Sci. Rep.* 6:27883
90. Sun W, Gao Z, Wang J, Huang Y, Chen Y, et al. 2019. Cotton fiber elongation requires the transcription factor GhMYB212 to regulate sucrose transportation into expanding fibers. *New Phytol.* 222:864–81
91. Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS. 2006. Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *PNAS* 103:18054–59
92. Tang C, Min LF, Zhang TZ, Pan JJ, Jing SR, et al. 1996. Genetic analysis for Hai-1 strain of glandless cotton (*G. barbadense* L.): interaction between Gl_2^e and *Gl1*. *Cotton Sci. Sin.* 8:138–40
93. Tao XY, Xue XY, Huang YP, Chen XY, Mao YB. 2012. Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide. *Mol. Ecol.* 21:4371–85
94. Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, et al. 2015. An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. *Nature* 517:571–75

70. Multiple-fold ploidization in the *Gossypium* lineage conferred large gene duplications and genetic complexity.

76. A linear cell-growth mode for cotton fiber cell elongation.

95. Teh BT, Lim K, Yong CH, Ng CCY, Rao SR, et al. 2017. The draft genome of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.* 49:1633–41
96. Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, et al. 2017. A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. *Plant J.* 90:111–21
97. Tian X, Ruan JX, Huang JQ, Fang X, Mao Y, et al. 2016. Gossypol: phytoalexin of cotton. *Sci. China Life Sci.* 59:122–29
98. Tian X, Ruan JX, Huang JQ, Yang CQ, Fang X, et al. 2018. Characterization of gossypol biosynthetic pathway. *PNAS* 115:E5410–18
99. Tian Y, Du J, Wu H, Guan X, Chen W, et al. 2020. The transcription factor MML4_D12 regulates fiber development through interplay with the WD40-repeat protein WDR in cotton. *J. Exp. Bot.* 71:3499–511
100. Tiwari SC, Wilkins TA 1995. Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. *Can. J. Bot.* 73:746–57
101. Udall JA, Long E, Hanson C, Yuan DJ, Ramaraj T, et al. 2019. *De novo* genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. *G3* 9:3079–85
102. Udall JA, Long E, Ramaraj T, Conover JL, Yuan DJ, et al. 2019. The genome sequence of *Gossypioides kirkii* illustrates a descending dysploidy in plants. *Front. Plant Sci.* 10:1541
103. Walford SA, Wu YR, Llewellyn DJ, Dennis ES. 2011. GhMYB25-like: a key factor in early cotton fibre development. *Plant J.* 65:785–97
104. Wan Q, Guan X, Yang N, Wu H, Pan M, et al. 2016. Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. *New Phytol.* 210:1298–310
105. Wang D, Fan W, Guo X, Wu K, Zhou S, et al. 2019. MaGenDB: a functional genomics hub for Malvaceae plants. *Nucleic Acids Res.* 48:D1076–84
106. Wang GD, Li QJ, Luo B, Chen XY. 2004. Ex planta phytoremediation of trichlorophenol and phenolic allelochemicals via an engineered secretory laccase. *Nat. Biotechnol.* 22:893–97
107. Wang HY, Wang J, Gao P, Jiao GL, Zhao PM, et al. 2009. Down-regulation of *GhADF1* gene expression affects cotton fibre properties. *Plant Biotechnol. J.* 7:13–23
108. Wang JA, Wang HY, Zhao PM, Han LB, Jiao GL, et al. 2010. Overexpression of a profilin (GhPFN2) promotes the progression of developmental phases in cotton fibers. *Plant Cell Physiol.* 51:1276–90
109. Wang K, Huang G, Zhu Y. 2016. Transposable elements play an important role during cotton genome evolution and fiber cell development. *Sci. China Life Sci.* 59:112–21
110. Wang K, Wang D, Zheng X, Qin A, Zhou J, et al. 2019. Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. *Nat. Commun.* 10:4714
111. Wang K, Wang Z, Li F, Ye W, Wang J, et al. 2012. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44:1098–103
112. Wang LL, Cheng H, Xiong FJ, Ma SY, Zheng L, et al. 2020. Comparative phosphoproteomic analysis of BR-defective mutant reveals a key role of GhSK13 in regulating cotton fiber development. *Sci. China Life Sci.* 63:1905–17
113. Wang MJ, Tu LL, Lin M, Lin Z, Wang PC, et al. 2017. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat. Genet.* 49:579–87
114. Wang MJ, Tu LL, Yuan D, Zhu D, Shen C, et al. 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* 51:224–29
115. Wang MJ, Wang PC, Tu LL, Zhu ST, Zhang L, et al. 2016. Multi-omics maps of cotton fibre reveal epigenetic basis for staged single-cell differentiation. *Nucleic Acids Res.* 44:4067–79
116. Wang MY, Zhao PM, Cheng HQ, Han LB, Wu XM, et al. 2013. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiol.* 162:1669–80
117. Wang S, Wang JW, Yu N, Li CH, Luo B, et al. 2004. Control of plant trichome development by a cotton fiber MYB gene. *Plant Cell* 16:2323–34
118. Wang S, Wu XM, Liu CH, Shang JY, Gao F, Guo HS. 2020. *Verticillium dahliae* chromatin remodeling facilitates the DNA damage repair in response to plant ROS stress. *PLOS Pathog.* 16:e1008481
119. Wang Z, Yang Z, Li F. 2019. Updates on molecular mechanisms in the development of branched trichome in *Arabidopsis* and nonbranched in cotton. *Plant Biotechnol. J.* 17:1706–22

111. The first sequenced cotton genome.

120. Watt G. 1907. *The Wild and Cultivated Cotton Plants of the World*. London: Longmans
121. Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. 2009. Evolution and natural history of the cotton genus. In *Genetics and Genomics of Cotton*, ed. AH Paterson, pp. 3–22. New York: Springer-Verlag
122. Westengen OT, Huaman Z, Heun M. 2005. Genetic diversity and geographic pattern in early South American cotton domestication. *Theor. Appl. Genet.* 110:392–402
123. Wong GK, Soltis DE, Leebens-Mack J, Wickett NJ, Barker MS, et al. 2020. Sequencing and analyzing the transcriptomes of a thousand species across the tree of life for green plants. *Annu. Rev. Plant Biol.* 71:741–65
124. Wu H, Tian Y, Wan Q, Fang L, Guan X, et al. 2018. Genetics and evolution of *MIXTA* genes regulating cotton lint fiber development. *New Phytol.* 217:883–95
125. Wu Y, Liu F, Yang DG, Li W, Zhou XJ, et al. 2018. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Front. Plant Sci.* 9:376
126. Xia Y, Huang G, Zhu Y. 2019. Sustainable plant disease control: biotic information flow and behavior manipulation. *Sci. China Life Sci.* 62:1710–13
127. Xiao GH, Wang K, Huang G, Zhu YX. 2016. Genome-scale analysis of the cotton *KCS* gene family revealed a binary mode of action for gibberellin A regulated fiber growth. *J. Integr. Plant Biol.* 58:577–89
128. Xiao GH, Zhao P, Zhang Y. 2019. A pivotal role of hormones in regulating cotton fiber development. *Front. Plant Sci.* 10:87
129. Xiao YH, Li DM, Yin MH, Li XB, Zhang M, et al. 2010. Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. *J. Plant Physiol.* 167:829–37
130. Xu B, Gou JY, Li FG, Shangguan XX, Zhao B, et al. 2013. A cotton BURP domain protein interacts with α-expansin and their co-expression promotes plant growth and fruit production. *Mol. Plant* 6:945–58
131. Xue Y, Chen R, Qu L, Cao X. 2020. Noncoding RNA: from dark matter to bright star. *Sci. China Life Sci.* 63:463–68
132. Yan Q, Wang Y, Li Q, Zhang ZS, Ding H, et al. 2018. Up-regulation of *GhTT2-3A* in cotton fibres during secondary wall thickening results in brown fibres with improved quality. *Plant Biotechnol. J.* 16:1735–47
133. Yang CL, Liang S, Wang HY, Han LB, Wang FX, et al. 2015. Cotton major latex protein 28 functions as a positive regulator of the ethylene responsive factor 6 in defense against *Verticillium dahliae*. *Mol. Plant* 8:399–411
134. Yang H, Zhang D, Li X, Li H, Zhang D, et al. 2016. Overexpression of *ScALDH21* gene in cotton improves drought tolerance and growth in greenhouse and field conditions. *Mol. Breeding* 36:34
135. Yang Y, Chen T, Ling X, Ma Z. 2017. *Gbvdr6*, a gene encoding a receptor-like protein of cotton (*Gossypium barbadense*), confers resistance to verticillium wilt in *Arabidopsis* and Upland cotton. *Front. Plant Sci.* 8:2272
136. Yang Z, Ge X, Yang Z, Qin W, Sun G, et al. 2019. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* 10:2989
137. Yang Z, Qanmber G, Wang Z, Yang Z, Li F. 2020. *Gossypium* genomics: trends, scope, and utilization for cotton improvement. *Trends Plant Sci.* 25:488–500
138. Yang ZR, Zhang CJ, Yang XJ, Liu K, Wu ZX, et al. 2014. *PAG1*, a cotton brassinosteroid catabolism gene, modulates fiber elongation. *New Phytol.* 203:437–48
139. Yu Y, Wu S, Nowak J, Wang G, Han L, et al. 2019. Live-cell imaging of the cytoskeleton in elongating cotton fibres. *Nat. Plants* 5:498–504
140. Yuan DJ, Tang ZH, Wang MJ, Gao WH, Tu LL, et al. 2015. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* 5:17662
141. Zhang J, Huang GQ, Zou D, Yan JQ, Li Y, et al. 2018. The cotton (*Gossypium hirsutum*) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. *New Phytol.* 217:625–40
142. Zhang L, Xu Y, Zhang X, Ma X, Zhang L, et al. 2020. The genome of kenaf (*Hibiscus cannabinus* L.) provides insights into bast fibre and leaf shape biogenesis. *Plant Biotechnol. J.* 18:1796–809
143. Zhang M, Han LB, Wang WY, Wu SJ, Jiao GL, et al. 2017. Overexpression of *GhFIM2* propels cotton fiber development by enhancing actin bundle formation. *J. Integr. Plant Biol.* 59:531–34

139. Live-cell imaging in cotton confirmed that fiber cells elongate via the unique linear cell-growth or tip-biased diffuse growth mode.

144. Zhang M, Zheng XL, Song SQ, Zeng QW, Hou L, et al. 2011. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat. Biotechnol.* 29:453–58
145. Zhang T, Jin Y, Zhao JH, Gao F, Zhou BJ, et al. 2016. Host-induced gene silencing of the target gene in fungal cells confers effective resistance to the cotton wilt disease pathogen *Verticillium dahliae*. *Mol. Plant* 9:939–42
146. Zhang T, Zhao YL, Zhao JH, Wang S, Jin Y, et al. 2016. Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. *Nat. Plants* 2:16153
147. Zhang TZ, Hu Y, Jiang WK, Fang L, Guan XY, et al. 2015. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33:531–37
148. Zhao B, Cao JF, Hu GJ, Chen ZW, Wang LY, et al. 2018. Core *cis*-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *New Phytol.* 218:1061–75
149. Zhao T, Tao XY, Feng SL, Wang LY, Hong H, et al. 2018. LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol.* 19:195
150. Zhou Y, Zhang ZT, Li M, Wei XZ, Li XJ, et al. 2015. Cotton (*Gossypium hirsutum*) 14-3-3 proteins participate in regulation of fibre initiation and elongation by modulating brassinosteroid signalling. *Plant Biotechnol. J.* 13:269–80

146. Endogenous miRNAs are exported from cotton plants into the fungal pathogen *Verticillium dahliae* to activate fungus–host interactions.