Published March 13, 2015

A Transcriptome Profile for Developing Seed of Polyploid Cotton

Ran Hovav,* Adi Faigenboim-Doron, Noa Kadmon, Guanjing Hu, Xia Zhang, Joseph P. Gallagher, and Jonathan F. Wendel

Abstract

Cotton ranks among the world's important oilseed crops, yet relative to other oilseeds there are few studies of oil-related biosynthetic and regulatory pathways. We present global transcriptome analyses of cotton seed development using RNA-seq and four developmental time-points. Because Upland cotton (*Gossypium hirsutum* L.) is an allopolyploid containing two genomes (A/D), we partitioned expression into the individual contributions of each homeologous gene copy. Data were explored with respect to genic and subgenomic patterns of expression, globally and with respect to seed pathways and networks. The most dynamic period of transcriptome change is from 20–30 d postanthesis (DPA), with about 20% of genes showing homeolog expression bias. Co-expression analysis shows largely congruent homeolog networks, but also homeolog-specific divergence. Functional enrichment tests show that flavonoid biosynthesis and lipid related genes were significantly represented early and later in seed development, respectively. An involvement of new features in oil biosynthesis was found, like the contribution of *DGAT3* (diacylglycerol acyltransferase) to the total triglyceride expression pool. Also, catechin-based and epicatechin-based proanthocyanidin expression are reciprocally biased with respect to homeolog usage. This study provides the first temporal analysis of duplicated gene expression in cotton seed and a resource for understanding new aspects of oil and flavonoid biosynthetic processes.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2014.08.0041
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

ONE OF THE MOST important genera in global agriculture is the genus *Gossypium*, which includes the domesticated species of cotton. Unrivaled as natural source of textile fibers, produced in over 70 countries and playing a central role in global trade, four distinct species were independently domesticated some 4000 to perhaps 7000 yr ago, including two allopolyploids from the Americas (*G. hirsutum* and *G. barbadense* L.) and two African-Asian diploids (*G. arboreum* L. and *G. herbaceum* L.; Wendel et al., 2012). Among these, the vast majority of modern crop production is from *G. hirsutum*, or Upland cotton, due to its combination of high yield and superior fiber characteristics. Remarkably, domesticated cotton not only provides the world's most important source of textile fibers, but also is a vital oil- and protein-seed crop (Bewley, 2006; Liu et al., 2009). Cotton was initially introduced into the European market as an oilseed in the 19th century during the Industrial Revolution, due to rapid population expansion and shortages of fats and oils (O'Brien et al., 2005). Later, it became popular both in the Old and New World, and it was not until the end of World War II when cottonseed was outranked by soybean [*Glycine max* (L.) Merr.] as the world's leading source of vegetable oil. Today, cotton provides the sixth largest source of vegetable oil, and according to the National Cottonseed Products Association (http://www.cottonseed.com/, verified 2 Dec. 2014), annual cottonseed oil production in the

R. Hovav, A. Faigenboim-Doron and N. Kadmon, ARO (Volcani Center), Bet Dagan, Israel; G. Hu, X. Zhang, J.P. Gallagher, and J.F. Wendel, Iowa State Univ., Ames, IA. Received 25 Aug. 2014.*Corresponding author (ranh@agri.gov.il).

Abbreviations: ACP, acyl carrier protein; ANR, anthocyanidin reductase; ATP, adenosine triphosphate; BCCP, biotin carboxyl carrier protein; DAG, diacylglycerol; DE, differential expression; DGAT, diacylglycerol acyltransferase; DPA, days postanthesis; FA, fatty acid; FLS, flavonol synthase; GO, gene ontology; LAR, leucoanthocyanidin reductase; RPKM, reads per kilobase per million mapped reads; SNP, single nucleotide polymorphism; TAG, triglyceride.

United States averages more than 1 billion pounds, or about 5 to 6% of the total domestic fat and oil supply.

The oil fraction in mature upland cotton seeds is about 97% triglycerides (TAGs) and comprises approximately 22% of the dry seed weight. Dry seeds also contain about 20% protein and a low amount of starch (Hu et al., 2013; O'Brien, 2002). The oil is stored in the seed largely in the form of oil droplets in the cells of the cotyledons, which comprise most of the seed bulk, having consumed the endosperm during development and maturation. Similar to most members of the small Malvaceous tribe Gossypieae, which includes only eight genera, cotton embryos are large and complex in their development. The primary distinguishing feature is the large conduplicately folded cotyledons, which increase rapidly in size 15 to 25 d after anthesis, filling the embryo sac at about 30 d and completely enclosing an undifferentiated epicotyl meristem (Fryxell, 1979; Mauney, 1984). The endosperm, on the other end, is negligible in the mature seed. Cottonseed also contains a relatively thick seed coat which in some cases can reach ~32% of the total dry weight.

Cottonseed oil produced by upland cotton typically consists of 26% palmitic acid, 2% stearic acid, 13% oleic acid and 58% linoleic acid (Cherry, 1983). A modest level of variation in these percentages has been reported among natural accessions, as well as developmentally from 25 to 45 DPA (Jiao et al., 2013). In addition to these common fatty acids (FAs), cottonseed oil also contains low levels (0.5–1%) of minor FAs (palmitoleic, linolenic) and cyclopropenoid FAs, mainly malvalic, sterculic, and dihydrosterculic acids (Cherry, 1983; Jiao et al., 2013; Shenstone and Vickery, 1961). Genetic variability for seed oil content and FA composition has also been described within *Gossypium*. Gotmare et al. (2004) reported oil content variation from 10.3 to 22.9% in 22 domestic and wild species of *Gossypium*. Variation in oil content, FA composition, free FA, and phospholipid content have been found among *G. hirsutum*. Cotton germplasm accessions with higher oleic acid and stearic acid contents have been described as well (Liu et al., 2002).

Despite the relative importance of cottonseed as an oil crop, the molecular characterization of seed development in general, and oil related pathways in particular are lacking in this crop relative to other major oilseed plants. Several genes and gene families have been characterized, for example, SAD (Δ*9-stearoyl ACP desaturase*), FAD2 (Δ*-12 fatty acid desaturase 2*), and CPA (cyclopropane-fatty-acyl-phospholipid synthase; Liu et al., 2009; Yu et al., 2011). Global transcriptome analysis of developing seeds can provide fundamental molecular understanding of oil-related processes in cotton seeds, as well as more general processes such as embryogenesis, seed filling, maturation, and seed quality. This promise, using RNA-seq approaches, recently has been pioneered for several crop species, including soybean (Jones and Vodkin, 2013; Severin et al., 2010) and *Brassica* (Troncoso-Ponce et al., 2011), palm oil (Dussert et al., 2013), corn oil (Sekhon et al., 2013), and for cotton by Jiao et al. (2013). The latter study described the presence and relative abundance of over 17,000 genes in Upland cotton seeds at a single developmental stage (30 DPA), and separately for both the cotyledons and embryo axis. To extend this information over a broader developmental time frame, here we present the results of a global analysis of gene expression during seed development program in Upland cotton. Utilizing gene annotation data from the recently sequenced D-genome cotton species *G. raimondii* Ulbrich (Paterson et al., 2012), as well as other resources, we conducted transcriptomic analysis of seed pathways in developing upland cotton seeds using a time series of four, equally-spaced developmental stages. Because Upland cotton is an allopolyploid (AD-genome) derived from ancient hybridization (1 to 2 million year ago; Wendel et al., 2012) between two diploid genomes (A, D) and because we have access to diagnostic single nucleotide polymorphism (SNP) information from each model diploid progenitor genome, we were able to partition and separately characterize total duplicate gene expression for each gene pair into the individual contribution of each homeologous (duplicated) gene copy (i.e., A and D; Grover et al., 2012; Page et al., 2013a; Page et al., 2013b; Rambani et al., 2014; Yoo et al., 2013). Accordingly, the relative contribution of each subgenome to the global expression profile was estimated for the entire dataset and for specific pathways.

Table 1. The number of expressed genes (RPKM ≥ 2) in developing cotton seeds.[†]

Developmental stage, d postanthesis	Phenotypic characterization	No. of expressed genes	% of total
10	initial seed development	20,240	54
20	seed enlargement	20,396	55
30	fully developed	19,362	52
40	mature seed	18,573	49

[†] RPKM, reads per kilobase per million mapped reads.

Materials and Methods

Plant Material

Seeds of *G. hirsutum* Texas Marker Stock 1 (TM1) were collected from greenhouse-grown plants (12 h photoperiod; 22 and 28°C, respectively). Cotton flowers were tagged on the first day postanthesis and collected at four developmental stages, 10, 20, 30, and 40 DPA, representing different stages of seed filling and maturation (Table 1). At each time point seeds were extracted from developing fruits, and immediately thereafter fibers were manually removed from the seed surface, seeds were flash frozen, and were stored at –80°C for RNA extraction. Three biological replicates were taken from each time point.

RNA Extraction and Library Construction for Sequencing

Total RNA was extracted using the hot borate (sodium borate decahydrate) method, as previously described for peanut seed (Brand and Hovav, 2010), with minor modifications as follows. Tissue samples were ground in liquid N and were combined with 8 mL of borate buffer (0.2 M sodium borate decahydrate; 30 mM ethylene glycol tetraacetic acid; 1% (wt/vol) sodium dodecyl sulfate; 1% sodium deoxycholic acid; 10 mM dithiothreitol; 1% IGEPAL CA-630 [Nonidet P-40, NP-40]; 2% [wt/vol] polyvinylpyrrolidone-40) at 65°C. The homogenate was then incubated for 1.5 h in a 42°C, and 1 mL of 2 M potassium chloride was added to each sample. Subsequent to centrifugation, the supernatant was transferred to a 50-mL tube containing 8 M lithium chloride and incubated on ice overnight. Following a second centrifugation, the supernatant was discarded and the pellet was washed a few times with 2 M lithium chloride. The pellet was then suspended in 250 µL Tris-EDTA and warmed to room temperature. Each sample was centrifuged again, and the supernatant was transferred to a 1.5-mL tube containing 2 M potassium acetate. After an additional centrifugation, the pellet was discarded and the supernatant was transferred to a tube containing 3 M sodium acetate and 2.5 volumes of cold 100% ethanol. Following centrifugation, the supernatant was discarded, the pellet was washed with 1 mL of 70% ethanol, and the RNA was resuspended in 40 µL of diethylpyrocarbonate-treated water and stored at −80°C.

The quality and concentration of extracted RNAs were determined using the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). mRNA purification was performed using the MicroPoly(A) Purist kit (Ambion, Austin, TX), and then processed for RNA-Seq library construction following a suggested protocol (Nagalakshmi et al., 2010). The 12 constructed libraries, indexed (bar-coded) with six nucleotide sequences, were pooled together in equimolar amounts and were sequenced on the Illumina Hi-Seq 2000 sequencer (Illumina, San Diego, CA) with 100 base single reads at the Genomics Core Facility at Iowa State University. All reads are available for download on NCBI bioproject (http://www.ncbi.nlm.nih.gov/bioproject/179447, verified 17 Nov. 2014).

Analysis of RNA-seq Data: Mapping and Differential Expression

Raw reads were sorted into the corresponding genome accession and time point according to their individual barcodes using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html, verified 17 Nov. 2014). To remove low-quality nucleotides, indexed reads were cleaned by trimming off 10 bases from the 5′ end and 15 bases from the 3′ end using FASTX trimmer. High quality reads having at least 70% of bases with quality score $\geq$ 23 were retained. Poly-A tails were removed using the trimest function of the European Molecular Biology Open Software Suite (EMBOSS; v. 6.4.0). Reads were mapped to plant rRNA, chloroplasts, and mtDNA with BWA to discard noncoding RNA.

Cleaned reads were mapped to the reference cotton genome (D-genome *G. raimondii*; 37,223 genes; Paterson et al., 2012) using GSNAP (Miller et al., 2008) with this command: gsnap-N 1-n 1-Q-t 5-B 5-D/resources/gmap_db-d Dgenome2_13-v Dgenome2_13.snp1<fastq>-A sam>out.sam. This yielded a number of .sam alignment files (Li et al., 2009). Next, PolyCat software (Page et al., 2013a) was used to assigns reads into new bam files. Then the program "counter," part of the BamBam package (https://sourceforge.net/projects/bambam, verified 17 Nov. 2014), was used to count the reads per gene and generates the .counts files for each .bam file.

Exons of coding sequences for homeologs in allopolyploid cotton differ, on average, at slightly <1% of their nucleotides (Flagel et al., 2013). The genomic distribution and numbers of At vs. Dt SNPs has been previously described by Page et al. (2013a). The following values were calculated for each gene: A-genome reads (At), D-genome reads (Dt), reads with no SNPs (N), and total number of reads (T = At + Dt + N). These data were subjected to normalization by PolyCat as described (Page et al., 2013a), ending with reads per kilobase per million mapped reads (RPKM) values for each bin (Supplemental Table S2). These values were used for cluster analysis and for the differentially biased expression analysis (by comparing At and Dt).

Differential expression (DE) analysis was conducted using DESeq package in R software v.3.0.1 (R Foundation for Statistical Computing, Vienna, Austria) based on non-RPKM-normalized read counts. Differential expression was assessed between the time points and between subgenomes. The distribution of *P*-values was controlled for a false discovery rate by the BH method (Benjamini and Hochberg, 1995) at $\alpha = 0.05$. To remove the negative effect of background expression noise on DE calls, we restricted analysis to genes having read counts RPKM $\geq$ 2 in all accessions and biological replicates in each contrast. In the homeolog expression bias for duplicated genes associated with oil and anthocyanin biosynthesis analyses, the *p*-value cutoff at $\alpha < 0.05$ was used.

Functional Analysis of Differentially Expressed Genes

Gene ontology (GO) enrichment tests were performed by using the GO enrichment tool of the AgriGO toolkit for agricultural community (Du et al., 2010). Fisher exact tests were used for the determination of significant biological processes or cellular components (Qvalue < 0.1).

For co-expression network analysis and construction, we used a list of 9910 individual genes that were DE during seed development. Expression values of these genes across all developmental stages were used to calculate pairwise Pearson correlation coefficients. Two genes were regarded as being connected when the correlation coefficient < 0.9. Then the PageRank algorithm (Page et al., 1998) was employed to rank the importance of genes in this co-expression network. The α value (damping parameter) of the PageRank analysis was set to be 0.85, and all edges were weighted equally.

Hierarchical clustering of RPKM normalized gene expression was performed on log2 scale and visualized on heatmaps in R using the gplots package (http://www.R-project.org, verified 17 Nov. 2014), specifying complete linkage and Euclidean distance metric. Boundaries of the clustering were placed by applying a vertical cut on the dendrogram.

Oil Content and Fatty Acid Profiling

The oil content of developing seeds was determined as follows. From each of the developmental stages, three to 10 seeds (depending on the size and developmental stage of the seeds) were weighed and kept overnight in a 60°c oven. Samples were then reweighed and ground with a mortar and pestle. Samples were placed in preweighed 2 mL tubes, and 1.5 mL of hexane was added to each tube. Samples were mixed vigorously with a vortex for 20s and incubated with shaking (Intelii-Mixer RM-2, ELMI Ltd., Russia) for 1 h. Following incubation, the samples were centrifuged for 10 min at 13,000 rpm, and the upper suspensions were transferred to new preweighed tubes. This process was repeated for the cottonseed residue, followed by overnight evaporation of the hexane and a final weighing of the tubes.

From each sample, 0.06 g oil was diluted with 0.3 mL chloromethane and 2 mL sodium methoxide, heated to 50°C for 30 min, and cooled to room temperature. Two percent acetic acid and 2 mL internal standard solution (C17-methylhepatadecanoate) was added. Following centrifugation (5000 rpm, 5 min) the upper phase was recovered using a glass syringe and then filtered through a 0.45 μm PTFE filter. The extract was dried at 37°C under a slow stream of N_2 gas. The gas chromatography analyses were performed on an Agilent 6890 GC-FID apparatus equipped with an SP-2560 (Supelco, 100 m × 0.25 mm i.d., 0.20 μm film thickness) bicyanopropyl polysiloxane fused-silica capillary column. The injector was kept at 250°C and the detector at 260°C. Helium at a constant pressure of 294.4 kPa was used as a carrier gas with retention time locking. The column was maintained at 50°C for 2 min and then programmed to 190°C at 5°C/min anda c kept for 20 min at 190°C. Identification of compounds was performed by comparing their relative retention indices with those of authentic samples.

Results and Discussion

Oil Accumulation in Cotton Seed

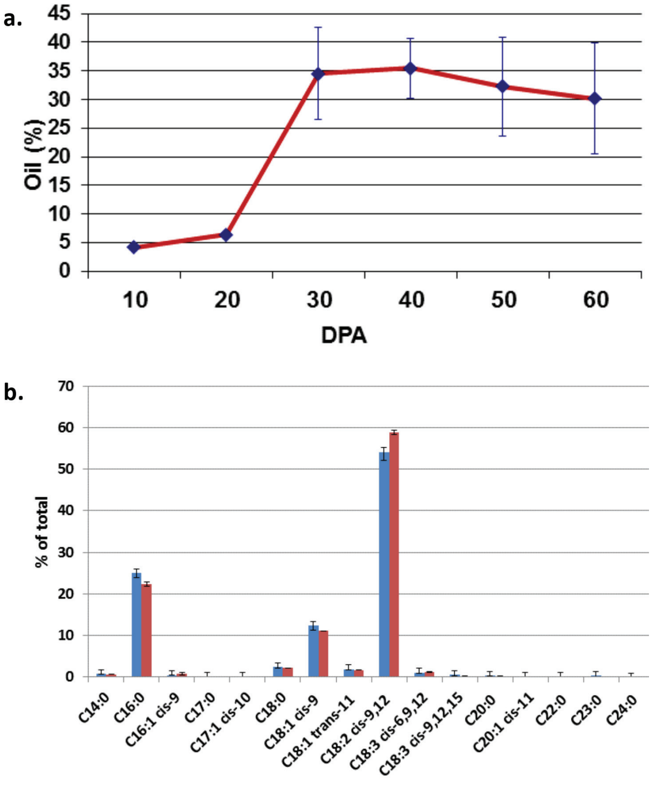

Figure 1. (a) Seed oil content of TM1 cotton during development. (b) Fatty acid profile of oil derived from 30 d postanthesis (DPA; blue) and 40 DPA (red) developing seeds.

To optimize experimental design for sampling developmental stages for seed oil and mRNA accumulation, total oil content of seeds derived from the cultivated cotton tetraploid accession TM1 was measured. As shown in Fig. 1a, cotton seed oil begins to accumulate as early as 20 DPA, and after 30 DPA the majority of oil has already accumulated in the seeds. This process of oil accumulation corresponds to rapid embryo growth, as described in previous studies (Dure, 1975; Reeves and Beasley, 1935). Also, at this timepoint, cotton seed trichomes (fibers) usually finish elongation (Hovav et al., 2008b), and in fact, oil accumulation appears more synchronous with fiber secondary cell wall thickening than with fiber elongation. As gene expression precedes oil accumulation, we decided to perform RNA expression analyses on four developmental stages (10, 20, 30, and 40 DPA), which encompass the gene expression changes involved with oil production and accumulation through the period of oil biosynthesis.

Overview of the Cotton Seed Transcriptome

To investigate transcript accumulation in allopolyploid cotton during seed development, 37,223 genes were surveyed for their expression using RNA-Seq. A total of 240 million (M) reads were generated from 12 Illumina libraries, with an average of 69% of reads mapped to the reference D-genome cotton assembly (Paterson et al., 2012). Based on the criterion of RPKM ≥ 2, on average, 52.5% of the 37,223 annotated cotton genes were expressed during seed development (Table 1). Notably, the number of expressed genes is slightly higher in the first two stages of seed development and decreases during the later stages (from 20,240 at 10 DPA to 18,537 at 40 DPA).

To profile the transcriptomic dynamics during seed development, the numbers of DE genes were evaluated between adjacent developmental time points (Table 2). Interestingly, over 65% of the DE genes were during the time phase from 20 to 30 DPA. In contrast, the temporal transition between 30 DPA and 40 DPA showed a low

Table 2. Differential gene expression during cotton seed development.[†]

Stage compared, d postanthesis	ADt		At		Dt		(A + D)t	
	No. of DE genes	% of total	No. of DE genes	% of total	No. of DE genes	% of total	No. of DE genes	% of total
10 vs. 20	2667	7.1	2330	6.2	2113	5.6	4430	5.9
20 vs. 30	5538	14.8	4772	12.8	4321	11.6	9040	12.1
30 vs. 40	238	0.6	143	0.4	133	0.35	275	0.7

[†] Shown are the numbers of differentially expressed (DE) genes when treated as duplicated homeologs (ADt column) or as single At or Dt homeologs. The totals of the At and Dt columns are shown in the (A + D)t column.

number of DE genes. This pattern is consistent with our observations showing little difference in oil content (Fig. 1a), FA profile (Fig. 1b), and general phenotypic appearance between these two developmental stages.

Since upland cotton is an allopolyploid species, wherein two duplicated copies (A and D homeologs) are present for nearly all genes, the actual number of genes is approximately twice that of the diploid genomes (~74,400). Based on the SNP information between A and D homeologs (Page et al., 2013a), we were able to map reads to their homeolog of origin based on genome-diagnostic SNPs, and thereby specify gene expression levels of each copy of duplicated genes. Therefore, in addition to total gene expression for each homeologous pair, DE within At and Dt subgenomes was determined separately, as shown in Table 2 (At and Dt columns). One limitation of this analysis is that not all reads from allotetraploid cotton cover a genome-diagnostic SNP; these reads, therefore, could be derived from transcripts of either the At or Dt copy. The number of such reads varies by gene, depending on the number of SNPs that accumulated during diploid divergence; some genes, therefore, have low SNP coverage, and accordingly, most reads from such genes are of unknown genomic origin. To inspect this effect of data reduction on the global gene expression, homeolog-specific RNA-seq reads were analyzed based on the allopolyploid gene models (74,446 genes). As shown in Table 2, the percentage of DE genes ([A+D]t columns) is 10 to 30% lower than that derived when using all mapped RNA-seq reads (ADt columns), which considers the total expression for each pair of homeologs (ADt columns) based on the diploid gene models (37,223 genes). However, in absolute numbers, analyzing each homeolog separately leads to a higher number of DE genes.

There are several key benefits of using the doubled (A+D)t dataset (74,446 genes) instead of the "total" counts (gene pairs), including the obvious one of having insight into actual gene copies used during development. One illustration of this power is shown in Supplemental Table S1, which presents a GO enrichment analysis for DE genes in seeds during development for ADt vs. (A+D)t. In general, more biological processes were found by using the (A+D)t counts than using the total counts. Moreover, the separation of the data set into At-specific and Dt-specific counts enabled the classification of each enriched GO term as At-specific, Dt-specific, or "general." For example, genes that were up-regulated at 20 DPA relative to 10 DPA in the (A+D)*t* test were found to be enriched with respect to auxin transport, response to sucrose stimulus, hydrolase activity; these were not significant when the ADt data were used. Within this group, auxin transport was specific to the Dt subgenome and hydrolase activity to the At subgenome (Supplemental Table S1). Therefore, we concluded that global gene expression analysis should be performed on the (A+D)t data, so all further analyses reflect this subgenome-specific dataset.

Table 3. Number of genes exhibiting homeolog expression bias during cotton seed development.

Developmental stage, d postanthesis	Biased gene pairs	% of total gene pairs	Bias toward A genome	Bias toward D genome
10	7595	20.4	3925	3670
20	7552	20.2	3882	3552
30	6649	17.9	3382	3267
40	6050	16.3	3105	2945

In addition to detailing the transcriptomic dynamics during seed development, the homeolog-level analysis also allowed us to examine unequal expression between homeologous genes, a well-appreciated phenomenon known as homeolog expression bias in allopolyploid species (Grover et al., 2012). At each developmental time point the numbers of genes that are differentially biased toward the At- or Dt- subgenome were tabulated based on normalized homeolog-specific read counts. A relatively high fraction of gene pairs exhibited biased expression toward one subgenome, ranging from 16.3 to 20.4% of the whole genome during seed development (Table 3). A slightly higher and significant level of expression bias toward the A-genome ($p < 0.05$; chi-squared tests) was found for the entire transcriptome, in all developmental stages, besides 30 DPA ($p = 0.1584$). Notably, this result is different from what was previously reported for cotton seed trichomes (fibers), which showed significantly more Dt-bias expressions in one study (Hovav et al., 2008a) and balanced biases toward subgenomes in another study (Yoo and Wendel, 2014), suggesting that the expression of homeologous gene pairs are differentially regulated between these two tissues. Although there was no biological process enriched between genes exhibiting At- and Dt- biases, interestingly, a significant enrichment of the cytoplasmic cellular component category (Q < 0.1) was found for the At-biased genes at 30 and 40 DAP (data not shown). This actually can explain the bias in expression toward the At genome.

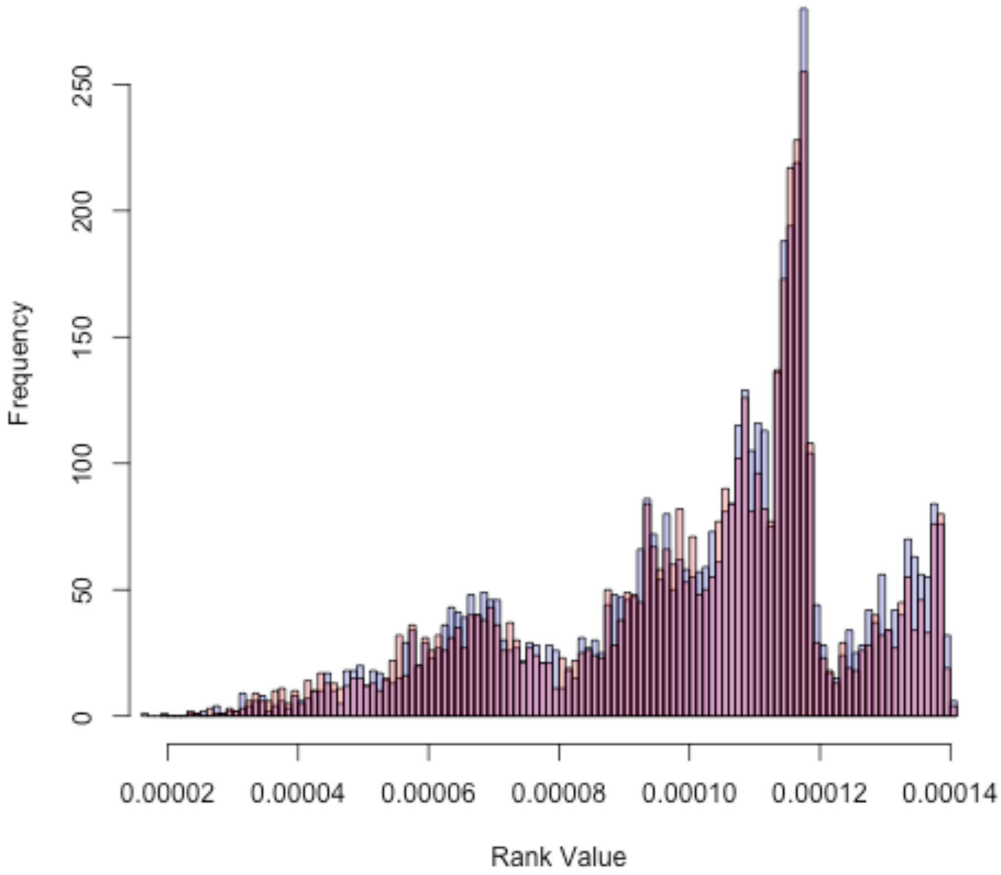

Figure 2. PageRank histogram of 9910 genes that were differentially expressed during seed development. Blue = At; pink = Dt; purple = overlap.

Co-expression Network Analysis of DE Genes during Seed Development

A co-expression network of DE genes during seed development was constructed and analyzed, to discriminate their interacting patterns between two subgenomes (Fig. 2). For this purpose, a list of 9910 individual genes that were DE in their homeologous expression levels during seed development was used to construct the co-expression network. PageRank analysis on the DE × DE co-expression network revealed a similar rank distribution for the At (the orange section in Fig. 2) and Dt genes (the blue section in Fig. 2) with a large overlapping area (the purple section in Fig. 2). Interestingly, the rank distribution of both subgenomes had three waves: 0.00012 to 0.00014 ("high-end"), 0.00008 to 0.00012 ("medium"), and 0.0002 to 0.0008 ("low-end"), with respect to their connectedness to other genes (Fig. 2). The actual meaning of the rank value is a probability associated with the importance of a node in a network. So the high-end wave may refer to the core of the network, medium might be intermediately spread genes, and low-end might be the outskirts of the network. The highly overlapping At and Dt data means that both subgenomes share the same network programming in the developing cotton seed. Indeed, GO enrichment analysis (Table 4) showed that both subgenomes share significant biological processes in the core section of the network (high-end, Group A in Table 4), including FA biosynthesis and response to hormone stimulus processes; the medium part (Group B in Table 4) was enriched with anthocyanin metabolic processes and receptor protein signaling pathway; and the low-end (Group C in Table 4) did not show particular processes in either subgenome.

One additional aspect of these results is that the specific gene list of each grouping for the At and the Dt subgenomes showed specific or local bias in the co-expressed network (Table 4). The total number of DE genes is not equal between At and Dt; of 9910 DE genes, 5129 belong to At and 4781 to the Dt subgenome, of which 59% overlapped between At and Dt with the remaining being only one homeolog, either At or Dt. Moreover, as shown in Table 4, many gene pairs did not fit the same network groups. For example, 70 genes were in Group A (high-end) in the At subgenome but in Group C (low-end) in the Dt subgenome, whereas 22 genes were in Group A in the Dt subgenome but in Group C in the At subgenome. Interestingly, the latter was enriched in lipid metabolic process-related genes, indicating homeolog-specific network divergence associated with this important biosynthetic pathway in oilseeds.

Table 4. Divergence of 9910 differentially expressed genes between three networking level groups. At, subgenome A; Dt, subgenome D; a, high-end networking group; b, middle networking group; c, low-end networking group. In each box, the number of shared genes for each comparison is indicated above. The significantly enriched processes (if they exist) are indicated below.

	At-a	At-b	At-c	Only Dt	All Dt
Dt-a	334 –	118 –	22 lipid metabolic processes	194 transcription factor activity	668 fatty acid biosynthesis; response to hormone stimulus
Dt-b	140 regulation of localization	1547 phenylpropanoid biosynthetic process	188 –	1235 –	3110 phenylpropanoid biosynthetic process; receptor protein signaling
Dt-c	70 –	181 meristem structural organization; transmembrane receptor tyrosine kinase signaling	331 –	420 –	1002 –
Only At	283 –	1397 –	518 regulation of transcription; determination of bilateral symmetry		
All At	827 fatty acid biosynthesis; response to hormone stimulus	3243 phenylpropanoid biosynthetic process; receptor protein signaling	1059 –		

Functional Analysis of Abundantly Expressed Genes

To explore key biological processes during seed development, one promising strategy is to characterize the most abundant transcripts and identify their associated pathways, such as recently applied to soybean seed transcriptomes (Jones and Vodkin, 2013). In our cotton seed transcriptomes, a total of 368 highly expressed genes with a RPKM > 500 in at least one of the four developmental stages were included for this analysis. The accumulated expression (RPKM read counts) of these abundant transcripts during seed development is shown in Fig. 3. In general, highly abundant transcripts were expressed later in seed development, and this pattern is consistent between A- and D- homeologs. However, this is not the case with respect to seed storage proteins (red bars in Fig. 3) and all other highly expressed genes (blue bars in Fig. 3). It is known that the major storage proteins in cotton seed include globulins (vicillin A, vicillin B, legumin A, and legumin B) and albumins (Hu et al., 2011). The genes encoding these storage proteins were barely expressed at 10 and 20 DPA, while at 30 and 40 DPA their transcripts accounted for 50% of the total pool (Fig. 3, top panel). Notably, a significant expression bias toward the D-genome homeologs was observed (Fig. 3, middle and bottom panels). For example, at 40 DPA, storage proteins comprise ~80% of total expression in the Dt subgenome, but only ~20% in the At subgenome. A similar observation was reported at the proteomic level (Hu et al., 2011), as well as in cotton embryos, by Jiao et al. (2013). In contrast to storage proteins, increased expression of other highly expressed genes was observed only for At homeologs, while expression of the Dt homeologs maintained the same along seed development and even slightly decreased.

To further categorize the DE profiles of these expressed genes, a hierarchical clustering analysis was performed on the nonpartitioned expression values and visualized on a heatmap (Fig. 4). Functional enrichment tests revealed that the flavonoid biosynthesis (Cluster 2) and lipid localization related genes (Cluster 3) were significantly represented at early and later seed developmental stages, respectively. These results emphasize the prevalence of the flavonoid and lipid processes in seed development, as discussed further below.

Analysis of Lipid Biosynthesis Related Genes in Cottonseed

As a prerequisite to examine expression patterns of lipid metabolic processes in cotton, a list of 673 oil-related genes belonging to 159 families assembled from cotton (Jiao et al., 2013) and other oilseed crops (http://aralip.plantbiology.msu.edu/pathways/pathways, verified 17 Nov. 2014; Troncoso-Ponce et al., 2011), was used to annotate lipid metabolic processes, including pyruvate biosynthesis, FA biosynthesis, and TAG biosynthesis processes, as well as oil localization, oil degradation, and oil regulatory factor genes (Supplemental Table S3). In addition, 424 cotton genes involved in lipid precursor synthesis from sucrose to pyruvate were also annotated based on Troncoso-Ponce et al. (2011) and included for analysis. The overall expression of these genes declined from 20 to 40 DPA (Fig. 5). However, this pattern was mainly represented by genes associated with the pyruvate precursors group, while expression of genes associated with FA biosynthesis and TAG assembly slightly increased during seed development. Genes from both the At and Dt subgenomes exhibited similar expression patterns, although higher ($p < 0.0001$; Chi-square tests)

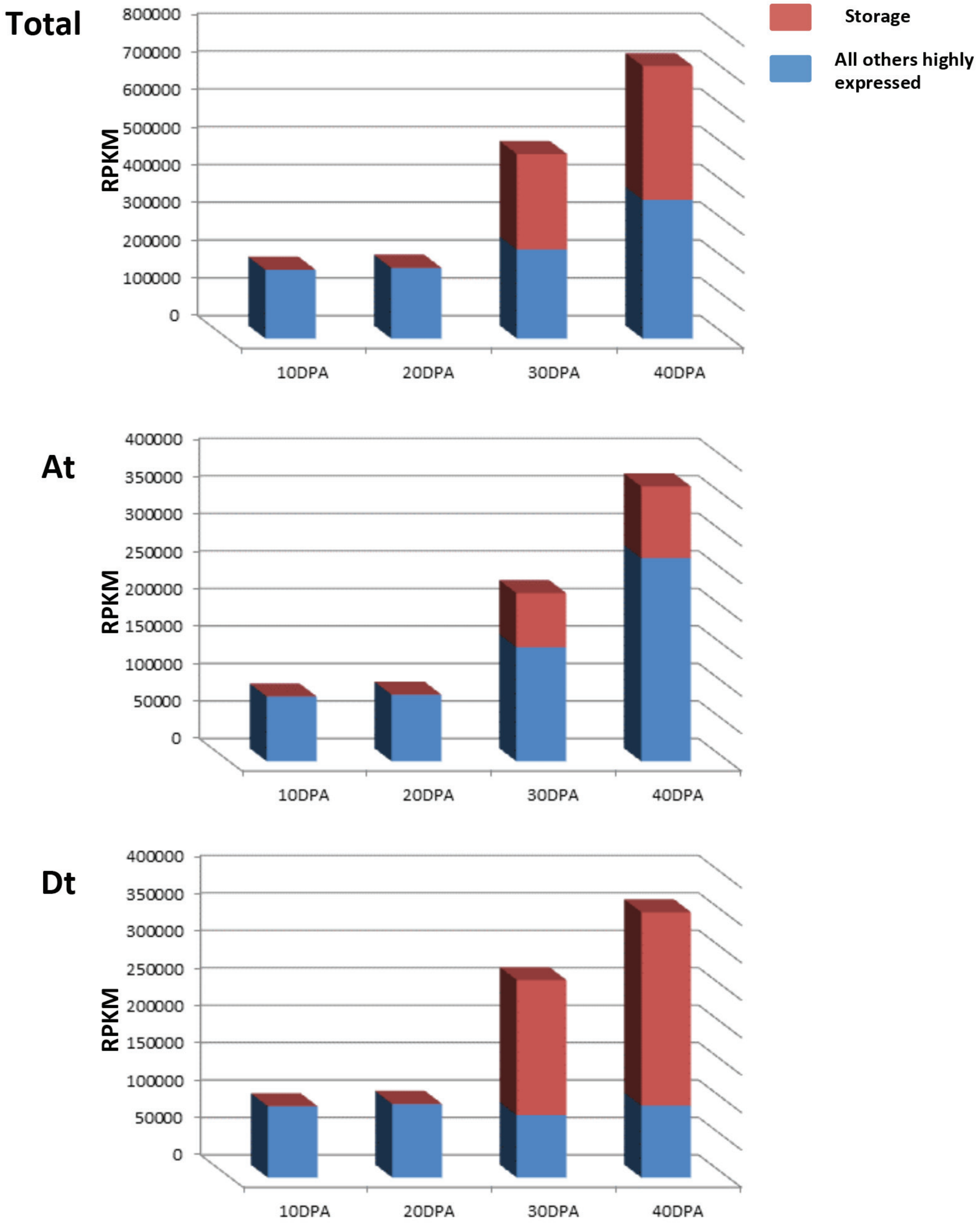

Figure 3. Global expression (reads per kilobase per million mapped reads, RPKM) of 368 highly expressed genes (RPKM > 500) in the nonpartitioned expression levels (Total) and for the homeologous subgenome (At and Dt, respectively). The column is divided into storage proteins (red) and all others highly expressed genes (blue).

expression levels for FA synthesis genes were found in the Dt homeologs (red bars in Fig. 5). For the At homeologs, lipid biosynthesis related genes comprised between 0.54 to 0.77% of the total expression pool, while the range for the Dt homeologs was between 0.70 to 1.00%.

Interestingly, about 27% of cotton genes annotated as "lipid metabolism related genes" were not expressed during seed development (Supplemental Table S3). Among the expressed genes, 28% exhibited biased expression (Table 5) in at least two of the four developmental stages sampled. In agreement with the global pattern of homeolog expression bias, more lipid related genes were biased toward the At than the Dt homeolog (Table 5). Yet, some specific lipid biosynthesis related processes that include FA biosynthesis (within and external to the plastid) and FA desaturation, had more genes that were biased toward the Dt homeolog.

Temporal Expression Patterns of FA and TAG Biosynthesis Genes in Cottonseed

Our suggested oil biosynthesis model of the cotton seed is presented in Fig. 6.

In many oilseed plants, the activity of the pyruvate dehydrogenase complex (PDHC), comprising four enzyme subunits (E1-a and -b, E2, and E3), provides the acetyl-CoA precursor required for de novo FA synthesis

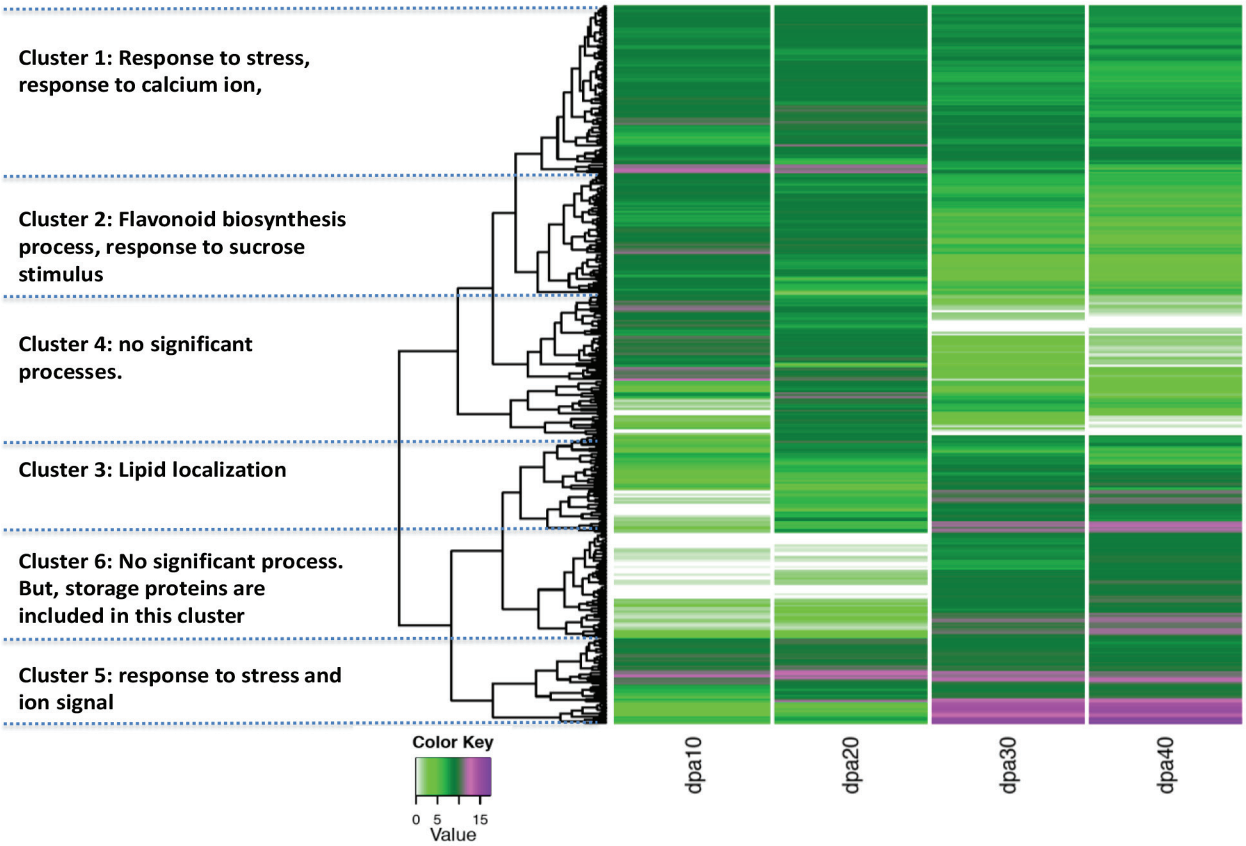

Figure 4. Hierarchical clustering of highly expressed genes. Enriched gene ontology categories were shown by each cluster (false discovery rate < 0.1). Value of the color key refers to the log base 2 of gene expression (reads per kilobase per million mapped reads, RPKM).

(Baud and Lepiniec, 2010). Alternative enzymes that might also provide acetyl units for FA synthesis were found, including adenosine triphosphate (ATP) citrate lyase and acetyl-CoA synthase (Lin and Oliver, 2008; Ratledge et al., 1997). Interestingly, in cotton seeds, ATP citrate lyase was expressed five times higher than the expression level of PDHC, particularly at the beginning of seed development (Fig. 7a), a developmental stage that was previously suggested to be important for acetyl-CoA synthesis in several oil crops (Troncoso-Ponce et al., 2011). This suggests an alternative route for acetyl-CoA biosynthesis in cotton seeds.

Acetyl-CoA Carboxylase (ACCase) catalyzes the first metabolic step of de novo FA biosynthesis in the plastid by adding one carboxyl group to the acetyl-CoA to form malonyl-CoA. Our analysis showed that the three subunits of the heteromeric ACCase enzyme, α-carboxyltransferase (CT), biotin carboxylase (BC), and biotin carboxyl carrier protein (BCCP) displayed a coordinated temporal pattern (Fig. 7b), as shown earlier for *Arabidopsis thaliana* (L.) Heynh. (Baud and Lepiniec, 2009) and other oilseeds (Troncoso-Ponce et al., 2011). The BCCP protein is the most abundant among the three ACCase subunits, and its BCCP2 isoform is significantly more abundant than BCCP1 (Fig. 7b), suggesting a single isoform expression in cotton seeds, unlike with what has been suggested for Arabidopsis (Thelen et al., 2001).

Plastidial acetyl-CoA and malonyl-CoA are converted into long-chain acyl molecule via a chain of reactions containing several enzymes with ACP (acyl carrier protein) as a cofactor. The temporal expression patterns for ACP and the FA synthesis enzymes mirrored oil accumulation in the seed, showing an increase in expression until 30 DPA (Fig. 7c and 7d). These results differ from those of some other studies that showed a steady decline in relative abundance of many oil biosynthesis expressed sequence tags during seed development (Troncoso-Ponce et al., 2011). This can be explained by the fact that, in our case, the bell-shaped expression pattern was based on samples that included seed coat and endosperm, similar to a previous study on developing seeds in Arabidopsis that showed the same bell-shaped expression pattern (Ruuska et al., 2004), and in contrast to Troncoso-Ponce et al. (2011), which used purified embryos. A small decrease in expression during seed development was actually observed for the two genes in the pathway that encode the two acyl-ACP thioesterases that terminate plastid FA synthesis, *FATA* and *FATB* (Voelker, 1996; Fig. 7e). In general, *FATA* expression level was higher than *FATB*, consistent with greater plastid production of unsaturated than saturated FAs in cotton seeds (Fig. 1b), since FATA has higher affinity for unsaturated fats than FATB (Jones et al., 1995).

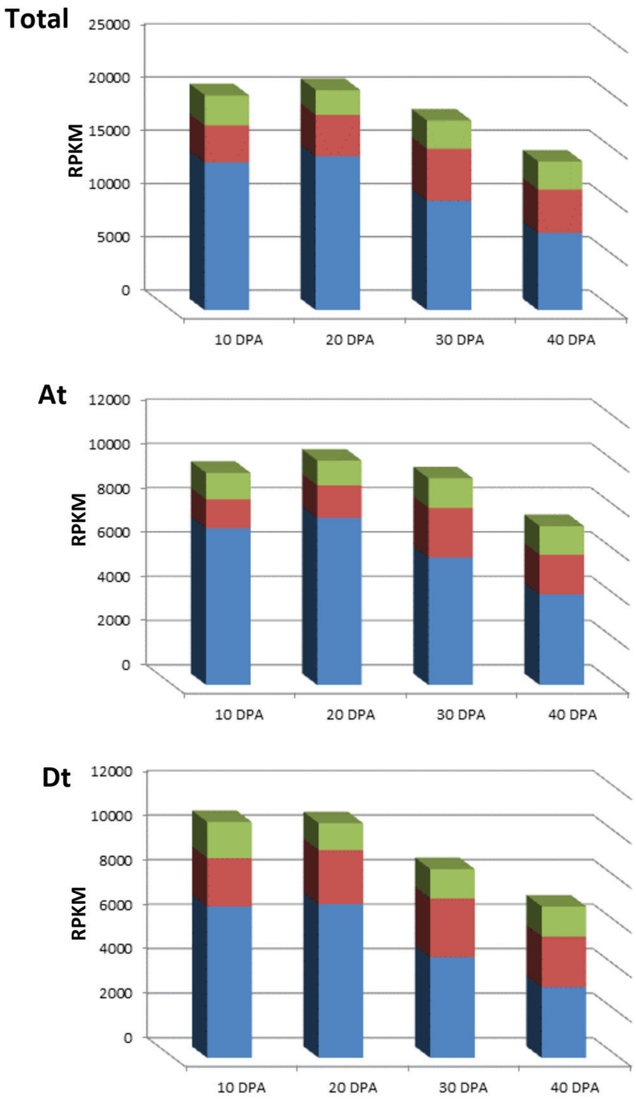

Figure 5. Global expression (reads per kilobase per million mapped reads, RPKM) of 1097 lipid biosynthesis related genes in the nonpartitioned expression levels (Total) and for the homeologous subgenomes (At and Dt). Each column is divided into genes related to pyruvate biosynthesis from sucrose (blue), fatty acid biosynthesis processes (red), and TAG and membrane biosynthesis (green).

Table 5. Number of significantly At- and Dt- biased genes in lipid-related biosynthetic processes. At- or Dt-bias is called when at least two of four developmental stages exhibit biased homeolog expression ($P < 0.05$).

	N significantly biased		% of total genes in the group	
Lipid biological process	A	D	A	D
Plastid fatty acid synthesis from pyruvate	15	22	21.4	31.4
Fatty acid desaturation (out of plastid)	0	2	0	33.3
Fatty acid elongation	4	3	17.4	13.0
Eukaryotic glycerolipid synthesis	13	11	19.1	16.1
Other acyl lipid related	7	3	35.0	15.0
TAG synthesis†	2	0	10.0	0
Oil storage	5	3	23.8	14.2
Plastidial glycerolipid, galactolipid, and sulfolipid synthesis	11	1	23.9	2.1
Plastid lipid trafficking	2	0	40.0	0
Mitochondrial fatty acid and lipoic acid synthesis	2	5	8.3	20.8
Sphingolipid synthesis	17	9	18.6	9.1
TAG degradation	2	2	12.5	12.5
β-oxidation	7	5	17.9	12.8
Lipid transfer proteins	19	11	21.8	12.5
Transcription factors	2	0	11.1	0
Miscellaneous	10	6	35.7	21.4
Total or mean	118	83	16.8	11.3

† TAG, triglyceride.

Another unexpected feature of the oil biosynthesis pathway in cotton, as revealed by our study, is the preferred usage of *DGAT3* (Acyl-CoA: Diacylglycerol Acyltransferase 3) for the total TAG biosynthesis gene expression pool (Fig. 7f). Diacylglycerol acyltransferase (EC 3.2.1.20) catalyzes the last acylation step from diacylglycerol (DAG) to TAG and has been suggested as an important step in plant storage lipid accumulation (Ichihara et al., 1988). Three different *DGAT* gene family members are described in oilseed plants, where *DGAT1* and *DGAT2* expression are more prominent in seeds (Liu et al., 2012; Lung and Weselake, 2006). The third, *DGAT3*, initially described in peanut (*Arachis hypogaea* L.), catalyzes the acylation of monoacylglycerol to DAG and the further acylation of DAG to TAG by the action of a cytosolic DGAT (Saha et al., 2006). This gene usually is expressed in vegetative tissue, particularly in developing seedlings (Hernandez et al., 2012), whereas its expression in developing seeds is not common among oilseed plants (Saha et al., 2006). An alternative enzyme catalyzing acyl-CoA-independent synthesis of TAG by the phospholipid:DAG transacylase activity has also been described in the *PDAT* (*Phospholipid:diacylglycerol acyltransferase*) gene family (Dahlqvist et al., 2000). Mutants analyses in Arabidopsis indicated that PDAT can partially compliments the function of DGAT in developing seeds (Mhaske et al., 2005; Zhang et al., 2009). In our study, *DGAT1* and *DGAT2* had minor expression levels in the developing seeds, while *DGAT3* had at least ten times more expression than *PDAT* (Fig. 7f), indicating the unique contribution of *DGAT3* to the alternative route for incorporation of FAs into TAG in the cotton seed.

Homeolog Expression Bias for Duplicated Genes Associated with Oil Biosynthesis and Accumulation

Because of the allopolyploid nature of cotton, we were interested in the specifics of homeologous gene expression associated with oil biosynthesis. For this purpose, we focused on FA biosynthesis, elongation, desaturation, and TAG biosynthesis genes. Only genes that were significantly up-regulated at 20 and 30 DPA relative to 10 DPA were studied, as were genes with a RPKM > 5 in at least one time point during seed development. These criteria

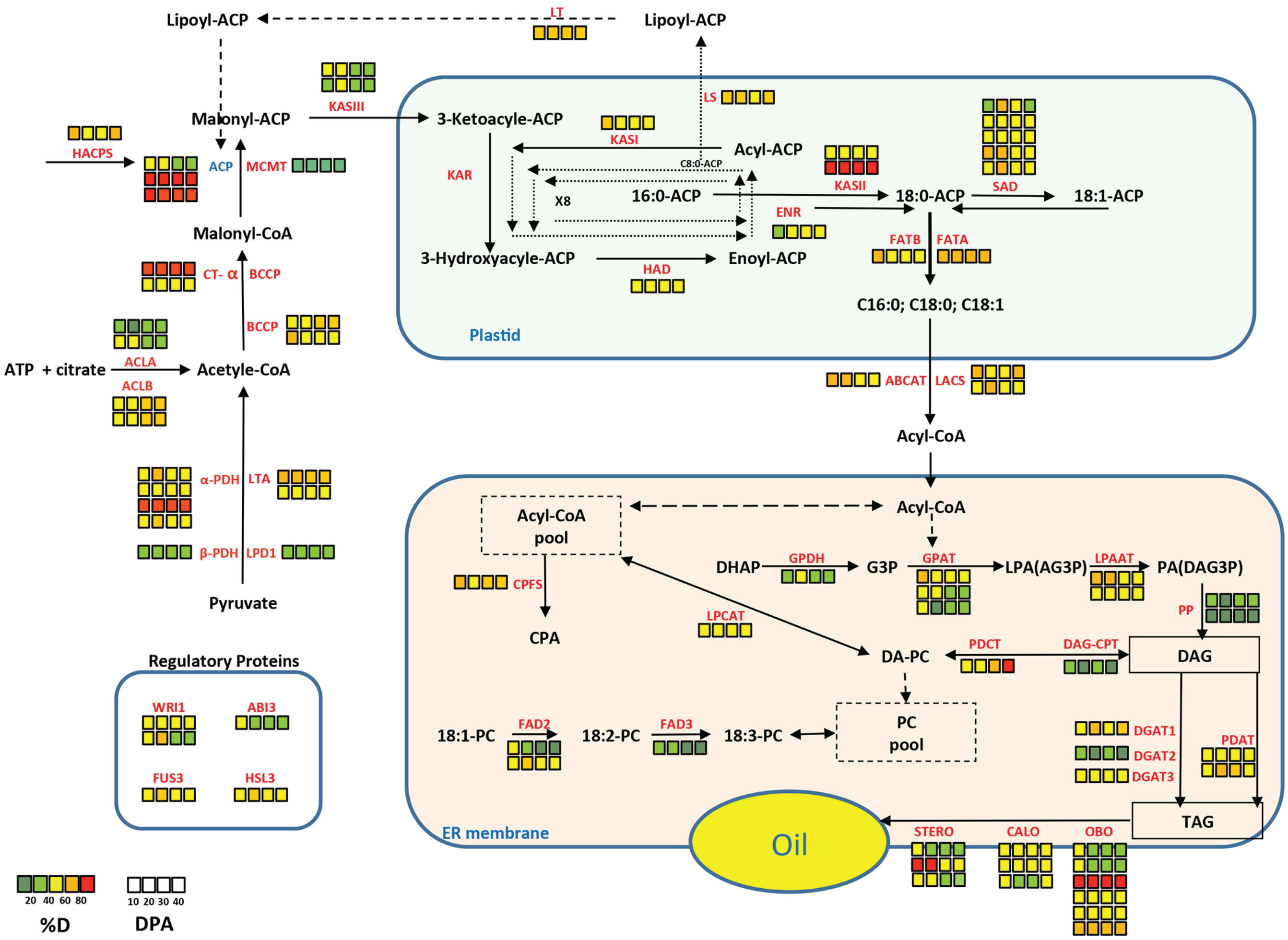

Figure 6. Partitioned expression of At and Dt subgenomes in fatty acid (FA) and oil biosynthesis related genes. The relative expression of each gene in the figure is represented by four squares that correspond to four developmental stages. Dark red and green colors indicate bias toward the Dt subgenome and At subgenomes, respectively. Full names of the genes are listed in Supplemental Table S3. ACP, acyl carrier protein, CPA, cyclopropane-fatty-acyl-phospholipid, D, D-genome; %D, percent bias towards D-genome; DAG, diacylglycerol, DPA, days postanthesis; PC, phosphatidylcholine, TAG, triacylglycerol.

resulted in 80 gene pairs for which reads were parsed by homeolog into their genome of origin. The relative bias of their expression toward the At or Dt subgenome is presented in Fig. 6. In general, both subgenomes contribute to gene expression along the oil pathway. Yet, some At- and Dt- biases were observed, in particular, parts of the pathway. For example, the de novo FA biosynthesis pathway in the cytosol and the plastid is generally more biased toward the Dt subgenome (red and orange squares in the figure). Alternatively, FA desaturation and TAG biosynthesis genes in the endoplasmic reticulum are more biased to the At subgenome (dark and light green squares in Fig. 6). Oil storage related genes (oleosin, caleosin, and steroleosin) that are highly expressed in the seeds are slightly biased toward the Dt subgenome, although both genomes are expressed in general (Fig. 6 and Supplemental Table S3).

Analysis of Flavonoid Biosynthesis Genes in Cottonseed

As noted above, the metabolic expression network and the highly expressed genes analyses indicate the possible importance of the flavonoid biosynthesis pathway in the cotton seed. Twenty-nine genes belonging to 13 key metabolic steps in the flavonoid biosynthesis pathway were chosen for global and homeolog-specific expression analyses (Supplemental Table S3). In general, genes are highly expressed early in seed development and their expression decreases after 20 DPA (Fig. 4). The only genes that showed higher expression during the late stages of seed development were flavonol synthase (FLS) and UDPG:flavonoid glucosyltransferase, which points to the accumulation of flavonols and anthocyanin glucosides later in seed development. FSII was not expressed at any point in development, suggesting that cotton seeds lack flavones.

Subgenomic expression bias was found to vary in its direction throughout the pathway, as summarized in Fig. 8 and Supplemental Table S3b. For chalcone synthase,

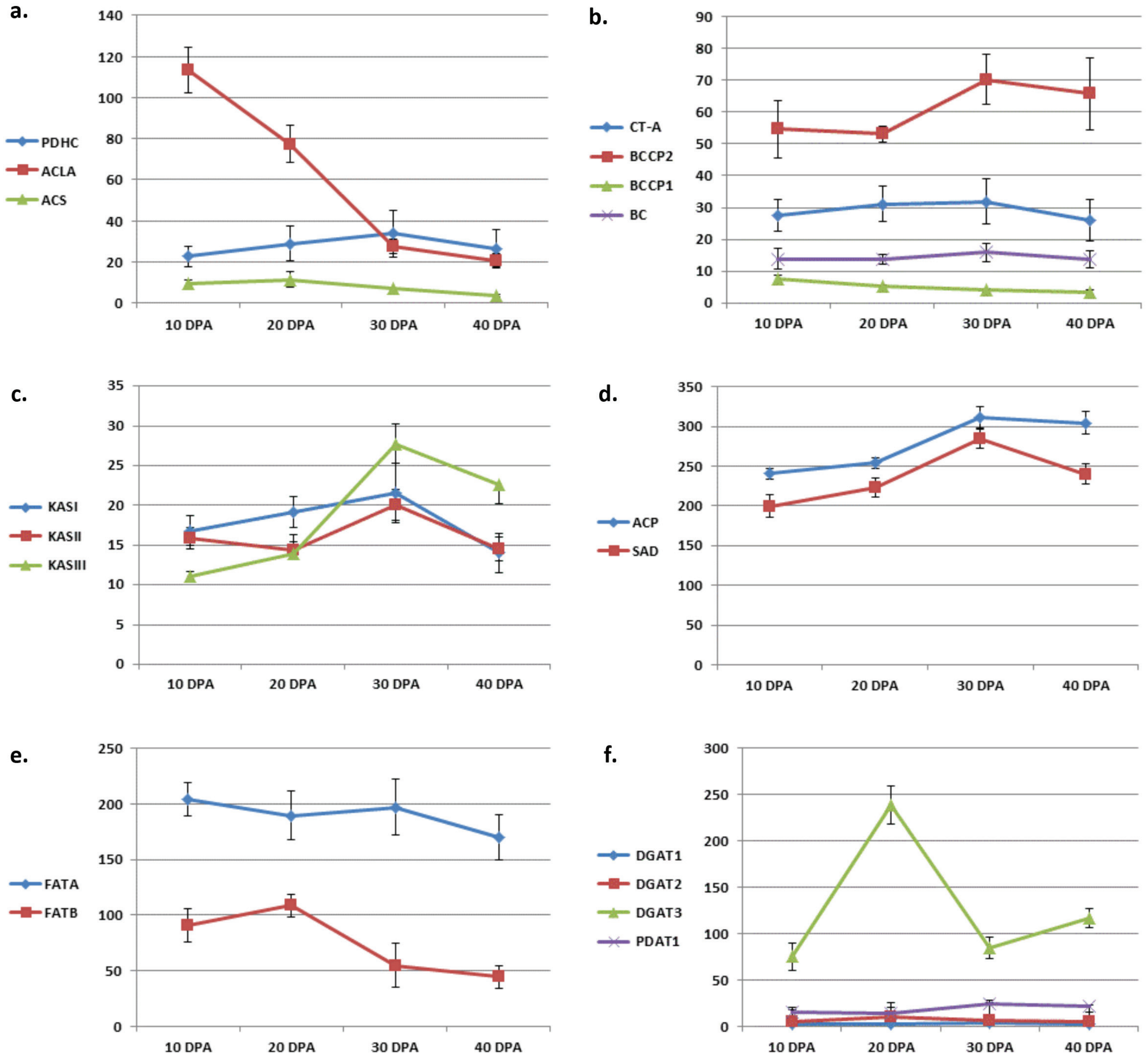

Figure 7. Nonpartitioned expression pattern (reads per kilobase per million mapped reads, RPKM) of several fatty acid (FA) and triglyceride (TAG) biosynthesis genes in developing cotton seeds. Values in each graph represent the mean expression of all isoforms for the indicated gene. (a) genes involved in aceyl-CoA biosynthesis. PDHC, pyruvate dehydrogenase complex; ACLA, ATP citrate lyase; ACS, acetyl-CoA synthase. (b) Heteromeric subunits of the acetyl-CoA carboxylase (ACCase) complex. CT-α-carboxyltransferase; BC, biotin carboxylase; BCCP, biotin carboxyl carrier protein. (c) The ketoacyl-ACP Synthase family. (d) Stearoyl-ACP Desaturase (SAD) and acyl carrier protein (ACP) gene families. (e) Acyl-ACP thioesterase (FATA, and FATB). (f) Acyl-CoA: diacylglycerol acyltransferase (DGAT) and Phospholipid:diacylglycerol acyltransferase (PDAT) gene families.

the entry point into the flavonoid biosynthesis pathway, the three gene pairs that contributed the vast majority of expression were biased in both directions at 10 DPA, but switch to being only Dt-biased at 20 DPA. However, *chalcone isomerase*, *flavonone 3-hydroxylase* (*F3H*), and *flavonol 3′,5′-hydroxylase* (*F3′5′H*), the next three genes pairs in the pathway, are At-biased. *Flavonol 3′-hydroxylase* (*F3′H*), which acts at the same step as *F3′5′H*, is Dt-biased, but is expressed at much lower levels. This DE suggests a higher ratio of the products of *F3H* and *F3′5′H* (dihydrokaempferol and dihydromyricetin) in relation to the products of *F3′H* (*dihydroquercetin*). *Dihydroflavonol 4-reductase* is strongly biased in both directions in the two gene pairs that produce its enzyme. *Leucoanthocyanidin dioxygenase* is Dt-biased in both of its gene pairs.

Flavonols and proanthocyanidins are the only flavonoids found in seeds of Arabidopsis (Lepiniec et al., 2006; Routaboul et al., 2006). Flavonols are initially produced by FLS, which is strongly biased toward the Dt homeolog from 30 to 40 DPA. Proanthocyanidins, also called condensed tannins, collect in the seed coat and protect the embryo and endosperm. *Leucoanthocyanidin*

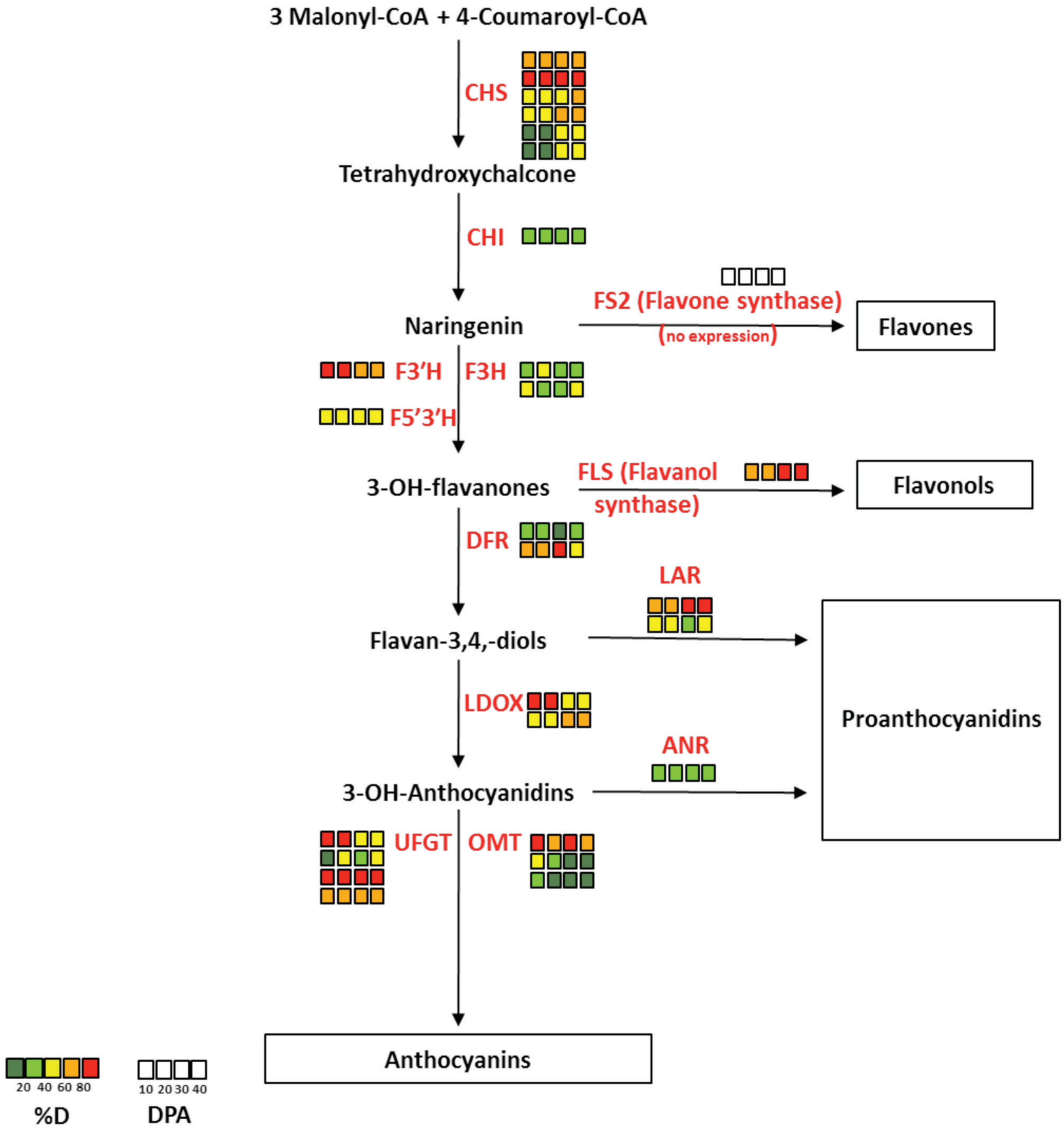

Figure 8. Partitioned expression of At and Dt subgenomes in flavonoid biosynthesis related genes. The relative expression of each gene in the figure is represented by four squares that correspond to four developmental stages. Dark red and green colors indicate bias toward the D subgenome and A subgenomes, respectively. Full names of the genes are listed in Supplemental Table S3.

reductase (*LAR*) and *anthocyanidin reductase* (*ANR*) produce proanthocyanidins from leucoanthocyanidins and anthocyanidins, respectively, but lead to differing stereochemistries (Lepiniec et al., 2006). In cotton seed, *LAR* is strongly biased toward the Dt homeolog in one gene pair, but is moderately biased toward the At homeolog in the other; *ANR* is strongly At-biased. Unlike in Arabidopsis, in which only ANR has been identified (Xie et al., 2003). This suggests the presence of both the *cis*-oriented epicatechin and the *trans*-oriented catechin in cotton seed. Taken together, these results lead to the hypothesis that the flavonol and catechin-based proanthocyanidin profiles will more closely mirror the D-genome cotton species, while the epicatechin-based proanthocyanidins profile will more closely mirror that of the A-genome species. Further metabolomic work will be necessary to determine the validity of this hypothesis.

Conclusions

While recent studies on the cotton transcriptome have shed some light on cotton seed and trichome biology, information on gene expression during different stages of cotton seed development has been lacking. In this study, by using gene annotation data from the recently sequenced D-genome cotton and an extensive RNA-seq dataset for four developmental stages, we provide baseline information and a resource that may be used to understand oil biosynthesis in cotton seeds. As a result, temporal patterns and expression levels are now available for thousands of genes, which can assist future protein and enzyme or other analyses, as well as facilitate cotton seed improvement efforts. In addition, we used genome-diagnostic SNP information based on data from each model diploid progenitor genome to enable characterization of expression for each homeologous gene copy of most important gene pairs. This has led to additional insights based on the individual subgenomic contributions for biological processes in developing seeds.

The study showed that the major time period for oil accumulation in cotton seeds under our experimental conditions occurs between 20 and 30 DPA. At this developmental stage, a significant portion of the seed RNA transcriptome is also highly dynamic. This may help inform future genetic and evolutionary studies (e.g., comparison between varieties, wild forms, mutant lines) that target gene expression differences concentrated on this particular developmental period. Gene ontology enrichment analysis of DE genes showed that biological processes such as FA biosynthesis, and regulation of meristem growth hormone-mediated signaling pathways are up-regulated during this developmental stage, whereas phenylpropanoid biosynthesis, regulation of cell size, and pigment accumulation processes are down-regulated (Supplemental Table S1). This change in global expression appears to be mediated by a small but significant bias toward the At subgenome. However, no particular biological processes were found to be overrepresented for this bias, and the overall gene expression network was similar for both subgenomes. This indicates that, for the majority of the transcriptome, an orchestrated and combined expression from both subgenomes is utilized in this allopolyploid crop plant. However, in some specific cases, such as for storage proteins, a strong and possibly biologically important bias occurs toward one genome.

Finally, we show a somewhat novel temporal expression pattern of lipid metabolism in cotton seeds. We report new and even surprising aspects like the involvement of ATP citrate lyase in acetyl-CoA synthesis and the contribution of *DGAT3* to the total TAG biosynthesis gene expression pool. These and other findings suggest aspects of oil seed development that may be unique to cotton. This aspect is now being studied using a broader set of cotton species and to detect the evolution of these unique patterns of FA and oil biosynthesis within the cotton tribe.

Acknowledgments

This research was supported by USA-Israel Bilateral Science Foundation (BSF) Grant No. 2009447— "The evolution of duplicate oil pathways in polyploid cotton." Joseph P. Gallagher is funded by a Graduate Research Fellowship from the National Science Foundation.

References

Baud, S., and L. Lepiniec. 2009. Regulation of de-novo fatty acid synthesis in maturing oilseeds of Arabidopsis. Plant Physiol. Biochem. 47:448–455. doi:10.1016/j.plaphy.2008.12.006

Baud, S., and L. Lepiniec. 2010. Physiological and developmental regulation of seed oil production. Prog. Lipid Res. 49:235–249. doi:10.1016/j.plipres.2010.01.001

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57:289–300.

Bewley, J.D. 2006. Cotton-an oilseed. In: J.D., Bewley, M. Black, and P. Halmer, editors, The encyclopedia of seeds: Science, technology and uses. CAB international, London. p. 106–109.

Brand, Y., and R. Hovav. 2010. Identification of suitable internal control genes for quantitative Real-Time PCR expression analyses in peanut (Arachis hypogaea). Peanut Sci. 37:12–19. doi:10.3146/PS09-014.1

Cherry, J.P. 1983. Cottonseed oil. J. Am. Oil Chem. Soc. 60:360–367. doi:10.1007/BF02543519

Dahlqvist, A., U. Stahl, M. Lenman, A. Banas, M. Lee, L. Sandager, H. Ronne, and H. Stymne. 2000. Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc. Natl. Acad. Sci. USA 97:6487–6492. doi:10.1073/pnas.120067297

Du, Z., X. Zhou, Y. Ling, Z. Zhang, and Z. Su. 2010. agriGO: A GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38:W64–W70. doi:10.1093/nar/gkq310

Dure, L. 1975. Seed formation. Annu. Rev. Plant Physiol. 26:259–278. doi:10.1146/annurev.pp.26.060175.001355

Dussert, S., C. Guerin, M. Andersson, T. Joet, T.J. Tranbarger, M. Pizot, G. Sarah, A. Omore, T. Durand-Gasselin, and F. Morcillo. 2013. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol. 162:1337–1358. doi:10.1104/pp.113.220525

Flagel, L.E., J.F. Wendel, and J.A. Udall. 2013. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genomics 13:302. doi:10.1186/1471-2164-13-302

Fryxell, P.A. 1979. The natural history of the cotton tribe. Texas A&M Univ. Press, College Station, TX.

Gotmare, V., P. Singh, C. Mayee, V. Deshpande, and C. Bhagat. 2004. Genetic variability for seed oil content and seed index in some wild species and perennial races of cotton. Plant Breed. 123:207–208. doi:10.1046/j.1439-0523.2003.00914.x

Grover, C.E., J.P. Gallagher, E.P. Szadkowski, M.J. Yoo, L.E. Flagel, and J.F. Wendel. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. 196:966–971. doi:10.1111/j.1469-8137.2012.04365.x

Hernandez, M.L., L. Whitehead, Z.S. He, V. Gazda, A. Gilday, E. Kozhevnikova, F.E. Vaistij, T.R. Larson, and I.A. Graham. 2012. A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued Arabidopsis seed oil catabolism mutants. Plant Physiol. 160:215–225. doi:10.1104/pp.112.201541

Hovav, R., J.A. Udall, B. Chaudhary, R. Rapp, L. Flagel, and J.F. Wendel. 2008a. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc. Natl. Acad. Sci. USA 105:6191–6195. doi:10.1073/pnas.0711569105

Hovav, R., J.A. Udall, E. Hovav, R. Rapp, L. Flagel, and J.F. Wendel. 2008b. A majority of cotton genes are expressed in single-celled fiber. Planta 227:319–329. doi:10.1007/s00425-007-0619-7

Hu, G., N.L. Houston, D. Pathak, L. Schmidt, J.J. Thelen, and J.F. Wendel. 2011. Genomically biased accumulation of seed storage proteins in allopolyploid cotton. Genetics 189:1103–1115. doi:10.1534/genetics.111.132407

Hu, G., J. Koh, M.J. Yoo, K. Grupp, S. Chen, and J.F. Wendel. 2013. Proteomic profiling of developing cotton fibers from wild and domesticated Gossypium barbandense. New Phytol. 200:570–582. doi:10.1111/nph.12381

Ichihara, K., T. Takahashi, and S. Fujii. 1988. Diacylglycerol acyltransferase in maturing safflower seeds—its influences on the fatty-acid composition of triacylglycerol and on the rate of triacylglycerol synthesis. Biochim. Biophys. Acta 958:125–129. doi:10.1016/0005-2760(88)90253-6

Jiao, X., X. Zhao, X.-R. Zhou, A.G. Green, Y. Fan, L. Wang, S.P. Singh, and Q. Liu. 2013. Comparative transcriptomic analysis of developing cotton cotyledons and embryo axis. PLoS ONE 8:e71756. doi:10.1371/journal.pone.0071756

Jones, A., H.M. Davies, and T.A. Voelker. 1995. Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell 7:359–371. doi:10.1105/tpc.7.3.359

Jones, S.I., and L.O. Vodkin. 2013. Using RNA-Seq to profile soybean seed development from fertilization to maturity. PLoS ONE 8:e59270. doi:10.1371/journal.pone.0059270

Lepiniec, L., I. Debeaujon, J.M. Routaboul, A. Baudry, L. Pourcel, N. Nesi, and M. Caboche. 2006. Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. 57:405–430. doi:10.1146/annurev.arplant.57.032905.105252

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Lin, M., and D.J. Oliver. 2008. The role of acetyl-coenzyme a synthetase in Arabidopsis. Plant Physiol. 147:1822–1829. doi:10.1104/pp.108.121269

Liu, Q., R.M.P. Siloto, R. Lehner, S.J. Stone, and R.J. Weselake. 2012. Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology. Prog. Lipid Res. 51:350–377. doi:10.1016/j.plipres.2012.06.001

Liu, Q., S. Singh, K. Chapman, and A. Green. 2009. Bridging traditional and molecular genetics in modifying cottonseed oil. In: A.H., Paterson, editor, Genomics of cotton, plant genetics and genomics; crops and models 3. Springer, New York. p. 353–384.

Liu, Q., S. Singh, and A. Green. 2002. High-oleic and high-stearic cottonseed oils: Nutritionally improved cooking oils developed using gene silencing. J. Am. Coll. Nutr. 21:205–211. doi:10.1080/07315724.2002.10719267

Lung, S.C., and R.J. Weselake. 2006. Diacylglycerol acyltransferase: A key mediator of plant triacylglycerol synthesis. Lipids 41:1073–1088. doi:10.1007/s11745-006-5057-y

Mauney, J.R. 1984. Cotton. Agron. Monogr. 24. ASA, CSSA, SSSA, Madison, WI.

Mhaske, V., K. Beldjilali, J. Ohlrogge, and M. Pollard. 2005. Isolation and characterization of an Arabidopsis thaliana knockout line for phospholipid: Diacylglycerol transacylase gene (At5g13640). Plant Physiol. Biochem. 43:413–417. doi:10.1016/j.plaphy.2005.01.013

Miller, N.A., S.F. Kingsmore, A. Farmer, R.J. Langley, J. Mudge, J.A. Crow, A.J. Gonzalez, F.D. Schilkey, R.J. Kim, J. van Velkinburgh, G.D. May, C.F. Black, M.K. Myers, J.P. Utsey, N.S. Frost, D.J. Sugarbaker, R. Bueno, S.R. Gullans, S.M. Baxter, S.W. Day, and E.F. Retzel. 2008. Management of high-throughput DNA sequencing projects. Alpheus. J. Comput. Sci. Syst. Biol. 1:132. doi:10.4172/jcsb.1000013

Nagalakshmi, U., K. Waern, and M. Snyder. 2010. RNA-Seq: A method for comprehensive transcriptome analysis. Curr. Protoc. Mol. Biol. 11:1–13. doi:10.1002/0471142727.mb0411s89.

O'Brien, R.D. 2002. Cottonseed oil. In: F.D. Gunstone, editor, Vegetable oils in food technology: Composition, properties and uses. Blackwell Publishing, Oxford. p. 203–230.

O'Brien, R.D., L.A. Jones, C.C. King, P.J. Wakelyn, and P.J. Wan. 2005. Cottonseed Oil. In: F. Shahidi, editor, Bailey's industrial oil and fat products. John Wiley & Sons, Hoboken.

Page, J.T., A.R. Gingle, and J.A. Udall. 2013a. PolyCat: A resource for genome categorization of sequencing reads from allopolyploid organisms. G3: Genes Genomes Genet. 3:517–525. doi:10.1534/g3.112.005298.

Page, K., S. Brin, R. Motwani, and T. Winograd. 1998. PageRank citation ranking: Bringing order to the web. Technical Report, Stanford Digital Library Technologies Project, Stanford, CA.

Page, J.T., M.D. Huynh, Z.S. Liechty, K. Grupp, D. Stelly, A. Hulse, H. Ashrafi, A. van Deynze, J.F. Wendel, and J.A. Udall. 2013b. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. G3: Genes Genomes Genet. 3:1809–1818. doi:10.1534/g3.113.007229.

Paterson, A.H., J.F. Wendel, H. Gundlach, H. Guo, J. Jenkins, et al. 2012. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492:423–427. doi:10.1038/nature11798

Rambani, A., J.T. Page, and J.A. Udall. 2014. Polyploidy and the petal transcriptome of Gossypium. BMC Plant Biol. 14:3. doi:10.1186/1471-2229-14-3

Ratledge, C., M.D. Bowater, and P.N. Taylor. 1997. Correlation of ATP/citrate lyase activity with lipid accumulation in developing seeds of Brassica napus L. Lipids 32:7–12. doi:10.1007/s11745-997-0002-7

Reeves, R.G., and J.O. Beasley. 1935. The development of the cotton embryo. J. Agric. Res. 51:935–944.

Routaboul, J.M., L. Kerhoas, I. Debeaujon, L. Pourcel, M. Caboche, J. Einhorn, and L. Lepiniec. 2006. Flavonoid diversity and biosynthesis in seed of Arabidopsis thaliana. Planta 224:96–107. doi:10.1007/s00425-005-0197-5

Ruuska, S.A., J. Schwender, and J.B. Ohlrogge. 2004. The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. Plant Physiol. 136:2700–2709. doi:10.1104/pp.104.047977

Saha, S., B. Enugutti, S. Rajakumari, and R. Rajasekharan. 2006. Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. Plant Physiol. 141:1533–1543. doi:10.1104/pp.106.082198

Sekhon, R.S., R. Briskine, C.N. Hirsch, C.L. Myers, N.M. Springer, C.R. Buell, N. de Leon, and S.M. Kaeppler. 2013. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE 8:e61005. doi:10.1371/journal.pone.0061005

Severin, A.J., J.L. Woody, Y.T. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Specht, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, and R.C. Shoemaker. 2010. RNA-Seq atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol. 10:160. doi:10.1186/1471-2229-10-160

Shenstone, F.S., and J.R. Vickery. 1961. Occurrence of cylco-propene acids in some plants of the order malvales. Nature 190:168–169. doi:10.1038/190168b0

Thelen, J.J., S. Mekhedov, and J.B. Ohlrogge. 2001. Brassicaceae express multiple isoforms of biotin carboxyl carrier protein in a tissue-specific manner. Plant Physiol. 125:2016–2028. doi:10.1104/pp.125.4.2016

Troncoso-Ponce, M.A., A. Kilaru, X. Cao, T.P. Durrett, J. Fan, J.K. Jensen, N.A. Thrower, M. Pauly, C. Wilkerson, and J.B. Ohlrogge. 2011. Comparative deep transcriptional profiling of four developing oilseeds. Plant J. 68:1014–1027. doi:10.1111/j.1365-313X.2011.04751.x

Voelker, T. 1996. Plant acyl-ACP thioesterases: Chain-length determining enzymes in plant fatty acid biosynthesis. Genet. Eng. 18:111–133. doi:10.1007/978-1-4899-1766-9_8

Wendel, J.F., L.E. Flagel, and K.L. Adams. 2012. Jeans, genes, and genomes: Cotton as a model for studying polyploidy. In: P.S. Soltis, and D.E. Soltis, editors, Polyploidy and genome evolution. Springer, New York. p. 181–207.

Xie, D.Y., S.B. Sharma, N.L. Paiva, D. Ferreira, and R.A. Dixon. 2003. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science 299:396–399. doi:10.1126/science.1078540

Yoo, M.J., E. Szadkowski, and J.F. Wendel. 2013. Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity (Edinb) 110:171–180. doi:10.1038/hdy.2012.94

Yoo, M.J., and J.F. Wendel. 2014. Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. PLoS Genet. 10:e1004073. doi:10.1371/journal.pgen.1004073

Yu, X.H., R. Rawat, and J. Shanklin. 2011. Characterization and analysis of the cotton cyclopropane fatty acid synthase family and their contribution to cyclopropane fatty acid synthesis. BMC Plant Biol. 11:97. doi:10.1186/1471-2229-11-97

Zhang, M., J.L. Fan, D.C. Taylor, and J.B. Ohlrogge. 2009. DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell 21:3885–3901. doi:10.1105/tpc.109.071795