PLOS ONE

RESEARCH ARTICLE

Phenomics and transcriptomics analyses reveal deposition of suberin and lignin in the short fiber cell walls produced from a wild cotton species and two mutants

Hee Jin Kim[ID]1*, Yongliang Liu2, Gregory N. Thyssen[ID]1, Marina Naoumkina[ID]1, James Frelichowski3

1 USDA-ARS, Southern Regional Research Center, Cotton Fiber Bioscience Research Unit, New Orleans, LA, United States of America, 2 USDA-ARS, Southern Regional Research Center, Cotton Structure and Quality Research Unit, New Orleans, LA, United States of America, 3 USDA-ARS-SPARC, Crop Germplasm Research Unit, College Station, TX, United States of America

* heejin.kim@usda.gov

Abstract

Fiber length is one of the major properties determining the quality and commercial value of cotton. To understand the mechanisms regulating fiber length, genetic variations of cotton species and mutants producing short fibers have been compared with cultivated cottons generating long and normal fibers. However, their phenomic variation other than fiber length has not been well characterized. Therefore, we compared physical and chemical properties of the short fibers with the long fibers. Fiber characteristics were compared in two sets: 1) wild diploid *Gossypium raimondii* Ulbrich (short fibers) with cultivated diploid *G. arboreum* L and tetraploid *G. hirsutum* L. (long fibers); 2) *G. hirsutum* short fiber mutants, Ligon-lintless 1 (Li_1) and 2 (Li_2) with their near isogenic line (NIL), DP-5690 (long fibers). Chemical analyses showed that the short fibers commonly consisted of greater non-cellulosic components, including lignin and suberin, than the long fibers. Transcriptomic analyses also identified up-regulation of the genes related to suberin and lignin biosynthesis in the short fibers. Our results may provide insight on how high levels of suberin and lignin in cell walls can affect cotton fiber length. The approaches combining phenomic and transcriptomic analyses of multiple sets of cotton fibers sharing a common phenotype would facilitate identifying genes and common pathways that significantly influence cotton fiber properties.

Introduction

Cotton (*Gossypium sp.*) is the most economically important natural fiber in the world [1]. In addition to the agronomic importance, cotton fibers are also utilized as an ideal biological model for studying molecular mechanisms involved in cell elongation and cell wall biogenesis because cotton fiber cells are unicellular and larger and longer than any other plant cell [2]. Cotton fiber development is divided into four overlapping stages: 1) initiation, 2) primary cell wall (PCW) biosynthesis characterized by fiber elongation, 3) secondary cell wall (SCW)

biosynthesis characterized by cellulose production, and 4) maturation process [3]. The overlapping period between PCW and SCW stages is often classified as a transition stage due to its importance for fiber cell wall biosynthesis. Genotypes and growth conditions greatly affect the period of each developmental stage. Cytochemical analyses showed that cultivated *G. hirsutum* fibers exhibit thickened SCW consisting of nearly pure cellulose [4, 5]. Chemical analyses detected very low levels (0.4–1% of dried weight) of lignin-like phenolics from cultivated *G. hirsutum* fibers [6].

Fiber length is a major physical property determining fiber quality [7]. Wild diploid *G. raimondii* (D_5 genome) produces short fibers that are agronomically inferior to cultivated diploid *G. arboreum* (A_2 genome) as well as cultivated polyploid *G. hirsutum* (AD_1 genome) [8]. Among the wild diploid cotton species, *G. raimondii* is considered as the closest extant relative to the D_T subgenome of the most widely cultivated polyploid *G. hirsutum* (AD_1 genome) [9]. It also shares cytological, morphological, and phenogenetical similarities to the polyploid cotton species. Thus, the D_5 genome of wild diploid *G. raimondii* was the first sequenced among cotton species [10–12]. *G. arboreum* is suggested to be the closest extant relative of the A_T subgenome of the polyploid cotton species [9]. The A_2 genome of cultivated diploid *G. arboreum* SXY1 was sequenced [13]. The AD_1 genome was first sequenced from cultivated polyploid *G. hirsutum* TM-1 [14, 15]. Those cotton genome sequences with their transcriptomic data have been available in public databases [16], and extensively utilized to study natural diversity and evolution of cotton species as well as cotton genetic research.

In polyploid *G. hirsutum*, two short fiber mutants, Ligon-lintless 1 (Li_1) [17] and Ligon-lintless 2 (Li_2) [18] have been reported. Li_1 mutant shows pleiotropic effects on both fiber and non-fiber tissues, whereas Li_2 mutant phenotype is specifically detected in fiber tissue. *G. hirsutum* short fiber mutant plants containing Li_1 [19] or Li_2 [20] gene were crossed with *G. hirsutum* DP-5690 plants producing long fibers. F_1 progeny were backcrossed for five generations (BC_5) by single seed decent (SSD) to DP-5690 which served as the recurrent parent in each backcross. The two short fiber mutants with their NIL, DP-5690 have been used as a model system to study fiber elongation [21–25]. Our group has shown that mutations in the *actin* [25] and in the putative *Ran Binding Protein 1* involved in nucleocytoplasmic transport [23] are responsible for the short fiber phenotypes of Li_1 and Li_2 mutants, respectively.

The short fiber phenotypes are commonly observed in polyploid *G. hirsutum* Li_1 and Li_2 mutants and wild diploid *G. raimondii* (Fig 1). However, it is not known if the short fibers share any common mechanisms to impair fiber elongation. In this study, we determined and compared the chemical compositions and transcriptomic profiles of two sets of cotton fibers. The first set (Fig 1A) consisted of three cotton species, wild diploid *G. raimondii* (short fibers), diploid cultivated *G. arboreum* (long fibers) and polyploid cultivated *G. hirsutum* (long fibers). The second set (Fig 1B) was composed of *G. hirsutum* Li_1 (short fibers), Li_2 (short fibers), and their NIL, *G. hirsutum* DP-5690 (long fibers). Integrations of chemical properties and transcriptomic profiles of the two sets showed that suberin and lignin are commonly associated with the short fiber phenotypes.

Materials and methods

Plant materials and growth conditions

This research was approved by the biosafety committee of USDA-ARS-SRRC. Field research was performed according to the policy and practices of USDA-ARS. The cottonseeds with Plant Introduction (PI) or Plant Variety Protection (PVP) numbers including diploid *G. arboreum* A_2-100 (PI 529728), and *G. raimondii* D_5-6 (PI530903) and D_5-31 (PI 530928) as well as polyploid *G. hirsutum* Texas Marker 1 (TM-1, PI 607172), Sure-Grow 747 (SG-747, PVP
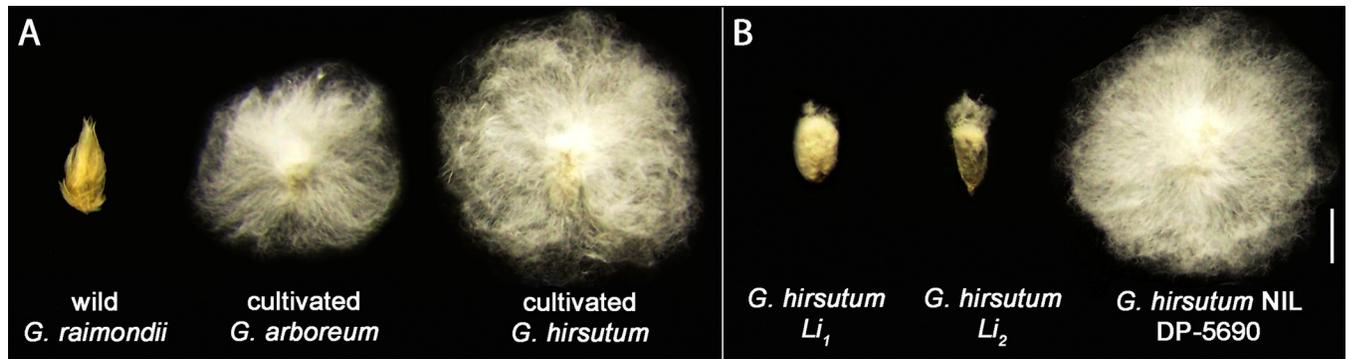
Fig 1. Morphology of two sets of cottonseeds for an integrative analysis. A. The 1st set is composed of three cotton species including wild diploid *G. raimondii* D_5-31 (D_5 genome), cultivated diploid *G. arboreum* A_2-100 (A_2 genome), and cultivated polyploid *G. hirsutum* TM-1 (AD_1 genome). B. The 2nd set consists of *G. hirsutum* Li_1 and Li_2 mutants producing short fibers along with their near isogenic line (NIL) DP-5690 producing long fibers. The single seeds with associating cotton fibers were harvested from the open bolls, combed and photographed. The bars represent 1 cm in length.

https://doi.org/10.1371/journal.pone.0282799.g001

9800118) and Delta Pine 5690 (DP-5690, PVP 9100116) were obtained from the U.S. National Cotton Germplasm Collection (NCGC). The cottonseeds of short fiber mutants, Ligon-lintless 1 (*Li₁*) mutant, Ligon-lintless 2 (*Li₂*), and diploid *G. arboreum* Shixiya1 (SXY1) were provided by Dr. Rickie Turley of USDA-ARS-SEA and Dr. Xianliang Song of Shandong Agricultural University, China.

Each variety of *G. arboreum* (A₂-100 and SXY1) and *G. hirsutum* (TM-1, SG-747, DP-5690, *Li₁*, and *Li₂*) were planted on two-row plots located at the Southern Regional Research Center (New Orleans, LA; 2017) with naturally neutral-day conditions. The soil type of the cotton plot was aquents dredged over alluvium in an elevated location to provide adequate drainage. Single row plots were 12 m long with approximately 40 plants per plot. The distance between two rows was 0.5 m, and the distance between two plants within a row was 0.3 m. To minimize environmental effects, boll samples were not collected from plants on the perimeter of the field and the end of each row. At harvest, approximately 60 naturally opened bolls were randomly collected from two plots for each cotton variety, and separated into two biological replicates with 30 bolls per biological replicate for further analyzing physical and chemical properties of each cotton variety.

To collect developing fibers at various developmental stages, wild diploid *G. raimondii* D_5-6 and D_5-31 along with *G. arboreum* and *G. hirsutum* were grown in a growth chamber (Percival Intellus Environmental Controller, Perry, IA) in 8 L pots at 28°C (day) / 24°C (night) with a short photoperiod condition (9h day light, 300 $\mu molm^{-2}\ s^{-1}$) during the vegetative stage, and reduced to 26°C (day)/ 18°C (night) during flowering and boll development stages. The pots were filled with Metro-Mix 350 soil. For fiber length measurement, two plants of wild diploid *G. raimondii* D_5-6 were grown in 167 L containers at an NCGC greenhouse located at College Station, Texas during a winter season for a short photoperiod condition. The two *G. raimondii* D_5-6 plants produced four bolls, and separated into two biological replicates with two bolls per biological replicate. To obtain sufficient *G. raimondii* fibers for fiber length and chemical analyses, *G. raimondii* D_5-31 was grown perennially at the cotton winter nursery at Tecoman, Colima, Mexico in association with the location of the Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias [26]. Three *G. raimondii* D_5-31 plants (240 days after planting) were transplanted on the ground of the cotton winter nursery. In the second year, they produced 400 bolls that were separated into two biological replicates with 200 bolls per biological replicate for further analyzing fiber length and chemical analyses. All *G. raimondii* grown

in the growth chamber, greenhouse, and cotton winter nursery produced a common phenotype demonstrating short and green colored fibers.

Cotton fiber length measurements

Maximum fiber lengths were estimated by placing ovules on a watch-glass and gently spraying fibers with a stream of distilled water as described by Schubert et al. [27]. Ten to thirty cotton bolls were randomly selected from each biological replicate samples of *G. arboreum* (A_2-100 and SXY1), *G. raimondii* (D_5-6 and D_5-31) and *G. hirsutum* (TM-1, SG-747, DP-5690, Li_1, and Li_2). Single cotton seeds were randomly selected from an individual cotton boll. The distance between the chalazal end of the selected seeds and the tip of the spread fibers were measured to the nearest 0.1 mm with a digital caliper. Mean maximum fiber length of each cotton variety was obtained by measuring the randomly selected seeds from two biological replicates.

Updegraff cellulose assay

Cellulose contents of developed cotton fibers were measured by the modified Updegraff method [28]. Five cotton bolls were randomly selected from each biological replicate samples of the cultivated *G. arboreum* (A_2-100 and SXY1) and *G. hirsutum* (TM-1, SG-747, and DP-5690) producing long fibers. Two to six cotton bolls were also randomly selected from each biological replicate samples of the *G. raimondii* (D_5-6 and D_5-31) and *G. hirsutum* mutants (Li_1 and Li_2) generating short fibers. Dried fiber samples of the selected bolls were manually harvested, and cut into small pieces. Ten milligrams of the blended fibers were placed in 5 mL Reacti-Vials™ (Thermol Fisher Scientific, Waltham, MA), and hydrolyzed with acetic-nitric reagent (a mixture of 73% acetic acid, 9% nitric acid and 18% water). The remaining cellulose was hydrolyzed with 67% sulfuric acid (v/v) and measured by a colorimetric assay with anthrone with Avicel PH-101 (FMC, Rockland, ME, USA) as a cellulose standard. Mean cellulose content of each cotton variety was obtained by measuring the randomly selected cotton bolls from two biological replicates.

Attenuated Total Reflection Fourier Transform Infrared (ATR FT-IR) spectral collection and data analysis

Five cotton bolls were randomly selected from each biological replicate samples of cultivated *G. arboreum* A_2-100 and *G. hirsutum* TM-1 and DP-5690 producing long fibers as well as wild *G. raimondii* D_5-31 and *G. hirsutum* Li_1 and Li_2 mutants generating short fibers. Dried fiber samples were manually harvested from the selected bolls, and divided into six portions that were directly scanned without further processing. Average spectra of each replicate samples were obtained from the spectra from the six portions of the sample. Mean spectra of each cotton variety were obtained by measuring the randomly selected cotton bolls from two biological replicates. All samples were scanned by an FTS 3000MX FT-IR spectrometer (Varian Instruments, Randolph, MA, USA) equipped with a ceramic source, KBr beam splitter, and deuterated triglycine sulfate (DTGS) detector and attenuated total reflection (ATR) attachment according to the methods that were previously described in Liu and Kim [29]. The spectra were normalized by dividing the intensity of an individual band in the 1800–600 cm^{-1} region with the average intensity in that 1800–600 cm^{-1} region, and subsequent principal component analysis (PCA) characterization was performed in the 3000–1200 cm^{-1} IR region with mean centering (MC), multiplicative scatter correction (MSC), and Savitzky–Golay first-derivative (13 points) spectral pretreatment and with leave-one-out cross-validation method.

Pyrolysis-molecular beam mass spectrometry lignin analysis

Five cotton bolls were randomly selected from each replicate samples of the cultivated *G. arboreum* A_2-100 and *G. hirsutum* TM-1 as well as wild *G. raimondii* D_5-31. Two replicated samples of the dried cotton fibers of *G. raimondii* D_5-31, *G. arboreum* A_2-100, and *G. hirsutum* TM-1 were cut in a Wiley mill into 20 mesh. Average contents of each cotton variety were obtained by measuring the randomly selected cotton bolls from two biological replicates. Lignin analysis was performed with pyrolysis molecular beam mass spectroscopy (pr-MBMS) by Complex Carbohydrate Research Center (CCRC) at University of Georgia. Duplicated cotton samples along with control samples including NIST 8492 (lignin content, 26.2%) and aspen standards were pyrolyzed at 500˚C and the volatile compounds were analyzed for lignin using a molecular beam mass spectrometer (Extrel Core Mass Spectrometers). The raw data were processed through UnscramblerX 10.1 software to obtain the principal components and raw lignin data. *G. arboreum* A_2-100 fibers exclusively composed of cellulose (95.6~100%) with the lowest lignin level among cotton species was also used as the lignin base line for all tested cotton samples.

Transcriptomic analyses

RNA-seq reads for the cotton materials shown in Table 1 were retrieved from the NCBI SRA database. These reads were aligned to the JGI *G. raimondii* reference genome [10] using gsnap software and reads that mapped to annotated genes were counted using bedtools software [30, 31]. RNA-seq expression analysis was conducted following the PolyCat pipeline as previously described [24, 32]. Briefly, all reads were aligned to the JGI *G. raimondii* reference genome, then the PolyCat software assigned each categorizable read to either the A_T or D_T subgenome based on an index of homeoSNPs. Using the retrieved RNA-seq reads (Table 1 and S1 Table), RPKM (reads per kilobase of transcript per million reads mapped) numbers were determined, and specifically or differentially expressed genes in developing D_5, Li_1 or Li_2 fibers at elongating PCW or wall-thickening SCW stage were identified and annotated based on the best hit by BLAST search with The *Arabidopsis* Information Resource version 10 (TAIR 10). The GO enrichment analysis was performed using agriGO v2.0 Singular Enrichment Analysis [33].

Table 1. Database of RNA-seq used in this study.

Cotton species (genome)	Fiber stages (DPA)	NCBI SRA RNA-seq Database	Reference
G. raimondii (D_5)	PCW (10)	SRX062246	[34]
	SCW (20)	SRX062252	
G. arboreum (A_2)	PCW (10)	SRX062247	
	SCW (20)	SRX062251	
G. hirsutum TM-1 (AD_1)	PCW (10)	SRX202895	[35]
	PCW (20)	SRX202896	
	SCW (28)	SRX15944380	[36]
G. hirsutum DP-5690 (AD_1)	PCW (8)	SRX853935, SRX853937, SRX 853939	[24]
G. hirsutum Li_1 in DP-5690 background (AD_1)	PCW (8)	SRX853916, SRX853922, SRX853932,	
G. hirsutum Li_2 in DP-5690 background (AD_1)	PCW (8)	SRX853891, SRX853905, SRX 853912	

Image and statistical analyses

Image composites were constructed using Adobe Photoshop 2022 software. Statistical analyses and construction of graphs were performed using t-test and Prism version 7.05 software (Graph-Pad Software, Inc., San Diego, CA). The p value cutoff for significance was 0.05 with four levels at <0.05 (*), <0.01 (**), <0.001 (***), and < 0.0001 (****).

Results

Phenotypic characterizations of cotton fibers

Fiber phenotypic characterizations of wild diploid *G. raimondii* and two cultivated cotton species (1ˢᵗ set). There were notable phenotypic variations of cottonseeds among three different cotton species including a wild diploid *G. raimondii*, cultivated diploid *G. arboreum*, and cultivated allotetraploid *G. hirsutum* (1ˢᵗ set, Fig 1A). The average maximum fiber lengths of diploid *G. raimondii* D_5-6 (11.7 mm) and D_5-31 (10.1 mm) were significantly (p<0.0001) shorter than the two diploid of *G. arboreum* A_2-100 (30.1 mm) and SXY1 (28.0 mm) as well as the two polyploid *G. hirsutum* TM-1 (41.7 mm) and SG-747 (40.1 mm) (Fig 2A). Similarly, the cellulose contents of the two *G. raimondii* fibers (75.0~78.0%) were significantly (p<0.01) less than those of the two *G. arboreum* (95.6~100%) and the two *G. hirsutum* (95.8~98.0%) fibers that were almost exclusively composed of cellulose (Fig 2B).

Fiber phenotypic variations of short fiber *G. hirsutum* mutants and their near isogenic line DP-5690 (2ⁿᵈ set). The fiber lengths and cellulose content of *G. hirsutum* Li_1 and Li_2 mutants were also compared with their NIL cotton, DP-5690 (2ⁿᵈ set, Fig 1B). Average maximum fiber lengths of *G. hirsutum* Li_1 (2.6 mm) and Li_2 (8.1 mm) mutants were significantly (p<0.0001) shorter than their NIL, DP-5690 (38.7 mm) (Fig 3A). Average cellulose content of Li_1 (86.9%) fibers was substantially lower than its NIL, DP-5690 (94.5%). However, the reduction was not statistically significant (p> 0.05) due to high variations among the Li_1 samples (Fig 3B). Average cellulose content of Li_2 fibers (85.1%) was significantly (p<0.041) lower than its NIL, DP-5690 (Fig 3B).
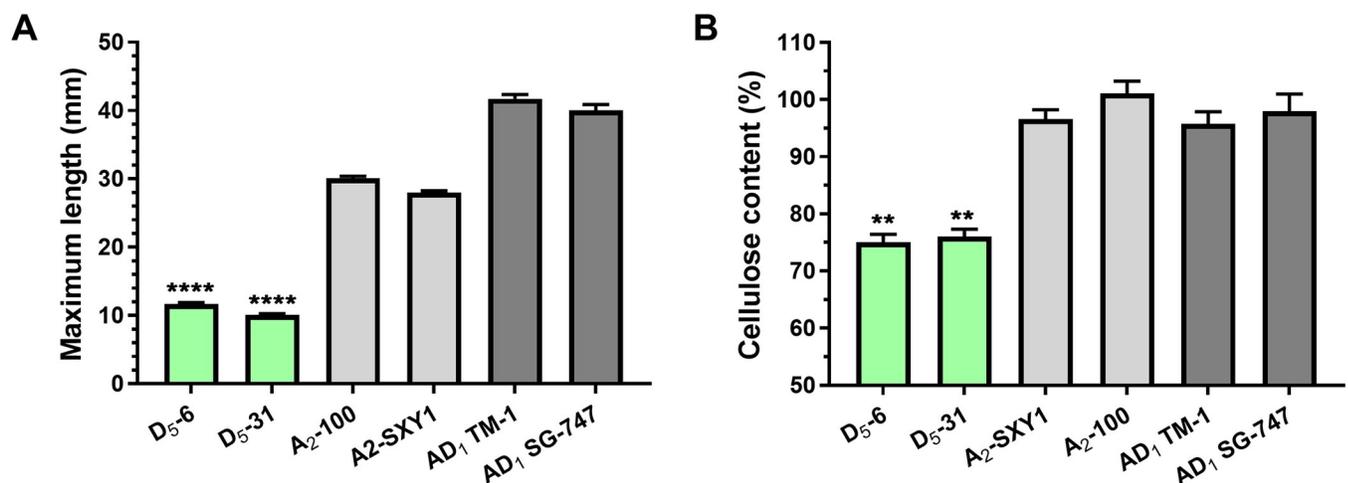
Fig 2. Fiber phenotypic variations between wild diploid *G. raimondii* and two cultivated cotton species (1ˢᵗ set). A. Maximum fiber length. Fiber length was manually measured from cottonseeds of wild *G. raimondii* (D_5-6 and 31), cultivated *G. arboreum* (A_2-100 and SXY1), and cultivated *G. hirsutum* (TM-1 and SG-747). B. Cellulose content. The error bars represent standard deviation. **p value < 0.01; ****p value <0.0001.

https://doi.org/10.1371/journal.pone.0282799.g002

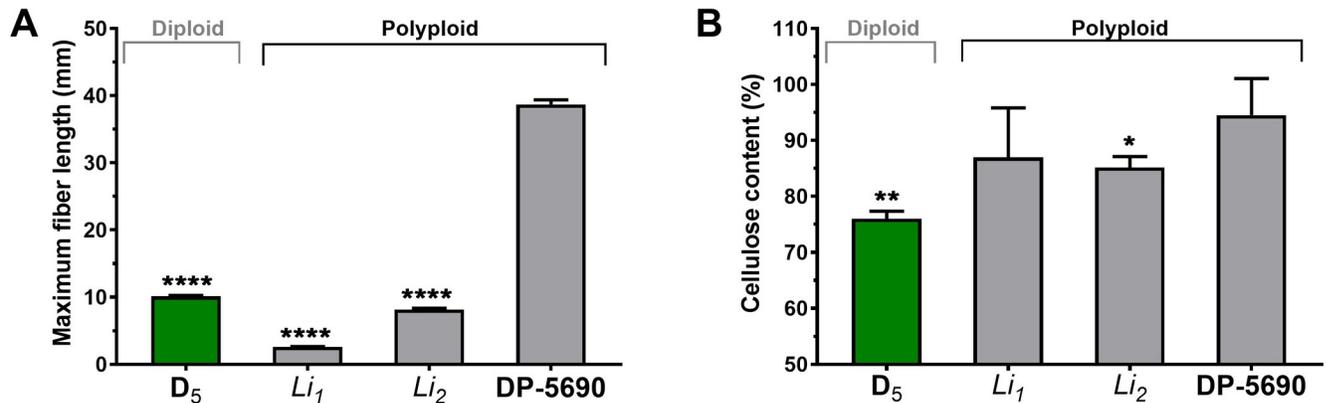**Fig 3. Fiber phenotype variations between short fiber *G. hirsutum* mutants and their near isogenic line DP-5690 (2nd set).** A. Maximum fiber length. Fiber length of *G. hirsutum* short fiber Li_1 and Li_2 mutants were manually measured and compared with those of their NIL, *G. hirsutum* DP-5690 and wild diploid *G. raimondii* D_5-31 (D_5). B. Cellulose content. The error bars represent standard deviation. *p value < 0.05; **p value < 0.01; ****p value < 0.0001.

Chemical analyses of cotton fibers

Suberin and lignin depositions in the wild diploid *G. raimondii* fibers among the 1st set of cotton materials. Among the three cotton species, wild *G. raimondii* produces naturally green colored fibers, whereas cultivated *G. arboreum* and *G. hirsutum* are white cotton (Fig 4A). Chemical fiber properties were monitored with ATR FT-IR spectroscopy in the 600–4000 cm^{-1} region (Fig 4B). Multiple IR spectral peaks known to be indicatives of suberin (1513, 1635, 1738, 2850, and 2920 cm^{-1}) and lignin (1514 and 1705–1720 cm^{-1}) components [37–42] were specifically identified from the wild *G. raimondii* fibers (S2 Table and Fig 4B).

Lignin levels were measured using a molecular beam mass spectrometer (Fig 4C). Among the three lignin units including 4-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units in the lignin polymer, the S lignin in *G. raimondii* fiber (1.8%) was significantly (p = 0.008) greater than cultivated *G. arboreum* (0%) and *G. hirsutum* (0.6%) fibers. In contrast, both G lignin (p = 0.18) and H lignin (p = 0.31) showed insignificant variation among the three cotton species.

Suberin and lignin depositions in the polyploid *G. hirsutum* Li_1 and Li_2 mutant fibers among the 2nd set of cotton materials. The unginned Li_1 and Li_2 fibers as well as their NIL fibers appear to be white (Fig 1B). On the contrary, the ginned Li_1 and Li_2 fibers showed a brown color that was visually different from their NIL fibers showing a white color (Fig 5A). Consistently, the IR spectra of Li_1 and Li_2 fibers were also different from those of DP-5690 fibers (Fig 5B). Multiple IR spectral peaks assigned as suberin (1738, 2850, and 2920 cm^{-1}) and lignin (1705–1720 cm^{-1}) components were specifically detected in the Li_1 and Li_2 short fibers (Fig 5B). In addition, a bigger bulge area (1580~1640 cm^{-1}) of IR spectrum was detected from the short mutant fibers as compared with the NIL fibers. The bulge area may be composed of four close IR peaks (1588, 1606, 1624, and 1635 cm^{-1}) that were assigned as suberin fractions from other plants [37, 41].

Transcriptomic profiles of developing cotton fibers

Classification of developmental stages of the 1st set of cotton fibers retrieved from the original RNA-seq analyses. As summarized in Table 1, RNA-seq data of developing *G. raimondii*, *G. arboreum*, and *G. hirsutum* fibers at 10, 20, or 28 DPAs were available from a public database. Because they are distinct species and grown in different environments [34–36], a
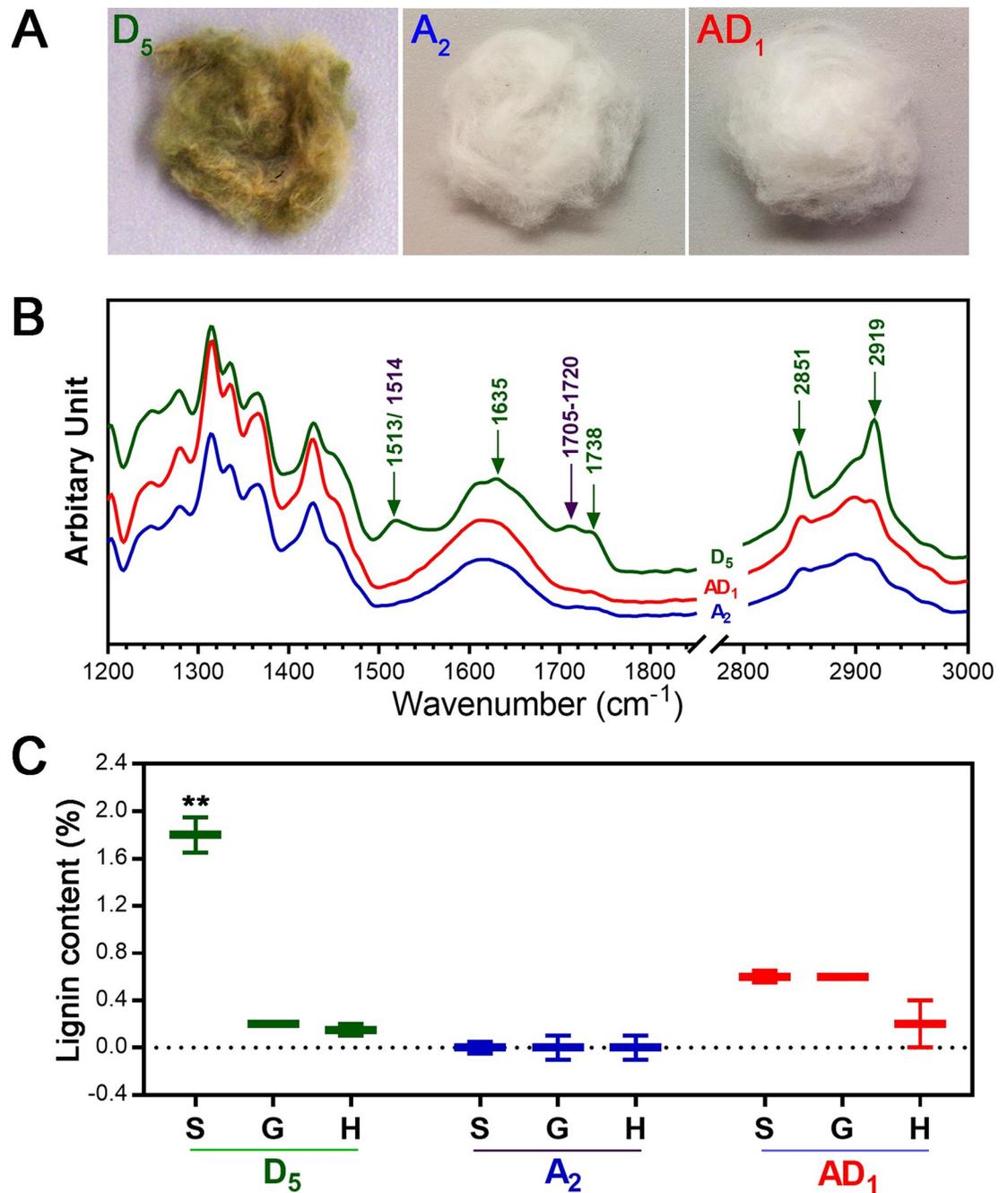
Fig 4. Suberin and lignin deposition in the wild diploid *G. raimondii* fibers among the 1st set of cotton materials. A. Color differences of ginned and fully developed fibers from diploid *G. raimondii* D_5-31 (D_5), diploid *G. arboreum* A_2-100 (A_2), and polyploid *G. hirsutum* TM-1 (AD_1). B. ATR FT-IR spectra. Each spectrum of the three cotton species fibers was normalized and compared. The wavenumbers representing suberin and lignin were labeled green and purple, respectively. C. Lignin quantification. Contents of syringyl (S), guaiacyl (G), and hydroxyphenyl (H) lignin were determined with pyrolysis-molecular beam mass spectrometry and compared among the three cotton species fibers. The error bars represent standard deviation. **p value < 0.01.

https://doi.org/10.1371/journal.pone.0282799.g004

closer examination of the RNA-seq data was needed to better align overall expression profiles and the fiber growth stages for more meaningful comparisons. The fiber developmental stages
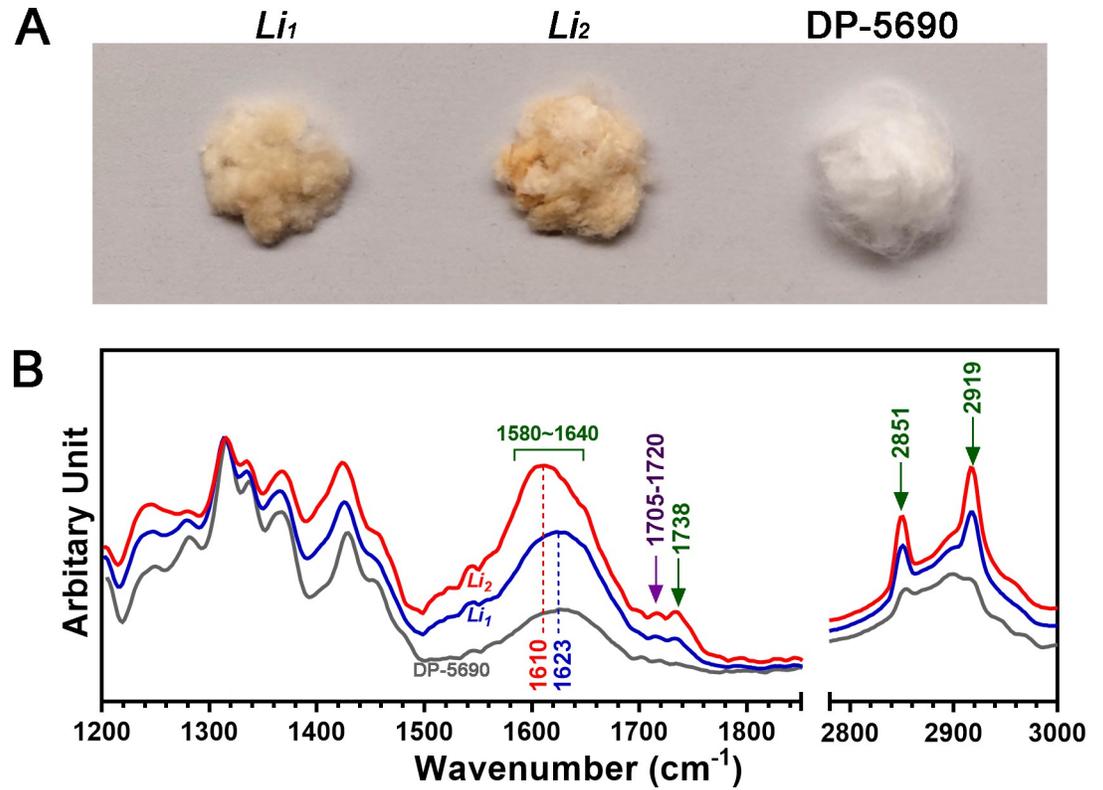
Fig 5. Suberin and lignin deposition in the polyploid *G. hirsutum Li₁* and *Li₂* mutant fibers among the 2nd set of cotton materials. A. Color detection from the developed *Li₁* and *Li₂* mutant fibers. B. Normalized ATR FT-IR spectra. The wavenumbers representing suberin and lignin were described with green and purple fonts, respectively.

of the three cotton species used in the original research [34–36] were first classified by monitoring the transcript abundance and patterns of the indicator genes including *fasciclin-like arabinogalactan* [43] and *expansin* [44] that are specifically up-regulated at PCW stage as well as *cellulose synthase* (*CesA*) genes that are specifically up-regulated at SCW stage [45]. Three *Fasciclin-like arabinogalactan* genes and four *expansin* genes were commonly up-regulated in the 10 DPA fibers of all three cotton species (Table 2), suggesting that developing *G. raimondii*, *G.*

Table 2. Classifications of cotton fiber developmental stages of *G. raimondii*, *G. arboreum*, and *G. hirsutum*. Ten indicator genes are shown for each species according to genome type and at 10, 20, or 28 DPA. The most up-regulated RPKM numbers of each indicator genes in the cotton species were written in bold fonts.

Cotton ID	TAIR10 best hit	TAIR10 Description	*G. arboreum*		*G. raimondii*		*G. hirsutum*					
			A_2-10	A_2-20	D_5-10	D_5-20	A_T-10	A_T-20	A_T-28	D_T-10	D_T-20	D_T-28
Gorai.013G152900	AT5G55730	Fasciclin-like arabinogalactan	**1389.5**	382.3	**162.1**	30.5	**363.8**	324.6	22	25.6	**96.5**	22.2
Gorai.004G026800	AT3G55820	Fasciclin-like arabinogalactan	**358.7**	296.3	**423.9**	144.5	101.5	**134.2**	138.6	197.9	**244.4**	166.4
Gorai.006G147500	AT5G06390	Fasciclin like arabinogalactan	**2095.5**	1241.8	**949.6**	743.1	**10112.5**	7192.8	1952.2	**6126.4**	3972	1003.8
Gorai.001G083300	AT3G03220	expansin A13	**543.6**	153.8	**815.3**	175.1	**2086.8**	1086.1	93.8	**3404**	1772.1	139.7
Gorai.005G142200	AT2G03090	expansin A15	**279.7**	63.3	**942.7**	22.4	**583.7**	285.6	23.9	73.3	**141.5**	32.3
Gorai.003G131000	AT2G39700	expansin A4	1078.2	278.2	**4945.1**	46.8	**5431.3**	4293.2	0	**1912.7**	1579.2	0
Gorai.006G172400	AT2G39700	expansin A4	**367.7**	165.1	**1839.1**	30.5	363.8	478.4	4.8	1491.3	**2518.3**	9.7
Gorai.001G238100	AT5G44030	cellulose synthase A4	0	1054	9.3	6427.3	0	17.1	**21908.6**	0	22.5	**35749.3**
Gorai.004G057400	AT5G44030	cellulose synthase A4	76.7	**21670.8**	162.1	**66777.2**	188.9	202.6	**97586.7**	183.2	231.6	**61131.7**
Gorai.001G044700	AT5G17420	Cellulose synthase A7	20.3	**18033.8**	0	**41768.4**	5.6	0	**4996.5**	7.3	25.7	**84363.2**

arboreum, and *G. hirsutum* fibers at 10 DPA used in the original research were at the PCW stage. Three SCW *CesA* genes were up-regulated in developing *G. arboreum* and *G. raimondii* fibers at 20 DPA. In contrast, they were up-regulated in developing *G. hirsutum* TM-1 fiber at 28 DPA, but not at 20 DPA. The RNA-seq data of *G. raimondii* (20 DPA), *G. arboreum* (20 DPA), and *G. hirsutum* (28 DPA) were compared for studying SCW stage.

Transcriptomic analyses of the 1ˢᵗ set of cotton species at PCW or SCW stage. The expressed gene (EGs) numbers in *G. raimondii* D_5 genome at PCW (18,698 EGs) and SCW (18,638 EGs) stages were similar to those in *G. hirsutum* D_T subgenome at PCW (15,485 EGs), and SCW (17,103 EGs) stages. The EG numbers in the *G. arboreum* A_2 genome at PCW (18,333 EGs) and SCW (17,520 EGs) stages were also similar to those in the *G. hirsutum* A_T subgenome at PCW (16,188 EGs), and SCW (17,472 EGs) stage. In this study, the focus was on identification of genes specifically expressed in developing *G. raimondii* fibers, but not expressed (zero RPKM) in developing fibers of the other cotton species. Transcriptomic profiles between the two diploid cotton species identified specifically expressed genes (SEGs) in developing *G. raimondii* fibers at PCW (2,663 SEGs) and SCW (3,193 SEGs) stage as shown in Fig 6A. When the transcriptomic profiles of the three cotton species compared, 1,385 and 1,520 SEGs were identified from developing *G. raimondii* at PCW and SCW stages, respectively (Fig 6A and S3 Table). Between PCW and SCW stages of developing *G. raimondii* fibers, 297 SEGs were overlappingly expressed (Fig 6B).

Of the 297 SEGs, glycerol-3-phosphate acyltransferase (GPAT, Gorai.008G098000) is required for suberin biosynthesis in *Arabidopsis* (S4 Table) [46]. Peroxidase (Gorai.013G005800) related to lignin biosynthesis [47]. Laccases (Gorai.007G376600 and 004G234200) are involved in phenylpropanoid pathway producing suberin and lignin [47]. Cytochrome P450 family genes (Gorai.001G222700, 002G097500, 003G077000, 006G217000 and 010G166100) shown in S4 Table play important roles in detoxification of xenobiotics and stress responses [48]. Gene ontology enrichment analysis using Singular Enrichment Analysis of agriGO v. 2.0 [33] identified four GO categories including O-methyltransferase activity (GO:0008171), terpene synthesis activity (GO:0010333), tetrapyrrole binding (GO:0046906), and response to endogenous stimulus (GO:0009719) in developing *G. raimondii* fibers at PCW stage (Table 3). Three GO categories (O-methyltransferase activity, terpene synthesis activity, and tetrapyrrole binding) were also found at SCW stage. Multiple O-methyltransferases were reported to be involved in redundant functions for lignin, suberin and flavonoids [49]. Terpene synthesis proteins also control synthesis of secondary metabolites and gossypol biosynthesis by responding to various abiotic stresses [12]. Tetrapyrroles are required for detoxification of reactive oxygen species (ROS), programed cell death, photosynthesis and respiration [50]. Consistent with the previous report [13], GO (GO:0009719) involved in disease resistance was also over-represented in *G. raimondii* fibers (Table 3 and S3 Table).

Transcriptomic analyses of the 2ⁿᵈ set consisting of *G. hirsutum* NILs differing in fiber length. The original RNA-seq analysis of the short fiber mutants (*Li₁* and *Li₂*) was performed with total RNAs extracted from developing fibers at PCW stage (8–12 DPA) grown in greenhouse or cotton fields [22, 24]. To identify differentially expressed genes (DEGs), developing fibers from field grown *Li₁* and *Li₂* fibers were compared to their NIL, *G. hirsutum* DP-5690, using a 2-fold difference as a threshold. In the *Li₁* mutant fibers, 4,043 genes were up-regulated whereas 2,536 genes were down-regulated (Fig 7A). In the *Li₂* mutant fibers, 2,419 genes were up-regulated, whereas 1,740 genes were down-regulated (Fig 7A). Identification of candidate genes producing the color pigments in the two mutant fibers focused on the 1,285 genes (S5 Table) that were commonly up-regulated in the developing *Li₁* and *Li₂* fibers (Fig 7B).

GO enrichment analysis of the 1,285 UGs identified six GO categories (Table 4). The two GO categories including transporter activity (GO:0005215) and cellular respiration
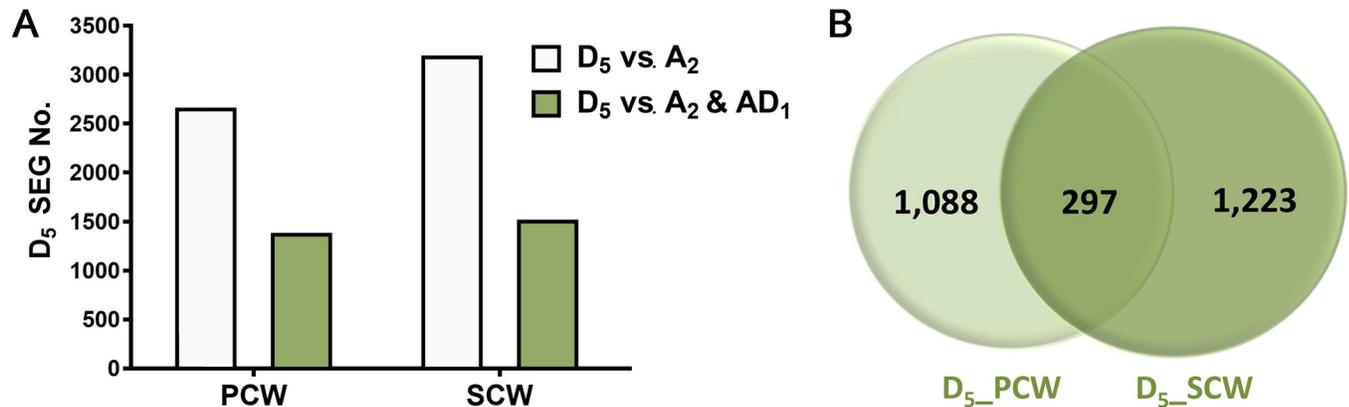
Fig 6. Transcriptomic analyses of wild diploid *G. raimondii* and two cultivated cotton species (1st set). A. Quantification of specifically expressed genes (SEGs) in developing *G. raimondii* fibers at primary or secondary cell wall (PCW or SCW) stages by comparing them with *G. arboreum* (D_5 vs A_2) as well as *G. arboreum* and *G. hirsutum* (D_5 vs A_2 & AD_1). B. Venn diagrams representing the common SEGs in *G. raimondii* fibers between PCW (D_5_PCW) and SCW (D_5_SCW) stages.

(GO:0045333) were also previously identified in the original analysis [24] by MapMan ontology [51]. Among the newly identified four GO categories, ADP binding (GO:0043531) and protein kinase activity (GO:0004672) composed of multiple nucleotide-binding leucine-rich repeat receptors (NLRs) and leucine-rich repeats receptor-like kinase (LRR-RLK) are involved in plant development and stress responses in other plants [52]. The other two GO categories, tetrapyrrole binding (GO:0046906) and response to endogenous stimulus (GO:0009719), were also over-represented in wild diploid *G. raimondii* fibers (Tables 3 and 4).

Integration of the chemical phenotypes and transcriptomic profiles between the two sets of cotton materials differing in fiber lengths and cellulose contents

For an examination of the quantitative and statistical significances of the spectral features showing different chemical components among the fiber samples used in the 1st and 2nd sets, a principal component analysis (PCA) was performed with the spectral region (1200–3000 cm^{-1}) composed of IR peak bands of suberin, lignin, and cellulose (Fig 8A). The analysis showed a dominant first principal component (PC1) accounting for 75.9% of the total variation, and revealed a distinction in PC1 score within the six tested samples (Fig 8A). The PC1 score increased in the order of *G. hirsutum* Li_2 < *G. hirsutum* Li_1 < *G. raimondii* D_5-31 < *G. hirsutum* TM-1 ≈ *G. hirsutum* DP-5690 ≈ *G. arboreum* A_2-100. The three cultivated cottons of *G. arboreum* A_2-100 (0.369), *G. hirsutum* DP-5690 (0.365), and *G. hirsutum* TM-1 (0.326) demonstrated similar positive PC1 scores with insignificant (p = 0.587) variations. In contrast, the other three short fiber cottons of *G. hirsutum* Li_2 (-0.738), *G. hirsutum* Li_1 (-0.279), and *G.*

Table 3. GO enrichment analyses of specifically expressed genes in PCW or SCW stage of developing *G. raimondii* (D_5) fibers. List and annotation of genes are described in S3 Table.

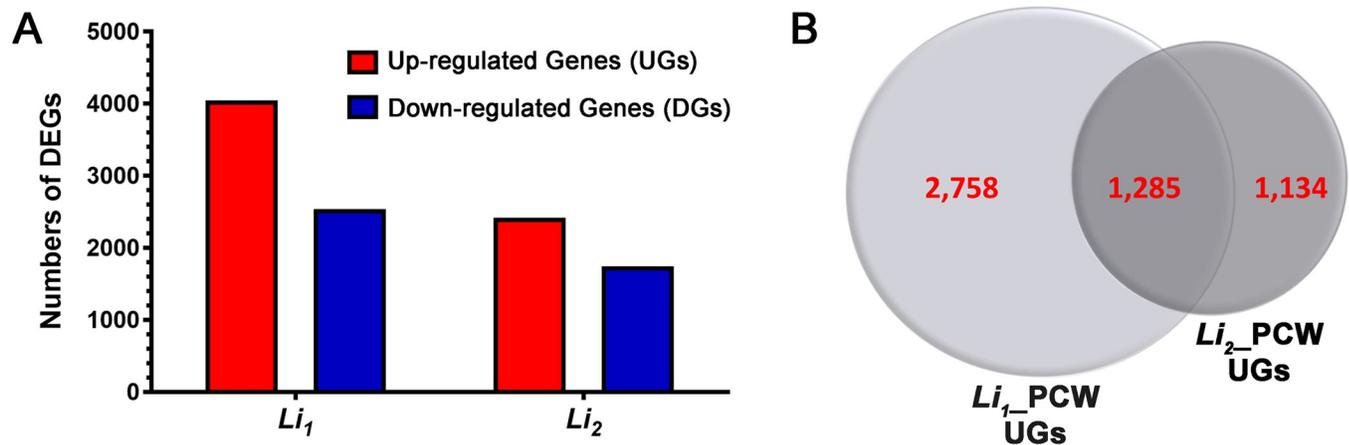| GO term	Common mechanism	D_5_PCW	D_5_SCW
GO:0008171	O-methyltransferase activity	12 genes (p<0.0001)	10 genes (p<0.0001)
GO:0010333	Terpene synthesis activity	8 genes (p<0.0001)	16 genes (p<0.0001)
GO:0046906	Tetrapyrrole binding	27 genes (p<0.0001)	44 genes (p<0.0001)
GO:0009719	Response to endogenous stimulus	10 genes (p<0.01)	

Fig 7. Transcriptomic analyses among three *G. hirsutum* NIL fibers differing in length. A. Summary of up- or down-regulated DEGs of *G. hirsutum* short fiber *Li₁* and *Li₂* mutant fibers at PCW stage with the corresponding NIL DP-5690 fiber. The DEGs annotated with JGI *Gossypium raimondii* reference genome were listed in S5 Table. B. Venn diagrams of the up-regulated genes (UGs) in developing *G. hirsutum* Li_1 and Li_2 mutant fibers.

https://doi.org/10.1371/journal.pone.0282799.g007

raimondii D_5-31 (-0.043) had all negative scores that were significantly (p<0.0001) variable from the cultivated cotton species. Among the three short fiber cottons, the PC1 scores also showed significant (p<0.0001) variations. These results suggested that the chemical compositions of *G. raimondii* D_5-31, *G. hirsutum* Li_1, and *G. hirsutum* Li_2 were different despite a sharing of suberin and lignin fractions in addition to cellulose.

Of the UGs in *G. hirsutum* Li_1 (4,043) and Li_2 (2,419) fibers (Fig 7), there were redundant and homeologous genes in polyploid *G. hirsutum* Li_1 (880) and Li_2 (376) mutants. To identify the commonly up-regulated orthologous genes between the diploid D_5 genome and polyploid AD_1 genome composed of A_T and D_T subgenomes, we compared the diploid *G. raimondii* 1,385 SEGs with non-redundant UGs in polyploid *G. hirsutum* Li_1 (3,163) and Li_2 (2,043) (Fig 8B). Among the three short cottons, 29 orthologs were commonly up-regulated (Table 5). Consistent with the co-existence of lignin and suberin components in the three short cottons, a laccase (Gorai.009G260600_D_T) and a peroxidase (Gorai.013G005800_A_T) involved in lignin polymerization by oxidizing lignin monomers (monolignols) [47, 53, 54] as well as ABC-2 type transporter family gene (Gorai.002G062500_D_T) required for suberin biosynthesis in *Arabidopsis* [55] were found. Jasmonate-zim-domain protein 8 (JAZ8, Gorai.009G154300) involved in flavonoid synthesis [56] was also found. A protein kinase protein (Gorai.011G182700) responsible for tetrapyrrole metabolism in a *Arabidopsis* color mutant was also up-regulated [57]. The other annotated genes including Fe(II)/ascorbate oxidase (SRG1) [58], cysteine-rich receptor-like protein kinase 10 (CRK10) [59], leucine-rich repeat protein kinase [60], FAD-binding berberine family protein [61], glutathione S-transferase [62], PLAT/

Table 4. GO enrichment analyses of up-regulated genes in *G. hirsutum* Li_1 and Li_2 short fibers at PCW stage. List and annotation of genes are described in S5 Table.

GO term	Description	Number of genes	p-value
GO:0005215	transporter activity	50	0.00065
GO:0045333	cellular respiration	7	0.00095
GO:0043531	ADP binding	28	9.5e-14
GO:0004672	protein kinase activity	75	1.2e-06
GO:0046906	tetrapyrrole binding	30	3.4e-07
GO:0009719	response to endogenous stimulus	9	7.6e-05

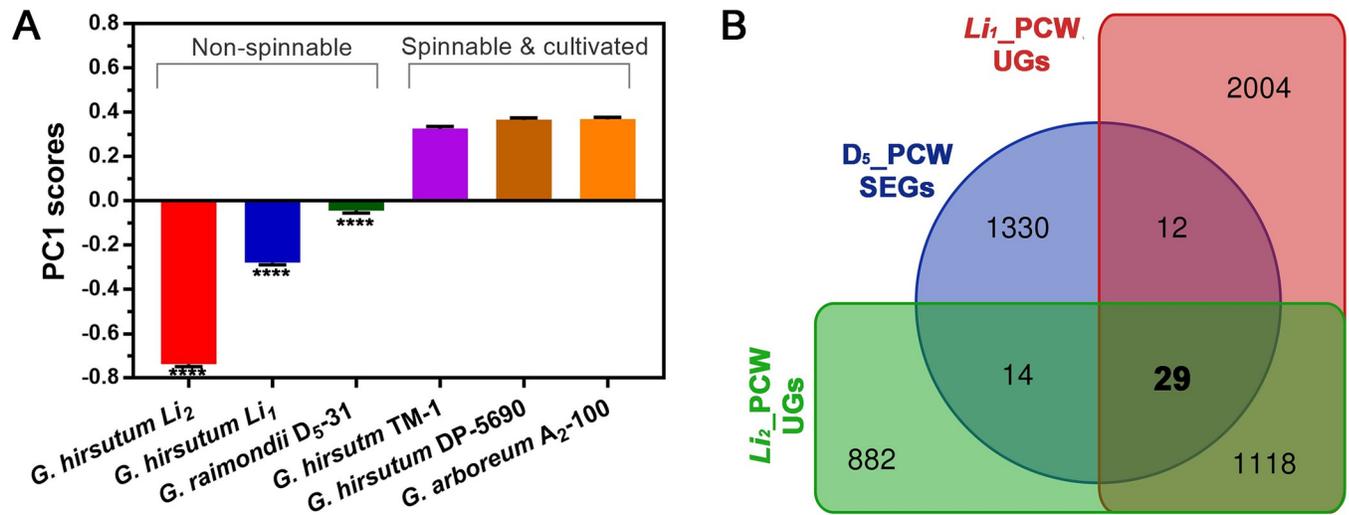https://doi.org/10.1371/journal.pone.0282799.t004

Fig 8. Integrations of chemical phenotypes and transcriptomic profiles between the two sets of cotton materials differing in fiber lengths and cellulose content. A. Classification of the six cotton samples from the first principal component (PC1) scores of principal component analysis operation on normalized ATR FT-IR spectra. B. Venn diagrams comparing the *G. raimondii* specifically expressed genes (SEGs) at PCW stage with the *G. hirsutum Li₁* and *Li₂* up-regulated genes (UGs) at PCW stage. The error bars represent standard deviation. ****p value <0.0001.

LH2 domain-containing lipoxygenase [63], FLOTILIN2 (FLOT2) [64], cytokinin response activator (ARR1) [65], PSBP-domain protein 6 (PPD6) [61], cyclic nucleotide gated channel 1 (CNGC) [66], and NAC014 [67] were all related to immunity or stress responses. Seven of them have not been annotated and the 22 annotated genes were not sufficient to perform the GO enrichment analysis.

Discussion

Common characteristics of physical properties of cotton fibers from wild diploid *G. raimondii* and polyploid *G. hirsutum Li₁* and *Li₂* mutants

Physical properties of cultivated cotton fibers are generally assessed by a High Volume Instrument (HVI) which is defined by the International Cotton Advisory Committee as a standardized instrument for cotton fiber quality measurements [7]. Cotton fibers with a length less than 12.7 mm are classified as short fibers that reduce the quality of spun yarns. The cotton fibers produced by *G. raimondii* and *G. hirsutum Li₁* and *Li₂* mutants were too short to be measured by HVI. Thus, we manually measured the maximum fiber lengths of the wet and relaxed cotton fibers from the chalezel end of cottonseeds [27]. The maximum fiber length of *G. raimondii* D_5-6 (11.7 mm) and D_5-31 (10.1 mm) as well as *G. hirsutum Li₁* (2.6 mm) and *Li₂* (8.1 mm) mutants were significantly shorter than those of cultivated diploid cotton, *G. arboreum* SXY1 (28.0 mm) and A_2-100 (30.1 mm) as well as cultivated polyploid cotton, *G. hirsutum* TM-1 (41.7 mm), SG-747 (40.1 mm) and DP-5690 (38.7 mm) as shown in Figs 2A and 3A. There are two different types of *G. hirsutum* fibers. Lint fibers differentiate from ovule epidermis on the day of anthesis and they grow approximately 25~35 mm based on HVI measurements. In contrast, linter or fuzz differentiate from the ovule epidermis around 5 to 10 DPA and they do not grow longer than 15 mm [68]. Wild diploid *G. raimondii* was often described as a lintless, non-fibered, or fiberless species [12, 69, 70]. The full names of the *Li₁* and *Li₂* mutants also contain "lintless" [17, 18]. However, the fiber initials of *G. raimondii* [8], *G. hirsutum Li₁* mutant [19], and *G. hirsutum Li₂* mutant [20] all differentiate on the day of anthesis.

Table 5. Commonly up-regulated orthologous genes among developing *G. raimondii* D_5, *G. hirsutum* Li_1 and Li_2 fibers at PCW stage.

No	*G. raimondii*		*G. hirsutum*			Description
	D5 gene ID	RPKM*	AD1 gene ID	Li_1 /DP-5690§	Li_2 /DP-5690§	
1	Gorai.010G058200	145.9	Gorai.010G058200_A_T	3.0	19.7	Fe(II)/ascorbate oxidase gene SRG1
2	Gorai.013G005800	132.0	Gorai.013G005800_A_T	17.2	16.0	Peroxidase superfamily protein
3	Gorai.002G062500	99.6	Gorai.002G062500_D_T	3.7	7.0	ABC-2 type transporter family protein
4	Gorai.001G266300	39.4	Gorai.001G266300_D_T	2.1	2.1	Lysine ketoglutarate reductase trans-splicing protein
5	Gorai.011G017700	37.1	Gorai.011G017700_A_T	2.3	2.1	CRK10 (cysteine-rich receptor-like protein kinase 10)
6	Gorai.007G338100	27.8	Gorai.007G338100_A_T	11.3	2.8	Disease resistance LRR family protein
7	Gorai.009G389300	25.5	Gorai.009G389300_D_T	13.9	6.1	Protein of unknown function (DUF2921)
8	Gorai.005G023800	20.9	Gorai.005G023800_A_T	9.2	59.7	FAD-binding Berberine family protein
			Gorai.005G023800_D_T	6.1	45.3	
9	Gorai.009G154300	13.9	Gorai.009G154300_A_T	4.6	4.6	Jasmonate-zim-domain protein 8 (JAZ8)
10	Gorai.005G013300	13.9	Gorai.005G013300_A_T	14.9	3.7	NAC 014
11	Gorai.005G163000	13.9	Gorai.005G163000_A_T	4.0	3.3	Lysine ketoglutarate reductase trans-splicing protein
12	Gorai.011G006900	7.0	Gorai.011G006900_D_T	4.3	3.5	Leucine-rich repeat protein kinase family protein
13	Gorai.013G097700	7.0	Gorai.013G097700_D_T	9.2	3.0	O-fucosyltransferase family protein
14	Gorai.011G163600	4.6	Gorai.011G163600_D_T	13.0	8.6	Glutathione S-transferase tau 7
15	Gorai.001G228100	4.6	Gorai.001G228100_A_T	2.1	5.7	PLAT/LH2 domain-containing lipoxygenase
16	Gorai.006G135600	4.6	Gorai.006G135600_D_T	39.4	4.0	FLOTILIN2 (FLOT2)
17	Gorai.009G260600	2.3	Gorai.009G260600_D_T	11.3	4.0	Laccase 14
18	Gorai.007G074700	2.3	Gorai.007G074700_A_T	4.0	8.0	TPX2 (targeting protein for Xklp2) protein
19	Gorai.009G006200	2.3	Gorai.009G006200_A_T	3.7	8.6	cytokinin response activator (ARR1)
20	Gorai.013G125600	2.3	Gorai.013G125600_A_T	4.0	1.5	PSBP-DOMAIN PROTEIN 6 (PPD6) thylakoid lumenal protein
			Gorai.013G125600_D_T	7.0	2.1	
21	Gorai.011G182700	2.3	Gorai.011G182700_A_T	2.6	14.9	Protein kinase superfamily protein
22	Gorai.001G155300	2.3	Gorai.005G223900_A_T	5.7	8.6	Cyclic nucleotide gated channel 1

*The RPKM values were detected in developing *G. raimondii* fibers, but not detected in developing *G. arboreum* and *G. hirsutum* fibers.

§ Fold difference

https://doi.org/10.1371/journal.pone.0282799.t005

Thus, the short fibers produced from *G. raimondii* and *G. hirsutum* Li_1 and Li_2 mutants can be classified as lint fibers according to the definition described by Lang [68].

In this study, we showed that the three short cottons of wild *G. raimondii* and two *G. hirsutum* mutants produced fibers containing color pigments composed of lignin and suberin (Figs 4A and 5A). The green coloration of *G. raimondii* fibers was reported by Hutchinson and his colleagues in 1947 [69], but its color pigment was not further characterized. A recent study showed that NIR spectra of *G. raimondii* fibers were similar to those measured from naturally green colored *G. hirsutum* fibers [71]. Despite the extensive studies of Li_1 and Li_2 mutant fibers, the brown color pigments of the short fiber mutants (Fig 5A) have been unnoticed. Green color of the cotton fibers is faded to tan color when they are exposed to light [72]. The light brown color of the short mutant fibers might have been overlooked because their NIL, DP-5690, produces white lint fibers.

Common chemical components among diploid *G. raimondii* and polyploid *G. hirsutum* mutants

Chemical analyses using cellulose assay, ATR FT-IR spectroscopy, and mass spectrometry consistently showed suberin and lignin components in the three short fibers. Average cellulose

contents of *G. raimondii* (75.0~78.0%) and *G. hirsutum Li$_1$* and *Li$_2$* mutants (85.1~86.9%) fibers were lower than cultivated *G. arboreum* (95.6~100%) and *G. hirsutum* (95.8~98.0%) fibers (Figs 2B and 3B). Non-cellulosic components of the *G. raimondii* (22.0~25.0%) and *G. hirsutum Li$_1$* and *Li$_2$* (13.1~14.9%) fibers were substantially greater than the cultivated fibers (0~4.4%). These results were consistent with the previous reports showing different cellulose contents between green and white upland cotton [71] and functional divergence of cellulose synthase orthologs between wild *G. raimondii* and cultivated *G. arboreum* [45].

The three short cottons of *G. raimondii* and *G. hirsutum Li$_1$* and *Li$_2$* mutants demonstrated the signature IR spectral peaks of suberin and lignin (Figs 4B and 5B). Another type of short fiber mutant *li$_y$* also showed the signature IR spectral peaks of suberin [73]. In naturally green colored *G. hirsutum*, suberin layers were observed in the secondary cell wall in cotton fibers [72, 74], and a major lignin precursor and its derivatives were deposited in the suberin layers [75]. Suberin and lignin can be produced with common precursors, i.e. phenolic components [76]. In contrast to suberin consisting of phenolics and aromatic polymers, lignin is purely composed of poly-aromatic components [76]. Generally, lignin is derived from three phenyl-propanoid monomers, the monolignols 4-coumaryl, coniferyl, and sinapyl alcohols, that produce the 4-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units in the polymer [77]. Our mass spectrometry lignin analysis showed significantly greater content of the S lignin in wild *G. raimondii* fiber (1.8%) than those of the other cultivated cotton fibers (0~ 0.6%). Recent studies suggest that lignin may play an important role in cotton fiber quality [78, 79].

The integrative PCA method of the two sets enabled classifying the six cotton fibers into two classes according to the PC1 scores (Fig 8A). All three cultivated cottons shared similar positive PC1 scores without any significant variation, whereas all three short cottons showed negative PC1 scores with significant and broad variation. During cotton fiber development, underdeveloped cotton fibers containing high levels of non-cellulosic components show negative PC1 scores [80]. As cellulose content increases during normal fiber development, PC1 scores also increase and become positive [81]. Notably, the pattern of the PC1 score in Fig 8A was consistent with previous reports that PC1 scores increased with the cellulose content during cotton fiber development [45]. The IR bulge area likely represents a macromolecule complex composed of suberized components whose IR signals can overlap. Interestingly, there were noticeable IR peak bands of the bulge area among *G. hirsutum Li$_1$* (1623 cm^{-1}), *G. hirsutum Li$_2$* (1610 cm^{-1}), and *G. raimondii* D5 (1635 cm^{-1}) (Figs 4B and 5B). These results along with the significant variation of their PC1 scores showed variation in the three short cottons although they shared suberin and lignin components (Figs 4B, 5B and 8A).

Commonly up-regulated orthologs among diploid *G. raimondii* and polyploid *G. hirsutum* mutants

To test if suberin and lignin genes were specifically up-regulated in developing *G. raimondii* fibers, we used the RNA-seq data performed with the RNAs extracted from developing fibers of *G. raimondii*, *G. arboreum*, and *G. hirsutum* (Table 1) [34–36]. The original transcriptomic analyses of the short fiber mutants (*Li$_1$* and *Li$_2$*) and their NIL DP-5690 were only performed with total RNAs extracted from developing fibers at PCW stage [22, 24]. Thus, we verified that suberin and lignin were specifically detected at the PCW stage of developing mutant fibers (S1 Fig). The short fiber phenotypes of *G. raimondii* fibers [70, 71] and mutants [19, 20, 22, 24] were consistent across various growing conditions. In this study, we used the JGI *G. raimondii* D$_5$ reference genome for analyzing transcriptomic profiles of the two sets because the *G. raimondii* genome sequence shows high homology (>96%) with the coding sequences of *G. hirsutum* A$_T$ and D$_T$ subgenomes. Thus, the D$_5$ reference genome sequence has been successfully

used for characterizing the Li_1 and Li_2 genomes by several groups [19, 20, 22, 24, 82]. Transcriptomic analysis of the 1st set identified that genes involved in suberin and lignin biosynthesis were specifically expressed in *G. raimondii* fibers (Fig 6, Table 3 and S3 Table). Among them, glycerol-3-phosphate acyltransferase 1 (GPAT), cytochrome P450 family genes, and laccases were reported to be involved in suberin or lignin biosynthesis in other plants [46, 47, 53, 54, 76]. Among the four GO categories over-represented in *G. raimondii* fibers (Table 3), O-methyltransferase activity is essential for biosynthesis of lignin, suberin and flavonoids [49, 83]. Integrative analyses of the two sets identified 29 genes that were commonly up-regulated in wild cotton species and short fiber mutant fibers (Table 5). Of the 22 annotated genes, four genes such as laccase, peroxidase, ABC-2 type transporter, and JAZ8 are involved in biosynthetic processes of lignin, suberin, and their derivatives [47, 54–56, 84]. The other 18 annotated genes are reported to be related to stress responses (Table 5). The mutations of an *actin* [25] and a putative *Ran Binding Protein 1* [23] cause the short fiber phenotypes of the Li_1 and Li_2 mutant respectively, and also up-regulate the genes involved in stress responses including lignin, suberin and flavonoid biosynthesis (Table 5).

Lignin deposition was suggested to reduce the extensibility of expanding fiber cell walls [79]. Suberin has been reported to be a major regulator of water and solute transport, and a pathogen barrier in plant cell walls [76]. A recent functional study of *Arabidopsis* mutants altered in suberin deposition clearly showed the reductions of the apoplastic transport of water and ions [85]. Hydrophobic suberin in the cotton fiber cell walls also negatively affect apoplastic transport activities in cotton fibers [72, 74, 86].

Conclusion

Here, we used both phenotypic and transcriptomic analyses for identifying common mechanisms reducing fiber elongation in the short fibers generated from *G. hirsutum* Li_1 and Li_2 mutants as well as wild *G. raimondii*. Chemical analyses identified a common deposition of suberin and lignin in the short fiber cell walls. The genes involved in suberin and lignin biosynthesis were also commonly up-regulated in the elongating cotton fibers of the three short cottons as compared with the cultivated and long *G. arboreum* and *G. hirsutum* fibers. These results support a notion that suberin and lignin deposition may affect cotton fiber elongation process negatively. They also provide insight on how suberin and lignin biosynthesis can affect fiber length and cellulose productions in wild and cultivated cotton species.

Supporting information

S1 Fig. Suberin and lignin deposition in developing *G. hirsutum* Li_2 mutant fiber at various developmental stages.
(DOCX)

S1 Table. Detailed information of the selected RNA-seq experiments in public database.
(XLSX)

S2 Table. Characteristic ATR-FTIR spectral peaks of suberins and lignins in cotton fibers.
(DOCX)

S3 Table. Annotation of specifically expressed genes in developing *G. raimondii* fibers at primary and secondary wall biosynthesis stages as compared with developing *G. arboreum* and *G. hirsutum* fibers.
(XLSX)

S4 Table. List of genes potentially involved in suberin and lignin biosynthesis in *G. raimondii* fibers.
(XLSX)

S5 Table. Annotation of differentially expressed genes in developing *G. hirsutum Li₁* and *Li₂* mutant fibers at primary wall biosynthesis stage as compared with their near isogenic line, *G. hirsutum* DP-5690 fibers.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Hee Jin Kim.

Data curation: Hee Jin Kim.

Formal analysis: Hee Jin Kim, Yongliang Liu, Gregory N. Thyssen.

Funding acquisition: Hee Jin Kim.

Investigation: Hee Jin Kim, Yongliang Liu, Gregory N. Thyssen, Marina Naoumkina, James Frelichowski.

Methodology: Hee Jin Kim, Yongliang Liu, Gregory N. Thyssen, Marina Naoumkina, James Frelichowski.

Project administration: Hee Jin Kim.

Supervision: Hee Jin Kim.

Validation: Hee Jin Kim, Gregory N. Thyssen.

Visualization: Hee Jin Kim.

Writing – original draft: Hee Jin Kim, Yongliang Liu.

Writing – review & editing: Hee Jin Kim, Yongliang Liu, Gregory N. Thyssen, Marina Naoumkina, James Frelichowski.

References

1. Wakelyn PJ, Bertoniere NR, French AD, Thibodeaux DP, Triplett BA, Rousselle M-A, et al. Cotton fiber chemistry and technology. M. L, editor. New York, NY: CRC Press; 2010.

2. Kim HJ, Triplett BA. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant Physiology. 2001; 127(4):1361–6. PMID: 11743074

3. Haigler CH, Betancur L, Stiff MR, Tuttle JR. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Frontiers in Plant Science. 2012; 3:1–7.

4. Weis KG, Jacobsen KR, Jernstedt JA. Cytochemistry of developing cotton fibers:: A hypothesized relationship between motes and non-dyeing fibers. Field crops research. 1999; 62(2):107–17.

5. Macmillan C, Birke H, Bedon F, Pettolino F. Lignin Deposition in Cotton Cells–Where is the lignin. J Plant Biochem Physiol. 2013; 1:e106.

6. Fan L, Shi WJ, Hu WR, Hao XY, Wang DM, Yuan H, et al. Molecular and Biochemical Evidence for Phenylpropanoid Synthesis and Presence of Wall-linked Phenolics in Cotton Fibers. Journal of Integrative Plant Biology. 2009; 51(7):626–37. https://doi.org/10.1111/j.1744-7909.2009.00840.x PMID: 19566641

7. Frydrych I, Thibodeaux DP. Fiber quality evaluation-current and future trends/ intrinsic value of fiber quality in cotton. In: Wakelyn PJ, Chaudhry MR, editors. Cotton: technology for the 21st century. Washington DC: International Cotton Advisory Committee; 2010. p. 251–96.

8. Applequist WL, Cronn R, Wendel JF. Comparative development of fiber in wild and cultivated cotton. Evolution & development. 2001; 3(1):3–17. https://doi.org/10.1046/j.1525-142x.2001.00079.x PMID: 11256432

9. Wendel JF, Flagel LE, Adams KL. Jeans, genes, and genomes: cotton as a model for studying polyploidy. Polyploidy and Genome Evolution: Springer; 2012. p. 181–207.

10. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, et al. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature. 2012; 492(7429):423–7. https://doi.org/10.1038/nature11798 PMID: 23257886

11. Udall JA, Long E, Hanson C, Yuan D, Ramaraj T, Conover JL, et al. De novo genome sequence assemblies of Gossypium raimondii and Gossypium turneri. G3: Genes, Genomes, Genetics. 2019; 9 (10):3079–85.

12. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, et al. The draft genome of a diploid cotton Gossypium raimondii. Nature Genetics. 2012; 44(10):1098–103. https://doi.org/10.1038/ng.2371 PMID: 22922876

13. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, et al. Genome sequence of the cultivated cotton Gossypium arboreum. Nature Genetics. 2014; 46(6):567–72. https://doi.org/10.1038/ng.2987 PMID: 24836287

14. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology. 2015; 33:531–7. https://doi.org/10.1038/nbt.3207 PMID: 25893781

15. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nature Biotechnology. 2015; 33 (5):524–30.

16. Yu J, Jung S, Cheng C-H, Lee T, Zheng P, Buble K, et al. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. Plants. 2021; 10(12):2805. https://doi.org/10. 3390/plants10122805 PMID: 34961276

17. Griffee F, Ligon LL. Occurrence of" lintless" cotton plants and the inheritance of the character" lintless". Journal of the American Society of Agronomy. 1929; 21(7):711–7.

18. Narbuth E, Kohel R. Inheritance and linkage analysis of a new fiber mutant in cotton. Journal of Heredity. 1990; 81(2):131–3.

19. Gilbert MK, Turley RB, Kim HJ, Li P, Thyssen G, Tang Y, et al. Transcript profiling by microarray and marker analysis of the short cotton (Gossypium hirsutum L.) fiber mutant Ligon lintless-1 (Li 1). BMC Genomics. 2013; 14(1):403. https://doi.org/10.1186/1471-2164-14-403 PMID: 23767687

20. Hinchliffe DJ, Turley RB, Naoumkina M, Kim HJ, Tang Y, Yeater KM, et al. A combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (Gossypium hirsutum L.). BMC Genomics. 2011; 12(1):445.

21. Fang DD, Naoumkina M, Kim HJ. Unraveling Cotton Fiber Development Using Fiber Mutants in the Post-Genomic Era. Crop Science. 2018; 58(6):2214–28.

22. Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD. Comparative Transcriptome Analysis of Short Fiber Mutants Ligon-Lintless 1 And 2 Reveals Common Mechanisms Pertinent to Fiber Elongation in Cotton (Gossypium hirsutum L.). PloS one. 2014; 9(4):e95554. https://doi.org/10.1371/journal.pone. 0095554 PMID: 24748059

23. Naoumkina M, Thyssen G, Fang DD, Florane CB, Li P. A deletion/duplication in the Li2 locus generates siRNAs from a putative RanBP1 that inhibits fiber cell elongation. Plant Physiology. 2022;in press.

24. Naoumkina M, Thyssen GN, Fang DD. RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and–2 (Li 2) revealed important role of aquaporins in cotton (Gossypium hirsutum L.) fiber elongation. BMC plant biology. 2015; 15(1):1–14. https://doi.org/10.1186/s12870-015-0454-0 PMID: 25848981

25. Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, Mattison CP, et al. A Gly65Val substitution in an actin, GhACT_LI1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. The Plant Journal. 2017; 90(1):111–21. https://doi.org/10.1111/tpj.13477 PMID: 28078746

26. Kothari N, Hague SS, Frelichowski J, Nichols RL, Jones DC. Utilization of cotton gemplasm in the winter nursery at Tecoman, Mexico for plant breeding training and research. Journal of Cotton Science. 2011; 15:271–3.

27. Schubert A, Benedict C, Berlin J, Kohel R. Cotton fiber development-kinetics of cell elongation and secondary wall thickening. Crop Science. 1973; 13(6):704–9.

28. Updegraff DM. Semimicro determination of cellulose in biological materials. Analytical Biochemistry. 1969; 32(3):420–4. https://doi.org/10.1016/s0003-2697(69)80009-6 PMID: 5361396.

29. Liu Y, Kim H-J, Delhom CD, Thibodeaux DP. Investigation of fiber maturity measurement by cross-sectional image analysis and Fourier transform infrared spectroscopy on developing and developed upland cottons. Cellulose. 2019; 26:5865–75.

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

31. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010; 26(7):873–81. https://doi.org/10.1093/bioinformatics/btq057 PMID: 20147302

32. Page JT, Gingle AR, Udall JA. PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3: Genes| Genomes| Genetics. 2013; 3(3):517–25. https://doi.org/10.1534/g3.112.005298 PMID: 23450226

33. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic acids research. 2017; 45(W1):W122–W9. https://doi.org/10.1093/nar/gkx382 PMID: 28472432

34. Flagel LE, Wendel JF, Udall JA. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC genomics. 2012; 13(1):302. https://doi.org/10.1186/1471-2164-13-302 PMID: 22768919

35. Yoo M-J, Wendel JF. Comparative Evolutionary and Developmental Dynamics of the Cotton (Gossypium hirsutum) Fiber Transcriptome. PLoS Genetics. 2014; 10(1):e1004073. https://doi.org/10.1371/journal.pgen.1004073 PMID: 24391525

36. Thyssen GN, Fang DD, Zeng L, Song X, Delhom CD, Condon TL, et al. The Immature Fiber Mutant Phenotype of Cotton (Gossypium hirsutum) Is Linked to a 22-bp Frame-Shift Deletion in a Mitochondria Targeted Pentatricopeptide Repeat Gene. G3: Genes| Genomes| Genetics. 2016; 6(6):1627–33. https://doi.org/10.1534/g3.116.027649 PMID: 27172184

37. Cordeiro N, Belgacem M, Silvestre A, Neto CP, Gandini A. Cork suberin as a new source of chemicals.: 1. Isolation and chemical characterization of its composition. International Journal of Biological Macromolecules. 1998; 22(2):71–80. https://doi.org/10.1016/s0141-8130(97)00090-1 PMID: 9585884

38. Boeriu CG, Bravo D, Gosselink RJ, van Dam JE. Characterisation of structure-dependent functional properties of lignin with infrared spectroscopy. Industrial crops and products. 2004; 20(2):205–18.

39. Socrates G. Infrared and Raman characteristic group frequencies: tables and charts: John Wiley & Sons; 2004.

40. Boza Barducci T, Madoo RM. Investigaciones acerca del parentesco de la especie peruana de algodonero Gossypium raimondii, Ulbrich. Estacion Experimental Agricola de la Molina, Lima, Peru. Boletin. 1941;(22).

41. Järvinen R. Cuticular and suberin polymers of edible plants. Analysis by gas chromatographic-mass spectrometric and solid state spectroscopic methods [Dissertation]: University of Turku; 2010.

42. Stewart D. Fourier transform infrared microspectroscopy of plant tissues. Applied Spectroscopy. 1996; 50(3):357–65.

43. Huang GQ, Gong SY, Xu WL, Li W, Li P, Zhang CJ, et al. A Fasciclin-Like Arabinogalactan Protein, GhFLA1, Is Involved in Fiber Initiation and Elongation of Cotton. Plant Physiology. 2013; 161(3):1278–90. https://doi.org/10.1104/pp.112.203760 PMID: 23349362

44. An C, Saha S, Jenkins JN, Scheffler BE, Wilkins TA, Stelly DM. Transcriptome profiling, sequence characterization, and SNP-based chromosomal assignment of the EXPANSIN genes in cotton. Molecular Genetics and Genomics. 2007; 278(5):539–53. https://doi.org/10.1007/s00438-007-0270-9 PMID: 17724613

45. Kim HJ, Thyssen GN, Song X, Delhom CD, Liu Y. Functional divergence of cellulose synthase orthologs in between wild Gossypium raimondii and domesticated G. arboreum diploid cotton species. Cellulose. 2019; 26(18):9483–501.

46. Beisson F, Li Y, Bonaventure G, Pollard M, Ohlrogge JB. The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of Arabidopsis. The Plant Cell. 2007; 19(1):351–68. https://doi.org/10.1105/tpc.106.048033 PMID: 17259262

47. Hoffmann N, Benske A, Betz H, Schuetz M, Samuels AL. Laccases and peroxidases co-localize in ligni-fied secondary cell walls throughout stem development. Plant Physiology. 2020; 184(2):806–22. https://doi.org/10.1104/pp.20.00473 PMID: 32699027

48. Pandian BA, Sathishraj R, Djanaguiraman M, Prasad P, Jugulam M. Role of cytochrome P450 enzymes in plant stress response. Antioxidants. 2020; 9(5):454. https://doi.org/10.3390/antiox9050454 PMID: 32466087

49. Do C-T, Pollet B, Thévenin J, Sibout R, Denoue D, Barrière Y, et al. Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lig-nin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. Planta. 2007; 226(5):1117–29. https://doi.org/10.1007/s00425-007-0558-3 PMID: 17594112

50. Larkin RM. Tetrapyrrole signaling in plants. Frontiers in Plant Science. 2016; 7:1586. https://doi.org/10.3389/fpls.2016.01586 PMID: 27807442

51. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, et al. mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal. 2004; 37(6):914–39. https://doi.org/10.1111/j.1365-313x.2004.02016.x PMID: 14996223

52. Dufayard J-F, Bettembourg M, Fischer I, Droc G, Guiderdoni E, Périn C, et al. New insights on leucine-rich repeats receptor-like kinase orthologous relationships in angiosperms. Frontiers in Plant Science. 2017:381. https://doi.org/10.3389/fpls.2017.00381 PMID: 28424707

53. Hilgers R, Vincken J-P, Gruppen H, Kabel MA. Laccase/mediator systems: their reactivity toward phe-nolic lignin structures. ACS sustainable chemistry & engineering. 2018; 6(2):2037–46. https://doi.org/10.1021/acssuschemeng.7b03451 PMID: 29430340

54. Hu Q, Min L, Yang X, Jin S, Zhang L, Li Y, et al. Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. Plant physiol-ogy. 2018; 176(2):1808–23. https://doi.org/10.1104/pp.17.01628 PMID: 29229698

55. Yadav V, Molina I, Ranathunge K, Castillo IQ, Rothstein SJ, Reed JW. ABCG transporters are required for suberin and pollen wall extracellular barriers in Arabidopsis. The Plant Cell. 2014; 26(9):3569–88. https://doi.org/10.1105/tpc.114.129049 PMID: 25217507

56. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, et al. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabi-dopsis thaliana seedling. The Plant Journal. 2007; 50(4):660–77. https://doi.org/10.1111/j.1365-313X.2007.03078.x PMID: 17419845

57. Tanaka R, Kobayashi K, Masuda T. Tetrapyrrole metabolism in Arabidopsis thaliana. The Arabidopsis book/American Society of Plant Biologists. 2011; 9. https://doi.org/10.1199/tab.0145 PMID: 22303270

58. Cui B, Xu S, Li Y, Umbreen S, Frederickson D, Yuan B, et al. The Arabidopsis zinc finger proteins SRG2 and SRG3 are positive regulators of plant immunity and are differentially regulated by nitric oxide. New Phytologist. 2021; 230(1):259–74. https://doi.org/10.1111/nph.16993 PMID: 33037639

59. Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, et al. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in Arabidopsis. BMC plant biology. 2010; 10(1):1–19. https://doi.org/10.1186/1471-2229-10-95 PMID: 20500828

60. Hok S, Danchin EG, Allasia V, Panabières F, Attard A, Keller H. An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. Plant, cell & environment. 2011; 34(11):1944–57. https://doi.org/10.1111/j.1365-3040.2011.02390.x PMID: 21711359

61. Ascencio-Ibáñez JT, Sozzani R, Lee T-J, Chu T-M, Wolfinger RD, Cella R, et al. Global analysis of Ara-bidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant physiology. 2008; 148(1):436–54. https://doi.org/10.1104/pp.108.121038 PMID: 18650403

62. Ugalde JM, Lamig L, Herrera-Vásquez A, Fuchs P, Homagk M, Kopriva S, et al. A dual role for glutathi-one transferase U7 in plant growth and protection from methyl viologen-induced oxidative stress. Plant physiology. 2021; 187(4):2451–68. https://doi.org/10.1093/plphys/kiab444 PMID: 34599589

63. Marcos R, Izquierdo Y, Vellosillo T, Kulasekaran S, Cascón T, Hamberg M, et al. 9-Lipoxygenase-derived oxylipins activate brassinosteroid signaling to promote cell wall-based defense and limit patho-gen infection. Plant physiology. 2015; 169(3):2324–34. https://doi.org/10.1104/pp.15.00992 PMID: 26417008

64. Daněk M, Angelini J, Malínská K, Andrejch J, Amlerová Z, Kocourková D, et al. Cell wall contributes to the stability of plasma membrane nanodomain organization of Arabidopsis thaliana FLOTILLIN2 and

HYPERSENSITIVE INDUCED REACTION1 proteins. The Plant Journal. 2020; 101(3):619–36. https://doi.org/10.1111/tpj.14566 PMID: 31610051

65. Karunadasa S, Kurepa J, Smalle JA. Gain-of-function of the cytokinin response activator ARR1 increases heat shock tolerance in Arabidopsis thaliana. Plant Signaling & Behavior. 2022; 17 (1):2073108. https://doi.org/10.1080/15592324.2022.2073108 PMID: 35535663

66. Moon JY, Belloeil C, Ianna ML, Shin R. Arabidopsis CNGC family members contribute to heavy metal ion uptake in plants. International journal of molecular sciences. 2019; 20(2):413. https://doi.org/10.3390/ijms20020413 PMID: 30669376

67. Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, et al. Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proceedings of the National Academy of Sciences. 2011; 108(23):9709–14. https://doi.org/10.1073/pnas.1100958108 PMID: 21593420

68. Lang AG. The origin of lint and fuzz hairs of cotton. Journal of Agricultural Research. 1938; 56:507–21.

69. Hutchinson JB, Silow RA, Stephens SG. The evolution of Gossypium and the differentiation of the cultivated cottons. London: Oxford University Press; 1947. 175 p.

70. Xu Z, Yu J, Kohel RJ, Percy RG, Beavis WD, Main D, et al. Distribution and evolution of cotton fiber development genes in the fibreless Gossypium raimondii genome. Genomics. 2015; 106(1):61–9. https://doi.org/10.1016/j.ygeno.2015.03.002 PMID: 25796538

71. Kim HJ, Liu Y, Dowd M, Frelichowski J, Delhom C, Rodgers J, et al. Comparative phenotypic analysis of Gossypium raimondii with Upland cotton. Journal of Cotton Science. 2016; 20(2):132–44.

72. Ryser U, Meier H, Holloway P. Identification and localization of suberin in the cell walls of green cotton fibres (Gossypium hirsutum L., var. green lint). Protoplasma. 1983; 117(3):196–205.

73. He Z, Nam S, Fang DD, Cheng HN, He J. Surface and thermal characterization of cotton fibers of phenotypes differing in fiber length. Polymers. 2021; 13(7):994. https://doi.org/10.3390/polym13070994 PMID: 33804984

74. Ryser U, Holloway P. Ultrastructure and chemistry of soluble and polymeric lipids in cell walls from seed coats and fibres of Gossypium species. Planta. 1985; 163(2):151–63. https://doi.org/10.1007/BF00393501 PMID: 24249333

75. Sun J, Sun Y, Zhu Q-H. Breeding Next-Generation Naturally Colored Cotton. Trends in Plant Science. 2021. https://doi.org/10.1016/j.tplants.2021.03.007 PMID: 33858744

76. Serra O, Geldner N. The making of suberin. New Phytologist. 2022; 235:848–66. https://doi.org/10.1111/nph.18202 PMID: 35510799

77. Mutuku JM, Cui S, Hori C, Takeda Y, Tobimatsu Y, Nakabayashi R, et al. The structural integrity of lignin is crucial for resistance against Striga hermonthica parasitism in rice. Plant Physiology. 2019; 179 (4):1796–809. https://doi.org/10.1104/pp.18.01133 PMID: 30670602

78. Zhang S, Jia T, Zhang Z, Zou X, Fan S, Lei K, et al. Insight into the relationship between S-lignin and fiber quality based on multiple research methods. Plant Physiology and Biochemistry. 2020; 147:251–61. https://doi.org/10.1016/j.plaphy.2019.12.025 PMID: 31884241

79. Gao Z, Sun W, Wang J, Zhao C, Zuo K. GhbHLH18 negatively regulates fiber strength and length by enhancing lignin biosynthesis in cotton fibers. Plant Science. 2019; 286:7–16. https://doi.org/10.1016/j.plantsci.2019.05.020 PMID: 31300144

80. Liu Y, Kim H-J. Fourier Transform Infrared Spectroscopy (FT-IR) and Simple Algorithm Analysis for Rapid and Non-Destructive Assessment of Developmental Cotton Fibers. Sensors. 2017; 17(7):1469. https://doi.org/10.3390/s17071469 PMID: 28640185

81. Kim HJ, Liu Y, French AD, Lee CM, Kim SH. Comparison and validation of Fourier Transform Infrared spectroscopic methods for monitoring secondary cell wall cellulose from cotton fibers. Cellulose. 2018; 25(1):49–64.

82. Patel JD, Huang X, Lin L, Das S, Chandnani R, Khanal S, et al. The Ligon lintless-2 short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. Plant Physiology. 2020; 183(1):277–88. https://doi.org/10.1104/pp.19.01531 PMID: 32102829

83. Goujon T, Sibout R, Pollet B, Maba B, Nussaume L, Bechtold N, et al. A new Arabidopsis thaliana mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. Plant molecular biology. 2003; 51(6):973–89. https://doi.org/10.1023/a:1023022825098 PMID: 12777055

84. Zhang X, He Y, Li L, Liu H, Hong G. Involvement of the R2R3-MYB transcription factor MYB21 and its homologs in regulating flavonol accumulation in Arabidopsis stamen. Journal of experimental botany. 2021; 72(12):4319–32. https://doi.org/10.1093/jxb/erab156 PMID: 33831169

85. Calvo-Polanco M, Ribeyre Z, Dauzat M, Reyt G, Hidalgo-Shrestha C, Diehl P, et al. Physiological roles of Casparian strips and suberin in the transport of water and solutes. New Phytologist. 2021; 232 (6):2295–307. https://doi.org/10.1111/nph.17765 PMID: 34617285

86. Ryser U. Ultrastructure of the epidermis of developing cotton (Gossypium) seeds: suberin, pits, plasmo-desmata, and their implication for assimilate transport into cotton fibers. American journal of botany. 1992; 79(1):14–22.