

1 Conservation and divergence in duplicated fiber coexpression networks accompanying
2 domestication of the polyploid *Gossypium hirsutum* L.

3
4 Joseph P. Gallagher^{*,1}, Corrinne E. Grover^{*}, Guanjing Hu^{*}, Josef J. Jareczek^{*}, and Jonathan F.
5 Wendel^{*}

6 ^{*} Ecology, Evolution, and Organismal Biology Dept., Iowa State University, Ames, IA, 50010

7 ¹ current address; Biology Dept., University of Massachusetts, Amherst, MA, 01003

8

9 ORCID (email):

10 MPG: 0000-0003-0378-7112 (jpg@umass.edu)

11 CEG: 0000-0003-3878-5459 (corrinne@iastate.edu)

12 GH: 0000-0001-8552-7394 (hugj2006@iastate.edu)

13 JJJ: 0000-0001-5609-6281 (jareczek@iastate.edu)

14 JFW 0000-0003-2258-5081 (jfw@iastate.edu)

15

16 Data reference: PRJNA530448, <http://dx.doi.org/10.5061/dryad.256hn>

17

18

19

20 **Running Title:** cotton fiber coexpression networks

21 **Keywords:** cotton, domestication, gene coexpression, differential gene expression, fiber,
22 polyploidy

23 **Corresponding Author:**

24 Jonathan F. Wendel

25 Department of Ecology, Evolution, and Organismal Biology

26 345A Bessey Hall, 2200 Osborn Drive

27 Ames IA 50011-4009

28 jfw@iastate.edu

29 515-294-7172

30

Abstract

31 *Gossypium hirsutum* L. (Upland cotton) has an evolutionary history involving inter-genomic
32 hybridization, polyploidization, and subsequent domestication. We analyzed the developmental
33 dynamics of the cotton fiber transcriptome accompanying domestication using gene coexpression
34 networks for both homoeolog-pair and homoeolog-specific networks. Remarkably, most genes
35 exhibited expression for at least one homoeolog, confirming previous reports of widespread gene
36 usage in cotton fibers. Most coexpression modules comprising the joint network are preserved in
37 each subgenomic network and are enriched for similar biological processes, showing a general
38 preservation of network modular structure for the two co-resident genomes in the polyploid.
39 Interestingly, only one fifth of homoeologs co-occur in the same module when separated, despite
40 similar modular structures between the homoeolog-specific networks. These results suggest that
41 the genome-wide divergence between homoeologous genes is sufficient to separate their co-
42 expression profiles at the intermodular level, despite conservation of intramodular relationships
43 within each subgenome. Most modules exhibit D-homoeolog expression bias, although specific
44 modules do exhibit A-homoeolog bias. Comparisons between wild and domesticated
45 coexpression networks revealed a much tighter and denser network structure in domesticated
46 fiber, as evidenced by its fewer modules, 13-fold increase in the number of development-related
47 module member genes, and the poor preservation of the wild network topology. These results
48 demonstrate the amazing complexity that underlies the domestication of cotton fiber.

49

Introduction

50 Cotton (*Gossypium*) is the most important source of natural textile fibers globally. Among the
51 four cultivated species, *G. hirsutum* L., also known as Upland cotton, is the most widely grown,
52 and is responsible for more than 90% of cotton production worldwide. Wild *G. hirsutum*, native
53 to coastal Yucatan, Mexico and more sparsely elsewhere in nearby regions (extending as far
54 north as the Florida Keys), was domesticated approximately 5,000 years ago (Wendel *et al.*
55 1992; Brubaker and Wendel 1994; d'Eeckenbrugge and Lacape 2014). Following initial
56 domestication in or around the Yucatan Peninsula, *G. hirsutum* spread rapidly throughout Central
57 America, where semi-domesticated or “door-yard” forms are still found today. In the last several
58 hundred years, strong directional selection for enhanced fiber and other agronomic traits led to
59 the modern forms of Upland cotton, which are grown globally today (Fig. 1).

60

61 Cotton fibers are single-celled seed trichomes that differentiate from the ovular epidermis as
62 early as three days before anthesis (Haigler *et al.* 2012). Fiber cells first elongate rapidly through
63 the synthesis of a thin primary cell wall, which is followed by secondary cell wall thickening.
64 During this transitional stage, around 15 to 20 days post-anthesis (dpa), an intermediary cell wall
65 “winding” layer is deposited, which substantially increases fiber strength. This secondary cell
66 wall, composed of 98-99% cellulose in domesticated cotton, continues to thicken until ~40 dpa
67 (Applequist *et al.* 2001; Haigler *et al.* 2012). During the last stage of fiber development, mature
68 fibers dehydrate, form spiral twists, and the capsules (colloquially “bolls”) dehisce. This
69 developmental programme requires the regulation and coordination of hundreds to thousands of
70 genes involved in cell wall and cytoskeletal formation (Szymanski and Cosgrove 2009; Haigler
71 *et al.* 2012; Yoo and Wendel 2014; Fang 2018).

72

73 Although cotton fibers are single-cells primarily composed of cellulose, a high proportion of the
74 ~72,000 genes in the genome are expressed at some point during fiber development (Hovav *et al.*
75 2008c; Yoo and Wendel 2014; Tuttle *et al.* 2015). Previous microarray and RNA-seq studies
76 have shown that the fiber transcriptomes vary dramatically between wild and domesticated *G.*
77 *hirsutum* (Rapp *et al.* 2010; Yoo and Wendel 2014; Bao *et al.* 2019). Bao *et al.* (2019) showed
78 that 15% of the genes have experienced some form of regulatory alteration between a pair of
79 wild and domesticated *G. hirsutum* accessions (TX2094 and Maxxa, respectively) in 10 and 20
80 dpa fibers, developmental time points that represent primary wall synthesis and the transition to
81 secondary wall synthesis, respectively. They and others (e.g. Applequist *et al.* 2001; Yoo and
82 Wendel 2014) suggested that the transcriptomes of modern elite lines have been reprogrammed
83 so that the duration of fiber elongation is lengthened and that resources from stress response
84 pathways have been reallocated towards enhanced fiber growth. While several classes of
85 domestication-related genes have been identified [e.g., master transcription factors (Shan *et al.*
86 2014); cellulose synthase (Rapp *et al.* 2010; Yoo and Wendel 2014), profilin (Bao *et al.* 2011),
87 reactive oxygen species regulation-related enzymes (ROS) (Hovav *et al.* 2008b)], the underlying
88 mechanisms and key components targeted by human-mediated selection remain elusive.

89

90 In addition to domestication, polyploidy itself also has had pervasive effects on gene expression
91 underlying fiber development. *G. hirsutum* and six other tetraploid species (Krapovickas and
92 Seijo 2008; Grover *et al.* 2015b, 2015a; Gallagher *et al.* 2017) represent the descendants of a
93 single allopolyploidization event *ca.* 1-2 million years ago between an African, A-genome
94 diploid species and an American, D-genome diploid (Wendel and Grover 2015). This

95 hybridization and subsequent genome doubling reunited genomes that evolved independently for
96 5-10 million years on separate continents and created genome-wide gene duplicates
97 (homoeologs). While these homoeologs exist in a shared *trans* regulatory environment, they may
98 have divergent *cis* regulatory controls (Wendel and Grover 2015; Bao *et al.* 2019). The
99 interactions among these newly reunited genes may contribute, in part, to the phenomenon
100 known as “genomic shock”, whose myriad possible effects are summarized elsewhere (Jackson
101 and Chen 2010; Grover *et al.* 2012; Yoo *et al.* 2014; Hu and Wendel 2019).

102
103 Adding to the evolutionary interest surrounding cotton fibers, only one parent of the polyploid
104 species, the maternal A-genome progenitor, possesses long spinnable fiber. Accordingly, one
105 might expect that the A-derived subgenome of the domesticated polyploids would carry most
106 targets for selection of fiber-related traits. Contrary to this expectation, however, there is
107 abundant evidence of D-genome recruitment into the developmental program of domesticated
108 allopolyploid cotton fiber. Numerous QTL studies have found fiber-related loci in the D-
109 subgenome of polyploidy cotton (Jiang *et al.* 1998; Lacape *et al.* 2005, 2010; Ulloa *et al.* 2005;
110 Han *et al.* 2006; Rong *et al.* 2007; Qin *et al.* 2008; Said *et al.* 2013, 2015; Grover *et al.* 2019).
111 Initial assessments of gene expression mirrored these observations, with 20% of homoeolog pairs
112 exhibiting biased homoeolog expression in cotton fiber favoring homoeologs from the non-fiber
113 (D genome) producing parent (Hovav *et al.* 2008b). Subsequent research has found a general
114 balance between A- and D-subgenome expression but with considerable variation among genes
115 with respect to the direction (A or D) of homoeolog bias (Yoo and Wendel 2014; Zhang *et al.*
116 2015; Fang *et al.* 2017b). The importance and contribution of the D-subgenome to allopolyploid
117 cotton has recently been reiterated using whole genome resequencing screens for signatures of

118 selection (Fang *et al.* 2017a, 2017b, 2017c; Wang *et al.* 2017). Collectively, the foregoing
119 studies support the notion that polyploidy-related expression alterations underlie the
120 transgressive and enhanced properties of fiber from allopolyploid relative to diploid cotton.

121

122 Whereas differential gene expression (DGE) is commonly used to evaluate transcriptomic
123 changes among species and genotypes, coexpression analyses, including differential correlation
124 (DC) and network construction, uncover how expression among genes is coordinated and how
125 these coordinated relationships are evolutionarily altered by either natural or human-mediated
126 selection. As demonstrated in tomato, wheat, maize, and cotton (Swanson-Wagner *et al.* 2012;
127 Ichihashi *et al.* 2014; Pfeifer *et al.* 2014; Hu *et al.* 2016), coexpression networks provide a
128 framework for testing preservation of coexpression patterns between wild and domesticated
129 accessions or between diploid and polyploid species (Gallagher *et al.* 2016), while also
130 highlighting coordinated changes among closely connected genes. Here we characterize fiber
131 gene expression in wild and domesticated *G. hirsutum* across four time points representing key
132 stages in fiber development. We examine the effects of polyploidy and domestication on gene
133 coexpression, confirming previous reports of dramatic alterations in gene expression dynamics
134 under domestication. While comparisons between homoeologous coexpression networks (i.e., A-
135 and D- subgenome networks) show general conservation in homoeologous network structure,
136 comparisons between wild and domesticated networks show a distinct lack of conservation and a
137 high level of differentially correlated genes. These results suggest a general preservation of
138 duplicate gene function in polyploid *Gossypium* and highlight tighter co-regulation of fiber
139 development genes following domestication, as observed for domesticated cottonseed (Hu *et al.*
140 2016).

141
142 Materials and Methods
143 **Plant materials, mRNA sequencing and mapping**
144 We selected three wild and three domesticated accessions of *G. hirsutum* to represent the wild to
145 domestication transformation (Fig. 1). These accessions have previously been shown to span the
146 known genetic diversity of wild and domesticated cotton (Brubaker and Wendel 1994; Wendel et
147 al. 1992; Grover et al. 2017; Yoo and Wendel 2014). Plants were grown in a common
148 greenhouse environment in the Pohl Conservatory at Iowa State University. Cotton ovules were
149 collected at 5, 10, 15, and 20 days post anthesis (dpa) and immediately frozen in liquid nitrogen
150 for fiber RNA extraction. Briefly, frozen ovules were vortexed with glass beads to shear the fiber
151 cells, and total RNA was extracted using the Sigma-Aldrich Spectrum Plant Total RNA kit
152 (Sigma-Aldrich, St. Louis, MO) followed by purification as in Hovav et al (2008). Illumina
153 RNA-seq libraries were constructed and subsequently sequenced as single-end 100-nt reads on
154 the Illumina HiSeq2500 at the DNA Facility of Iowa State University
155 (<http://dna.biotech.iastate.edu/>). A total of twenty-four RNA-seq libraries were generated with an
156 average of 30 million reads per library.
157
158 RNA-seq reads were mapped onto the diploid reference genome sequence of D-genome diploid
159 *G. raimondii* (Paterson et al. 2012) using the single nucleotide polymorphism (SNP)-tolerant
160 mapping option of GSNAp (Wu and Nacu 2010) and a homoeolog-specific SNP database that
161 distinguishes the A- and D- subgenomes in allopolyploid cotton (Page et al. 2014). Mapped
162 reads containing the diagnostic SNPs were partitioned to estimate A- and D-subgenome
163 homoeolog-specific expression by PolyCat v. 1.3 (Page et al. 2013, 2014). Read count data were
164 generated via samtools (Li et al. 2009) and HTseq-count (Anders et al. 2015) for (1) the

165 combined expression of both A- and D- homoeologous genes, and (2) the expression of
166 individual homoeologs based only on subgenome specific reads. These two datasets are hereafter
167 referred to as “joint” and “homoeologous”, respectively.

168

169 **Differential gene expression analysis**

170 Using the R package DEseq2 (Love *et al.* 2014), a multifactor design of \sim *domestication* +
171 *development* + *domestication:development* was applied to both the joint and homoeologous
172 datasets, together with pairwise comparisons of appropriate contrasts with respect to
173 developmental stage and domestication status. Read counts greater than an average of 1 across
174 all raw samples were considered expressed. The significant effect of *domestication:development*
175 interaction term was determined by contrasting the full model against a reduced model of \sim
176 *domestication* + *development* with the likelihood ratio test (LRT) built in DESeq2. Significant
177 statistical results were considered at a Benjamini-Hochberg adjusted *P*-value < 0.05 (Benjamini
178 and Hochberg 1995). Functional enrichment analysis was performed using the topGO package in
179 R 3.3.1 (Alexa and Rahnenfuhrer 2016; R Core Team 2017) with Fisher’s exact test. Gene
180 ontology (GO) annotations for the *G. raimondii* reference genes was downloaded from
181 CottonGen (Paterson *et al.* 2012; Yu *et al.* 2014).

182

183 Homoeolog expression bias was calculated in a similar fashion. Using the R package DEseq2
184 (Love *et al.* 2014), the multifactor design was updated to include “subgenome” as a factor (i.e., \sim
185 subgenome + domestication + development). The data set was limited to gene pairs that showed
186 an average expression of 1 read in at least half of the partitioned libraries. Differential expression
187 between homoeologous copies was calculated by contrasting the A and D subgenome.

188 Statistically significant homoeolog bias (within module or in total) was considered at a chi-
189 square P-value <0.05.

190

191 **Weighted gene coexpression network construction**

192 Gene coexpression networks were constructed using the R package WGCNA (Langfelder and
193 Horvath 2008). After removing genes with zero expression or without variation across samples, a
194 total of 29,706 homoeologous gene pairs were used to construct the joint networks, and 50,996
195 individual homoeologs (25,474 At and 25,522 Dt) were used to constructed the homoeologous
196 networks. Read counts across different RNA-seq libraries were first normalized using the *rlog*
197 function of DESeq2 (Love *et al.* 2014), and then subjected to automatic network construction
198 using the WGCNA function *blockwiseModules* with default settings. Briefly, Pearson
199 correlations were calculated between each pair of genes, and the resulting correlation matrix was
200 raised to a default power of $\alpha = 12$ to generate an adjacency matrix representing the connection
201 strengths among genes. Adequate fit to the scale-free topology of the biological network was
202 verified for each adjacency matrix (with the fit index above 0.8, or the highest fit index
203 achieved). Next, the topological overlap matrix (TOM) was calculated to measure network
204 interconnectedness for each pair of genes relative to all the other genes within the network. By
205 performing average linkage hierarchical clustering with a dynamic tree cutting algorithm on $1 -$
206 TOM (the measure of topological overlap dissimilarity), highly interconnected genes were
207 grouped together into coexpression modules, representing subnetwork structure and
208 organization.

209

210 The module eigengene (ME), whose expression represents the members of a given coexpression
211 module, was calculated as the first principal component of the scaled expression profiles of all
212 module gene members. As previously done by Hu *et al.* (2016), module expression levels of
213 member genes were summarized by their module eigengene value in correlation with the sample
214 conditions (here, 2 cultivation conditions x 4 developmental stages = 8 conditions). Module
215 eigengene-based connectivity (kME), also known as module membership, was calculated for
216 each gene by the Pearson correlation between the gene expression profile and its corresponding
217 ME. To determine whether a set of genes (e.g. a list of differentially expressed genes or genes
218 belonging to a Gene Ontology functional category) was significantly enriched within a specific
219 module, ranked lists of module kME were subjected to gene set enrichment analysis (GSEA)
220 using the *Preranked* function (Subramanian *et al.* 2005). That is, for a given gene set (i.e. genes
221 corresponding to a certain GO term or a differential expression category), the distribution of its
222 gene members were examined in a full network gene list ranked by kME, thereby calculating an
223 enrichment score to reflect the degree to which this gene set is overrepresented at the top of the
224 entire ranked list. Compared to overlap based enrichment tests such as Fisher's exact test, this
225 approach is more robust to the threshold parameters used for defining WGCNA modules. The
226 results of GSEA Preranked GO enrichment were visually summarized using REVIGO with an
227 allowed GO term similarity of 0.7 and the *Arabidopsis thaliana* GO term size database (Supek *et*
228 *al.* 2011).

229

230 **Conservation and divergence of gene coexpression networks**

231 Network preservation tests were performed as previously described (Hu *et al.* 2016). To assess
232 how well the intra-modular structure of a reference network is preserved in a test network, the

233 WGCNA function *modulePreservation* calculates two types of preservation statistics, i.e.,
234 Z_{summary} and medianRank scores, for each module (Langfelder and Horvath 2008). The
235 medianRank is a composite module preservation statistic calculated to compare relative
236 preservation among modules, such that lower medianRank score of a module indicates higher
237 preservation relative to other modules. Z_{summary} is derived from the Z statistic; modules with
238 $Z_{\text{summary}} > 10$ are interpreted as strongly preserved, whereas Z_{summary} between 2 and 10 indicates
239 weak to moderate preservation, and $Z_{\text{summary}} < 2$ indicates no preservation (Horvath 2011). Given
240 that the Z_{summary} statistic is sensitive to module size, while medianRank is not, both statistics were
241 considered collectively for inference of network modular structure preservation. This test was
242 performed for (1) each homoeologous network versus the joint gene expression modules and (2)
243 for the domesticated homoeologous network versus the homoeologous network from the wild
244 accessions.

245

246 **Differential coexpression analysis**

247 Differential coexpression (DC) analysis was performed by calculating the Pearson correlation
248 coefficients for all gene pairs, followed by comparisons of corresponding gene pairs between
249 wild and domesticated datasets, as previously conducted for cottonseed (Hu et al., 2016).
250 Differential correlation was evaluated based on Fisher's z-test using the R package DiffCorr with
251 a local FDR < 0.05 (Fukushima 2013). Differentially coexpressed genes were identified if the
252 number of DC gene pairs was significantly higher than expected. That is, for a gene identified
253 with k DC pairs among all gene pairs n , the probability P of this gene to be significantly
254 coexpressed follows the binomial distribution model. The resulting P values were further

255 corrected by the Benjamini-Hochberg method (Benjamini and Hochberg 1995) for multiple
256 testing to detect significant DC genes (adjusted $P < 0.05$).

257

258 **Data availability**

259 The raw sequencing data are available in the NCBI short read archive (SRA) under
260 PRJNA530448 and on Dryad (DOI: <http://dx.doi.org/10.5061/dryad.256hn>). Supplemental
261 material in support of this work can be found on figshare and includes: Fig S1, RNA-seq library
262 PCA by timepoint; Fig. S2, REVIGO plots summarizing homoeologous gene coexpression
263 network module GO term enrichment; Fig. S3, patterns of module eigengene expression for each
264 homoeologous gene coexpression network module; Fig. S4, gene membership correspondence
265 between the wild and domesticated gene coexpression networks; Table S1, RNA-seq library read
266 mapping and counts summary; Table S2, GO enrichment results for differentially expressed
267 genes; Table S3, basic statistics for the joint and homoeologous gene coexpression networks;
268 Table S4, GO enrichment results for the homoeologous gene coexpression network modules;
269 Table S5, basic statistics for the wild and domesticated gene coexpression networks; Table S6,
270 GO enrichment results for the differentially coexpressed genes; Table S7, genes showing both
271 differential expression and differential coexpression; Table S8, genes showing both differential
272 expression and differential coexpression related to cell wall biosynthesis and the cytoskeleton.
273 Custom R scripts of the analysis performed in this paper are available at
274 <https://github.com/Wendellab/AD1FiberDom>.

275

276 **Results**

277 **Transcriptome dynamics accompanying fiber development and domestication**

278 Here, we compared gene expression in fiber for three wild and three domesticated accessions of
279 *G. hirsutum* across four developmental time points. These accessions have previously been
280 shown to encompass the genetic diversity within the wild and domesticated *G. hirsutum* gene
281 pools (Wendel et al. 1992; Brubaker and Wendel 1994; Grover et al. 2017; Yoo and Wendel
282 2014). These time points represent primary fiber cell elongation (5 and 10 dpa) and the transition
283 to secondary cell wall synthesis (15 and 20 dpa; Applequist et al. 2001; Rapp et al. 2010; Haigler
284 et al. 2012). A total of 24 RNA-seq libraries were generated with an average of 30 million reads
285 per library (Table S1). About 83.6% of raw reads were mapped to the SNP-tolerant *G. raimondii*
286 reference genome (Paterson et al. 2012), with approximately equal proportions of homoeolog-
287 specific reads assigned to the A- and D- subgenomes (Table S1). Principal components analysis
288 (PCA) of the fiber transcriptome profiles revealed that the first component accounts for 76.1% of
289 the total variance and mainly clusters fiber samples by developmental stage (Fig. 2); there was
290 no apparent clustering based on domestication. Clustering based on domestication was observed,
291 however, when we performed PCA on each individual timepoints (Fig. S1). We noted that a
292 single 10 dpa library of *G. hirsutum* var. *yucatanense* clustered with the remaining 20 dpa
293 samples (Fig. 2; red arrow), indicating a potentially mislabeled library; therefore, we replaced
294 this sample with a previously sequenced RNA-seq library of 10 dpa *G. hirsutum* var.
295 *yucatanense* fibers (SRX062250; Fig. 2, blue circle) for all subsequent analyses.
296
297 Consistent with previous reports, (Hovav et al. 2008c; Yoo and Wendel 2014; Fang et al.
298 2017b), the majority of all cotton genes (including homoeologs) were expressed during fiber
299 development. A total of 29,706 homoeologous gene pairs were jointly expressed during these
300 timepoints, accounting for 79.2% of the 37,505 reference gene models in the diploid *G.*

301 *raimondii* genome. When considering the homoeologs individually, expression is slightly lower;
302 i.e., 68.0% of homoeologs were expressed in developing fibers. The number of homoeologs
303 expressed from each of the two subgenomes was approximately evenly distributed, 25,474 and
304 25,522 in the A- and D- subgenomes, respectively.

305

306 In order to disentangle the fiber transcriptomic changes due to the effects of *domestication*,
307 *development*, or their complex interactions, we conducted differential gene expression (DGE)
308 analyses by employing a multivariate design of “~ *domestication* + *development* +
309 *domestication:development*”s (Table 1, top) in addition to univariate, pairwise comparisons
310 (Table 1, bottom). By blocking the accessions selected specifically so that they represented the
311 endpoints “wild” vs. “domesticated”, thereby treating accessions as pseudo-biological replicates,
312 we are able to better identify regulatory evolution that has accompanied the domestication
313 process (Yoo and Wendel 2014). Common to both wild and domesticated accessions, significant
314 effects of *development* were found for 29,672 individual homoeologs (58.2% of those
315 expressed). Pairwise comparisons of adjacent developmental stages (Table 1) showed that the
316 largest DGE changes occur between 15 and 20 dpa, which represents the transition from primary
317 elongation to secondary wall synthesis (Haigler *et al.* 2012). The lowest amount of DGE was
318 found between 10 and 15 dpa fibers. Gene Ontology (GO) term enrichment analysis suggests
319 that these DGE genes with developmental effects are primarily involved in metabolism,
320 biosynthesis, and transport (adjusted $P < 0.05$; Table S2). During primary wall synthesis (here, 5
321 and 10 dpa), regulation-related GO terms (adjusted $P < 0.05$; Table S2) were significantly down-
322 regulated for both wild and domesticated fiber. Towards the end of elongation (i.e., 10 to 15
323 dpa), fatty acid and lipid biosynthesis and metabolism related GO terms were enriched for genes

324 upregulated only in the domesticated accessions, (adjusted $P < 0.05$; Table S2), which may
325 suggest that lipids, as components of cell membranes and vesicles, are trafficking necessary
326 molecular components to the expanding fiber cell wall. On the contrary, genes upregulated in
327 wild fibers at 10 to 15 dpa were enriched for cellulose biosynthesis and metabolism, suggesting
328 an earlier initiation of secondary cell wall biosynthesis, as previously shown (Table S2;
329 Butterworth et al. 2009). DGE was highest during the transition to secondary cell wall
330 biosynthesis (between 15 and 20 dpa), where a staggering number of GO categories were
331 enriched (Table S2); these include various biosynthesis, transport, and molecule modification
332 terms, including those related to cell wall biosynthesis.

333

334 Following the effect of *development*, a smaller number of homoeologs (10,218; 20% of
335 expressed genes) were identified to show significant *domestication* effect, regardless of dynamic
336 changes during fiber development. Here, direct comparisons between wild and domesticated
337 accessions at each developmental stage revealed the most DGE at the latest stage examined, i.e.,
338 20 dpa, whereas the least DGE was observed at the beginning of the transition stage (15 dpa;
339 Table 1). These *domestication* DGE genes were found to be associated with a large number of
340 GO categories, including localization, transport, and detection of stimuli (adjusted $P < 0.05$;
341 Table S2). For pairwise comparisons, the highest number of enriched GO terms was observed at
342 20 dpa, including small molecule biosynthesis and metabolism-related processes, which were
343 upregulated in wild fiber (adjusted $P < 0.05$; Table S2).

344

345 Only 85 DGE genes exhibited a significant interaction effect of *domestication* over
346 *developmental stages* (Table 1). DGE was greatest between 15 and 20 dpa; however, given the
347 small number of detected genes, no GO categories were enriched for this interaction comparison.

348

349 **Highly conserved modular organization between A- and D- homoeologous networks in**
350 **allopolyploid cotton fiber**

351 Coexpression network analysis has an additional dimension in polyploid species where the
352 duplicated nature of the genome allows individual homoeologs to retain their original function,
353 to evolve independently (possibly acquiring new or partitioned functions), and to interact in a
354 variety of adaptive, maladaptive, and neutral ways. In considering polyploid coexpression
355 networks, one approach is to treat all homoeologous pairs just as one would treat alleles of a
356 single gene, regardless of parental origin. Alternatively, each homoeologous network
357 (comprising homoeologs of the same parental origin) might be considered independently to
358 reveal how networks respond to polyploidization. While the assumption that homoeologous pairs
359 are equivalently co-regulated will be violated for any number of genes, construction of a
360 coexpression network from the summed expression of all homoeologs into a single “gene pair”
361 (i.e., treating them as “alleles”) provides a null model (referred to as the “joint network”
362 hereafter; Table 2) for comparison to individual, homoeologous gene networks (referred to as the
363 “homoeologous network” hereafter, Table 2). Here, such comparisons were conducted to detect
364 differences that may be relevant to fiber development and domestication, and to explore
365 network-level responses to genome doubling.

366

367 Construction of the joint network resulted in partitioning the 29,706 gene pairs into 26 modules
368 (i.e., clusters of highly co-expressed genes; Table S3). Using this joint network as reference, we
369 asked whether its modular organization was separately preserved by both the A- and D-
370 homoeologous gene expression datasets. When the homoeologous gene expression data (i.e.,
371 from 25,474 A- and 25,522 D-homoeologs) was fit to the modular organization of the gene-pair
372 (i.e., joint) network, nearly all modules exhibited high preservation scores for both the A- and D-
373 homoeologous networks ($Z_{\text{summary}} > 10$, see Methods; Fig. 3). This indicates general preservation
374 of gene-pair coexpression relationships by both the A- and D- subgenomes. Notably, a lack of
375 preservation was observed for modules 15 and 26 in either one or both of the homoeologous
376 networks ($Z_{\text{summary}} < 10$ and high medianRank scores; Fig. 3). For module 26, this may be due to
377 a lack of statistical power, given it contains the smallest number of genes ($n = 72$). For module
378 15, however, poor preservation according to both Z_{summary} score and the medianRank score is
379 only observed in the D-homoeologous network, potentially representing expression and/or
380 functional divergence for genes contained in that module.

381
382 In addition to evaluating the topological similarity between each individual homoeologous
383 network relative to the joint network, we also assessed the expression contribution of each
384 homoeolog (i.e., homoeolog expression bias) across each module in the joint network (Table 3).
385 Of the 16,273 gene pairs that exhibit homoeolog expression bias (based on DGE, see methods),
386 significantly more pairs display higher expression of the D-homoeolog than the A-homoeolog
387 (8521 vs 7752; Chi-squared test $P < 0.05$; Table 3). At the module level, 16 modules contained
388 balanced numbers of A- and D- biased expression, while 9 modules exhibited D-biased
389 expression (not including Module 0, which contains unassigned genes) and only one module

390 exhibited a strong bias toward higher expression of A-homoeologs (i.e., module 15, which
391 contains 216 A-biased vs 24 D-biased gene pairs; Table 3). This was the same A-biased module
392 that showed asymmetrical preservation of the joint topology between the A- and D-
393 homoeologous expression datasets. Considering that almost all other modules in the joint
394 network were well preserved by homoeologous networks, the observed general and module-level
395 D-homoeolog biases do not appear to significantly alter gene co-expression relationships.

396

397 **The majority of homoeologous gene pairs are in separate modules in the polyploid network**

398 We next constructed the homoeologous coexpression network based on the expression of
399 individual homoeologs. A total of 50,996 homoeologs expressed during fiber development (see
400 above) were clustered into 52 coexpressed modules containing between 38 and 9,314 genes
401 (Table S3). Notably, this doubles the number of joint network modules (52 versus 26), indicating
402 that, while the general network topology is largely preserved, coexpression relationships within
403 each subgenome are distinct enough to generate separate modules. In support of this hypothesis,
404 only one fifth of all paired homoeologs (i.e., 7,561 out of 37,505 homoeologous gene pairs) were
405 placed into the same module. This implies that the coexpression divergence between
406 homoeologous genes mainly occurred at the intermodular level, whereas the intramodular
407 relationships are most likely preserved.

408

409 Among these 52 modules, 28 exhibited expression profiles that were significantly associated
410 with developmental stage and/or domestication status (ANOVA, $P < 0.05$; Fig. S3). DGE genes
411 exhibiting significant *development* and/or *domestication* effects were enriched in 22 modules
412 (42%; GSEA adjusted $P < 0.05$). Among those modules, 20 exhibited enrichment for

413 *development* DGE, with 14 modules also enriched for *development:domestication* interaction
414 DGE (Table S3). Enrichment for the *domestication* effects was found for 11 modules, which
415 usually also were enriched for *development* (8 modules) and *development:domestication* (10
416 modules) effects (GSEA adjusted $P < 0.5$; Table S3).

417

418 Each module was functionally annotated by enriched GO terms, where several modules were
419 identified with relevance to key biological processes of fiber development (Table S4, Fig. S2).
420 Module 6, for example, was enriched for cell wall modification, sucrose metabolic and
421 biosynthetic process, and regulation of meristem structural organization (GSEA, adjusted $P <$
422 0.05; Table S4) and showed higher expression at 15 and 20 dpa (Fig. 4). Module 8 also showed
423 enrichment for a large number of biological processes, including cellulose biosynthetic process
424 and cell wall macromolecule catabolic process (GSEA, adjusted $P < 0.05$; Table S4); this module
425 showed low expression at 10 and 15 dpa, but spiked at 20 dpa, when the secondary cell wall is
426 forming (Fig. 4). Module 41, enriched for cell wall modification, regulation of meristem
427 structural organization, and sucrose biosynthetic processes (GSEA, adjusted $P < 0.05$; Table S4),
428 showed higher expression at 10 dpa and 15 dpa in wild fiber, but much lower expression in
429 domesticated fiber (Fig. 4). Other modules also showed enrichment for GO terms related to the
430 cell wall (module 9), meristem structure and development (modules 7,12,17), cellulose (modules
431 28,33,37), and sucrose (modules 1, 9, 17, 23, 37, 44, 46); however these GO terms only represent
432 a subset of the enrichment observed (Table S4, Fig. S2).

433

434 **Domestication has substantially altered the fiber network topology**

435 Previous work on the cotton seed transcriptome revealed the extensive effects of domestication
436 on the coexpression relationships among genes and homoeologs (Hu *et al.* 2016), a key
437 conclusion being that interconnectivities became more tightly regulated in domesticated cotton.
438 Here, we observed a similar effect in the fiber transcriptome, by comparing the individual
439 coexpression networks each constructed from the wild and domesticated homoeolog expression
440 datasets (Table 2). Over twice as many gene modules were recovered from the wild versus
441 domesticated network (107 versus 47 modules, respectively; Table S5), suggesting tighter
442 coregulation of genes in domesticated fiber. While a similar number of modules were found
443 significantly associated with fiber development in both networks (10 and 9, respectively;
444 ANOVA, $p < 0.05$, Table S5), these modules do not contain similar sets of genes according to
445 the marginal correspondence analysis (Fig. S4, bolded modules). Instead, over 10 times more
446 genes were clustered in these development-related modules in the domesticated network (33,037
447 genes) than those in the wild network (2,426 genes). It appears that several modules not
448 significantly related to fiber development in wild cotton were recruited and became co-regulated
449 during domestication.

450

451 Module preservation tests (Fig. 5) showed that of the 108 modules present in the wild cotton
452 network, only 26 were strongly preserved when fitting the domesticated fiber gene expression
453 data to the wild network topology ($Z_{\text{summary}} > 10$, low medianRank score), indicating a large
454 reorganization of network relationships accompanying domestication. Moreover, preservation of
455 the development-related modules is minimal, with only 3 strongly preserved in domesticated
456 cotton. Taken together, these results show that domestication has condensed the coexpression

457 network in cotton fibers, resulting in tighter and denser connections among genes, in a similar
458 manner as for the cotton seed gene coexpression network (Hu *et al.* 2016).

459
460 In addition to the coexpression networks, differential correlation (DC) analysis was performed to
461 directly contrast the gene-to-gene correlations between wild and domesticated transcriptomes.
462 Nearly one third of all homoeologs (i.e., 16,503 genes or 32.4%) exhibited significant DC
463 change under domestication (adjusted $P < 0.05$; see Methods). GO enrichment analysis showed
464 that these DC genes were enriched for a wide variety of Biological Processes GO Terms ($p <$
465 0.05 , $q < 0.05$; Table S6; Alexa and Rahnenfuhrer 2016), including protein localization, protein
466 transport, catabolic processes, and DNA metabolic processes. Among the DC genes, 10,026
467 overlapped with the list of DGE genes that showed differential expression between wild and
468 domesticated cotton (i.e. a *domestication* effect), containing roughly equivalent number of genes
469 between subgenomes (4,941 A-homoeologs and 5,085 D-homoeologs; Table S7). This large list
470 of overlapping DGE/DC genes also supports the observed substantial change in the cotton fiber
471 transcriptome by domestication, and thus was also used to pinpoint the target domestication
472 genes involved in key biological processes and functions in fiber.

473

474 **The effects of domestication on fiber cell wall synthesis**

475 Changes in gene expression and co-regulation accompanying domestication represent the
476 molecular consequences of intense, directional selection on the fiber phenotype. Obvious targets
477 of selection include genes involved in specifying the composition and regulation of the
478 cytoskeleton and fiber cell wall. Previous work has revealed some of the genes involved in
479 cotton fiber synthesis, specifically those involved in cell wall biosynthesis and cytoskeletal

480 activities (Haigler *et al.* 2009; Szymanski and Cosgrove 2009; Taylor-Teeple *et al.* 2015).
481 While 150 genes related to the cytoskeleton and/or cell wall biosynthesis exhibited both
482 significant DGE and DC across the developmental timepoints (Table S8), these were distributed
483 among 11 (of 52) modules. These DGE/DC cell wall and cytoskeletal genes encode a diversity of
484 proteins, including cellulose synthases, expansins, actins, tubulins, and ethylene response
485 sensors, among others.

486 Examples of genes altered during domestication include several genes involved in
487 generating the cell wall, i.e., galacturonosyltransferase 3 (GAUT3), three glycosyl hydrolases, an
488 exostosin gene, and a CESA gene. Only GAUT3 was upregulated under domestication (at 5 and
489 10 DPA). The remaining five genes were all downregulated at various stages: the hydrolases at 5
490 DPA, the CESA gene at 15 DPA, and the exostosin gene at 10 and 15 DPA. This general trend of
491 downregulation was also observed for genes involved in lignin biosynthesis, which is also
492 relevant to cell wall synthesis, although the gene expression changes suggest a potential
493 reduction in lignin content. Both O-methyltransferase and cytochrome P450 are involved in
494 lignin synthesis and both were broadly downregulated (5/10/15 DPA and 20 dpa, respectively).
495 A single laccase gene, which is involved in lignin degradation, shows a more complicated
496 pattern, i.e., upregulated in domesticated cotton at 5 DPA but downregulated at both 15 and 20
497 DPA. Finally, tubulin β , a microtubule component of the cytoskeletal framework, was also
498 downregulated at 10 DPA. Interestingly, while one might expect other canonical cell-wall related
499 genes to exhibit significant DE and DC, such as additional CESA homologs, actins/actin-
500 regulating genes, other tubulins, and cytoskeletal motors, we did not observe any significant
501 changes in the stages evaluated for *G. hirsutum*.

502 Also important in fiber development are genes related to regulation and/or cell signaling.
503 We found several genes that have significant expression changes correlated with domestication.
504 These include three MYB transcription factors (i.e. MYB3, 7, and 66), which were
505 downregulated in domesticated cotton at 5 DPA (MYB3 and MYB7) and 20 DPA (MYB66) and
506 a heavy metal transport protein (potentially involved in signaling), which was upregulated at 10
507 and 20 DPA. Also upregulated are homologs of ERF1, ERS1, and ETR2, which function in
508 ethylene response and regulation, and have been associated with cotton fiber elongation (Li et al.
509 2007).

510 Finally, we find a large number of genes (Table S7) encoding reactive oxygen species
511 (ROS), which are known to function in diverse cellular processes, including cell expansion
512 (Cosgrove 2005), patterning (Gapper and Dolan 2006), and polar growth (Mori and Schroeder
513 2004). Previous research has shown upregulation of ROS producing and regulating genes in
514 domesticated cotton fiber (Hovav *et al.* 2008a; Chaudhary *et al.* 2009). Accordingly, we see
515 several peroxidase genes upregulated within different modules in domesticated fiber, in addition
516 to two NADPH oxidoreductase-encoding genes.

517 While these and the above mentioned genes are not an exhaustive list of biologically
518 relevant gene expression changes, they serve to underscore the complexity of the expression
519 changes that have altered the fiber developmental program as a result of domestication. A
520 complete list of fiber-related genes and their expression changes is in Table S8.

521

522

523 Discussion

524 The suite of coordinated changes in gene expression that underlie polyploidy and domestication
525 are of substantial interest from both evolutionary and agronomic standpoints. Decades of

526 research have elucidated the myriad changes in gene expression that are stimulated during
527 polyploidization and have highlighted the opportunities that a redundant genome might provide
528 for evolutionary innovation (Osborn *et al.* 2003; Adams and Wendel 2005; Chen and Ni 2006;
529 Chen 2007). Simultaneously, changes in gene expression under domestication have been studied
530 in multiple species, either on a gene-by-gene basis (Hirano *et al.* 1998; Konishi *et al.* 2006; Cong
531 *et al.* 2008) or genome-wide (Doebley *et al.* 2006; Hovav *et al.* 2008; Paran and van der Knaap
532 2007; Jin *et al.* 2008; Ramírez-González *et al.* 2018; Dong *et al.* 2019; Sauvage *et al.* 2017).
533 Allopolyploid cotton provides the opportunity to evaluate the consequences of both processes,
534 having been domesticated *ca.* 5000 years ago (Wendel *et al.* 2010).

535
536 Previous research in cotton has demonstrated that homoeologs commonly diverge in expression
537 pattern, and possibly function (Chaudhary *et al.* 2009; Guan *et al.* 2014; Lv *et al.* 2016; Liu *et al.*
538 2019), and that domestication has dramatically altered the fiber transcriptome (Hovav *et al.*
539 2008b; Rapp *et al.* 2010; Yoo and Wendel 2014). We evaluate the gene coexpression network in
540 developing fibers from both wild and domesticated representatives of *G. hirsutum* to characterize
541 the patterns of expression evolution that have been caused by polyploidy and domestication.

542
543 **Coordinated expression changes of cell wall-related genes during domestication**
544 Cotton fiber development is highly complex, involving myriad metabolic processes
545 incorporating most of the genes in the genome. A much longer, stronger, finer, whiter fiber, one
546 of the most obvious outcomes of directional selection under domestication, is likely controlled
547 by those genes related to cell wall biosynthesis and the cytoskeleton. Here we find many of these
548 genes are significantly differentially expressed *and* co-expressed across developmental

549 timepoints and under domestication, suggesting a coordinated role in conferring the domesticated
550 phenotype. For example, upregulation of GAUT3, results in an increase in pectin (Sterling et al.
551 2006) and thereby cell wall flexibility (Stiff and Haigler 2012), while a concomitant
552 downregulation of three glycosyl hydrolases reduces hydrolyzation of the glycosidic bonds
553 found within the cell wall (Xu *et al.* 2004), thereby increasing stability of the primary cell wall
554 during elongation. These exemplar genes represent distinct examples of a broader pattern of
555 coordinated regulation among genes. We found large numbers of genes involved in reactive
556 oxygen species (ROS) and ethylene biosynthesis/response that exhibited DE and DC, whose
557 coordination between and among pathways affects cell wall loosening and regulation of
558 elongation (Cosgrove 2005; Shi *et al.* 2006; Li *et al.* 2007; Hovav *et al.* 2008a; Qin *et al.* 2008;
559 Chaudhary *et al.* 2009; Zhang *et al.* 2016). Likewise, we found coordination of multiple genes
560 involved in regulating lignin production at the early stages of fiber development, which results in
561 reduced lignin production in the early stages of fiber development and absence of lignin in the
562 later stages, resulting in lighter and smoother fibers. Interestingly, although each fiber is
563 composed primarily of cellulose, only a single CESA gene (*CESA8*-homolog) exhibited
564 significant expression changes in our data and for only a single timepoint. Notably, no other
565 cellulose synthase paralogs exhibited expression differences under domestication here despite the
566 increased cellulose content of domesticated fiber, possibly indicating that regulation of cellulose
567 synthase isoforms may not occur at the transcriptional level.

568

569 **Evolution of networks duplicated by polyploidy**

570 Polyploidy, or whole genome duplication, is a recurrent and ongoing phenomenon that has
571 influenced the evolutionary history of all plants (Jiao *et al.* 2011; Soltis *et al.* 2014; Wendel

572 2015; Soltis et al. 2014; One Thousand Plant Transcriptomes Ini...). The gene expression and
573 regulatory consequences of polyploidization are many, and include extensive transcriptional
574 rewiring as the independently evolved regulatory environments merge and interact (Ramírez-
575 González et al. 2018; Yang et al. 2016; Yoo et al. 2014; Gallagher et al. 2016; Bao et al. 2019;
576 Hu et al. 2016; Edger et al. 2019). Specifically, novel interactions between the *cis* and *trans*
577 factors derived from each progenitor species have consequences for the polyploid expression
578 environment, reverberating throughout the network (De Smet and Van de Peer 2012; Gallagher
579 *et al.* 2016; Bao *et al.* 2019).

580

581 Previous work on cotton seed networks (Hu *et al.* 2016) demonstrated extensive rewiring of the
582 seed oil network as a consequence of hybridization between the two independently evolved
583 diploid progenitor species and subsequent genome doubling; however, this study was limited by
584 aggregating the expression of homoeologous gene pairs into a single cumulative value. Here we
585 extend this approach to compare the two co-resident homoeologous gene networks in
586 allopolyploid cotton. We find that in general, each homoeologous network (the A- and D-
587 subgenome networks) is largely similar to the joint network created by combining homoeologous
588 gene expression (*a la* Hu *et al.* 2016), indicating a general preservation of network topology
589 between the two homoeologous networks. Comparisons between the joint and homoeologous
590 coexpression networks mirrors this observation of general conservation, as the homoeologous
591 network contains the expected doubling of the number of modules (52 versus 26 modules in the
592 homoeologous versus joint network, respectively).

593

594 In contrast to preservation at the module level, the homoeologous networks themselves were
595 highly divergent in module membership. Comparisons between the A and D networks revealed
596 that only a quarter of the expressed homoeologs were found in the same module. Together, these
597 results indicate that the general topology of the fiber network is similar in the A- and D-
598 homoeologous networks, but that coexpression differences among homoeologs are somewhat
599 different. These observations are relevant to our understanding of polyploidy in general and to
600 cotton in particular. With respect to the former, our results demonstrate a dimension of
601 homoeologous coexpression evolution that has not previously been addressed, and which may
602 comprise a common feature of allopolyploid genomes. Recent work in cotton has demonstrated
603 complex interactions between independently evolved *cis* and *trans* acting factors (Bao *et al.*
604 2019), which become combined in a common *trans* environment with the onset of polyploidy.
605 We speculate that the myriad novel interactions between regulatory sequences and trans-acting
606 factors become superimposed on the pre-existing regulatory environments of the divergent
607 diploid progenitors, magnifying expression and coexpression differences in the allopolyploid.
608 Understanding how this new combinatorial complexity is shaped by the evolutionary process or
609 in response to particular selective regimes remains a promising avenue for understanding
610 polyploid evolution and the origin of new phenotypes.

611

612 With respect to cotton, our coexpression network results may offer insight into the transgressive,
613 superior fibers of cultivated allopolyploid cotton relative to their diploid progenitors. The
614 generalized conservation of coexpression modular structure for the A and D homoeologs
615 suggests that even though the D-genome progenitor does not produce spinnable fiber, the
616 underlying architecture of the developmental program for producing epidermal seed trichomes is

617 conserved. This realization may lead to an enhanced understanding of both the superior fiber
618 produced by polyploid cotton and previous observations of D-genome recruitment for fiber
619 production during domestication (Jiang *et al.* 1998; Lacape *et al.* 2005, 2010; Ulloa *et al.* 2005;
620 Han *et al.* 2006; Rong *et al.* 2007; Qin *et al.* 2008; Said *et al.* 2013, 2015; Grover *et al.* 2019).
621 Our results thus direct attention to unraveling the evolutionary and regulatory differences
622 responsible for the A vs. D expression and coexpression differences, and how these were altered
623 by polyploidization and domestication.

624

625 The contribution of the D-genome to polyploid cotton fiber is reiterated by the generalized bias
626 in expression toward this parent, both in overall gene expression and module expression.
627 Although there are many studies demonstrating homoeolog bias in cotton (Hovav *et al.* 2008b;
628 Yoo and Wendel 2014; Zhang *et al.* 2015), we provide a new perspective on this phenomenon
629 here, relating homoeolog expression bias to coexpression module membership, domestication
630 status, and developmental stage. Notably, our estimates of differential expression relative to
631 domestication and development all show bias towards higher expression of the D-subgenome,
632 but that differential coexpression analysis highlights the fact that an individual module may
633 reverse this trend. Specifically, module 15 exhibits higher expression in the A-subgenome where
634 it was also more strongly preserved. This module exhibits lower expression in domesticated
635 cotton that is generally restricted towards the later developmental timepoints. These observations
636 suggest that reducing expression of this module at specific timepoints may be important for the
637 domesticated fiber phenotype, making those genes suitable targets for agronomic improvement
638 via RNAi or other expression-reducing modification.

639

640 **Domestication recruits genes into more tightly regulated modules**

641 Coexpression networks provide a useful summary of complex, multidimensional data, such as
642 changes in transcriptional relationships among genes across developmental time and between
643 accessions. Coexpression analysis of cotton seed domestication demonstrated, quite remarkably,
644 that changes under domestication within a species were more extensive than natural evolutionary
645 changes between species (Hu *et al.* 2016), and that domestication resulted in a more highly
646 condensed, or tighter network. Similar to the latter, domestication has extensively rewired the
647 cotton fiber network, resulting in fewer and more densely connected modules. While the number
648 of modules was reduced by a little more than half in domesticated cotton, the number of genes
649 associated with domestication-relevant modules increased 13-fold to nearly half of the genes in
650 the allopolyploid genome. This results in nearly one-third of genes exhibiting evidence of
651 differential correlation. These genes were largely recruited from wild modules that were not
652 significantly associated with fiber development, and whose coordinated expression contributed
653 to the domesticated fiber phenotype. This high level of change following cotton domestication is
654 also concordant with the high number of QTL that are found in studies associated with wild and
655 domesticated Upland cotton (Jiang *et al.* 1998; Lacape *et al.* 2005, 2010; Ulloa *et al.* 2005; Han
656 *et al.* 2006; Rong *et al.* 2007; Hovav *et al.* 2008b; Qin *et al.* 2008; Said *et al.* 2013, 2015; Fang *et*
657 *al.* 2017a, 2017c; Wang *et al.* 2017; Grover *et al.* 2019). While the results discussed here only
658 represent only a small sampling of the wild and domesticated gene pools, the level of diversity
659 within wild and domesticated cotton populations suggests that our results would be reiterated in
660 an expanded sampling of accessions. Taken together, our results demonstrate the amazing
661 complexity that underlies the domestication of cotton fiber.

662

663 Enrichment of highly connected intramodular hub genes can show functions that act together
664 during fiber development. Gene set enrichment analysis (GSEA) was used to assess which
665 modules were associated with various biological processes. We detected enrichment for genes
666 within modules (Table S4, Fig. S2), with gene sets being enriched for processes such as cell wall
667 biogenesis, fatty acid biosynthesis, flavonoid biosynthesis, and many others, underscoring the
668 complexity of the cotton fiber transcriptome and its developmental dynamics. To some extent
669 this is not an unexpected result, given the observation that the majority of the genes in the
670 genome are expressed at some point during cotton fiber development; accordingly, we might
671 expect complex coexpression relationships that involve or invoke the majority of cellular
672 processes. We suggest that the modules described here as enriched for specific biological
673 processes represent starting points for functional genomic studies targeting intramodular hub
674 genes and for additional techniques designed to layer with transcriptomic coexpression data (e.g.,
675 protein interaction networks, transcription factor binding networks).

676

677

678

Acknowledgements

679 We thank Anna Tuchin and Simon Renny-Byfield for assistance with RNA extractions. This
680 work was funded by Cotton Incorporated and by the National Science Foundation Plant Genome
681 Program. Joseph P. Gallagher was supported by an NSF Graduate Research Fellowship
682 (DGE1247194). The authors acknowledge computational support and assistance from the Iowa
683 State University ResearchIT Unit (<http://researchit.las.iastate.edu/>).

684

685

686 **References**

687 Adams, K. L., and J. F. Wendel, 2005 Novel patterns of gene expression in polyploid plants.

- 688 Trends Genet. 21: 539–543.
- 689 Alexa, A., and J. Rahnenfuhrer, 2016 *topGO: Enrichment Analysis for Gene Ontology*.
- 690 Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq--a Python framework to work with high-
- 691 throughput sequencing data. Bioinformatics 31: 166–169.
- 692 Applequist, W. L., R. Cronn, and J. F. Wendel, 2001 Comparative development of fiber in wild
- 693 and cultivated cotton. Evol. Dev. 3: 3–17.
- 694 Bao, Y., G. Hu, L. E. Flagel, A. Salmon, M. Bezanilla *et al.*, 2011 Parallel up-regulation of the
- 695 profilin gene family following independent domestication of diploid and allopolyploid
- 696 cotton (*Gossypium*). Proc. Natl. Acad. Sci. U. S. A. 108: 21152–21157.
- 697 Bao, Y., G. Hu, C. E. Grover, J. Conover, D. Yuan *et al.*, 2019 Unraveling cis and trans
- 698 regulatory evolution during cotton domestication. Nat. Commun. 10: 5399.
- 699 Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and
- 700 Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. 57: 289–
- 701 300.
- 702 Brubaker, C. L., and J. F. Wendel, 1994 Reevaluating the origin of domesticated cotton
- 703 (*Gossypium hirsutum* ; Malvaceae) using nuclear restriction fragment length polymorphisms
- 704 (RFLPs). Am. J. Bot. 81: 1309–1326.
- 705 Butterworth, K. M., D. C. Adams, H. T. Horner, and J. F. Wendel, 2009 Initiation and Early
- 706 Development of Fiber in Wild and Cultivated Cotton. International Journal of Plant Sciences
- 707 170: 561–574.
- 708 Chaudhary, B., R. Hovav, L. Flagel, R. Mittler, and J. F. Wendel, 2009 Parallel expression
- 709 evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and
- 710 polyploid cotton (*Gossypium*). BMC Genomics 10: 378.

- 711 Chen, Z. J., 2007 Genetic and epigenetic mechanisms for gene expression and phenotypic
712 variation in plant polyploids. *Annu. Rev. Plant Biol.* 58: 377–406.
- 713 Chen, Z. J., and Z. Ni, 2006 Mechanisms of genomic rearrangements and gene expression
714 changes in plant polyploids. *Bioessays* 28: 240–252.
- 715 Cong, B., L. S. Barrero, and S. D. Tanksley, 2008 Regulatory change in YABBY-like
716 transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat.*
717 *Genet.* 40: 800–804.
- 718 Cosgrove, D. J., 2005 Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6: 850–861.
- 719 De Smet, R., and Y. Van de Peer, 2012 Redundancy and rewiring of genetic networks following
720 genome-wide duplication events. *Curr. Opin. Plant Biol.* 15: 168–176.
- 721 Doebley, J. F., B. S. Gaut, and B. D. Smith, 2006 The molecular genetics of crop domestication.
722 *Cell* 127: 1309–1321.
- 723 d'Eeckenbrugge, G. C., and J.-M. Lacape, 2014 Distribution and Differentiation of Wild, Feral,
724 and Cultivated Populations of Perennial Upland Cotton (*Gossypium hirsutum* L.) in
725 Mesoamerica and the Caribbean. *PLoS One* 9: e107458.
- 726 Fang, D. D., 2018 Cotton Fiber Genes and Stable Quantitative Trait Loci, pp. 151–178 in *Cotton*
727 *Fiber: Physics, Chemistry and Biology*, edited by D. D. Fang. Springer International
728 Publishing, Cham.
- 729 Fang, L., H. Gong, Y. Hu, C. Liu, B. Zhou *et al.*, 2017a Genomic insights into divergence and
730 dual domestication of cultivated allotetraploid cottons. *Genome Biol.* 18: 33.
- 731 Fang, L., X. Guan, and T. Zhang, 2017b Asymmetric evolution and domestication in
732 allotetraploid cotton (*Gossypium hirsutum* L.). *The Crop Journal* 5: 159–165.
- 733 Fang, L., Q. Wang, Y. Hu, Y. Jia, J. Chen *et al.*, 2017c Genomic analyses in cotton identify

- 734 signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49:
735 1089–1098.
- 736 Fukushima, A., 2013 DiffCorr: an R package to analyze and visualize differential correlations in
737 biological networks. *Gene* 518: 209–214.
- 738 Gallagher, J. P., C. E. Grover, G. Hu, and J. F. Wendel, 2016 Insights into the Ecology and
739 Evolution of Polyploid Plants through Network Analysis. *Mol. Ecol.* 25: 2644–2660.
- 740 Gallagher, J. P., C. E. Grover, K. Rex, M. Moran, and J. F. Wendel, 2017 A New Species of
741 Cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Syst. Bot.* 42: 115–123.
- 742 Gapper, C., and L. Dolan, 2006 Control of plant development by reactive oxygen species. *Plant*
743 *Physiol.* 141: 341–345.
- 744 Grover, C. E., J. P. Gallagher, J. J. Jareczek, J. T. Page, J. A. Udall *et al.*, 2015a Re-evaluating
745 the phylogeny of allopolyploid *Gossypium* L. *Mol. Phylogen. Evol.* 92: 45–52.
- 746 Grover, C. E., J. P. Gallagher, E. P. Szadkowski, M. J. Yoo, L. E. Flagel *et al.*, 2012 Homoeolog
747 expression bias and expression level dominance in allopolyploids. *New Phytol.* 196: 966–
748 971.
- 749 Grover, C. E., M.-J. Yoo, M. A. Gore, D. B. Harker, R. L. Byers *et al.*, 2019 Genetic analysis of
750 the transition from wild to domesticated cotton (*G. hirsutum*). *bioRxiv* 616763.
- 751 Grover, C. E., X. Zhu, K. K. Grupp, J. J. Jareczek, J. P. Gallagher *et al.*, 2015b Molecular
752 confirmation of species status for the allopolyploid cotton species, *Gossypium ekmanianum*
753 Wittmack. *Genet. Resour. Crop Evol.* 62: 103–114.
- 754 Guan, X., M. Pang, G. Nah, X. Shi, W. Ye *et al.*, 2014 miR828 and miR858 regulate
755 homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre
756 development. *Nat. Commun.* 5: 3050.

- 757 Haigler, C. H., L. Betancur, M. R. Stiff, and J. R. Tuttle, 2012 Cotton fiber: a powerful single-
758 cell model for cell wall and cellulose research. *Front. Plant Sci.* 3: 104.
- 759 Haigler, C. H., B. Singh, G. Wang, and D. Zhang, 2009 Genomics of Cotton Fiber Secondary
760 Wall Deposition and Cellulose Biogenesis, pp. 385–417 in *Genetics and Genomics of*
761 *Cotton*, edited by A. H. Paterson. Springer US, New York, NY.
- 762 Han, Z., C. Wang, X. Song, W. Guo, J. Gou *et al.*, 2006 Characteristics, development and
763 mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theor. Appl.*
764 *Genet.* 112: 430–439.
- 765 Hirano, H. Y., M. Eiguchi, and Y. Sano, 1998 A single base change altered the regulation of the
766 Waxy gene at the posttranscriptional level during the domestication of rice. *Mol. Biol. Evol.*
767 15: 978–987.
- 768 Hovav, R., J. A. Udall, B. Chaudhary, E. Hovav, L. Flagel *et al.*, 2008a The evolution of
769 spinnable cotton fiber entailed prolonged development and a novel metabolism. *PLoS*
770 *Genet.* 4: e25.
- 771 Hovav, R., J. A. Udall, B. Chaudhary, R. Rapp, L. Flagel *et al.*, 2008b Partitioned expression of
772 duplicated genes during development and evolution of a single cell in a polyploid plant.
773 *Proc. Natl. Acad. Sci. U. S. A.* 105: 6191–6195.
- 774 Hovav, R., J. A. Udall, E. Hovav, R. Rapp, L. Flagel *et al.*, 2008c A majority of cotton genes are
775 expressed in single-celled fiber. *Planta* 227: 319–329.
- 776 Hu, G., R. Hovav, C. E. Grover, A. Faigenboim-Doron, N. Kadmon *et al.*, 2016 Evolutionary
777 Conservation and Divergence of Gene Coexpression Networks in *Gossypium* (Cotton)
778 Seeds. *Genome Biol. Evol.* 8: 3765–3783.
- 779 Hu, G., and J. F. Wendel, 2019 Cis – trans controls and regulatory novelty accompanying

- 780 allopolyploidization. *New Phytologist* 221: 1691–1700.
- 781 Ichihashi, Y., J. A. Aguilar-Martínez, M. Farhi, D. H. Chitwood, R. Kumar *et al.*, 2014
- 782 Evolutionary developmental transcriptomics reveals a gene network module regulating
- 783 interspecific diversity in plant leaf shape. *Proc. Natl. Acad. Sci. U. S. A.* 111: E2616–21.
- 784 Jackson, S., and Z. J. Chen, 2010 Genomic and expression plasticity of polyploidy. *Curr. Opin.*
- 785 *Plant Biol.* 13: 153–159.
- 786 Jiang, C.-X., R. J. Wright, K. M. El-Zik, and A. H. Paterson, 1998 Polyploid formation created
- 787 unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. U. S.*
- 788 *A.* 95: 4419–4424.
- 789 Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr *et al.*, 2011 Ancestral
- 790 polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.
- 791 Jin, J., W. Huang, J.-P. Gao, J. Yang, M. Shi *et al.*, 2008 Genetic control of rice plant
- 792 architecture under domestication. *Nat. Genet.* 40: 1365–1369.
- 793 Konishi, S., T. Izawa, S. Y. Lin, K. Ebana, Y. Fukuta *et al.*, 2006 An SNP caused loss of seed
- 794 shattering during rice domestication. *Science* 312: 1392–1396.
- 795 Krapovickas, A., and G. Seijo, 2008 *Gossypium ekmanianum* (Malvaceae), algodon silvestre de
- 796 la Republica Dominicana. *Bonplandia* 55–63.
- 797 Lacape, J.-M., D. Llewellyn, J. Jacobs, T. Arioli, D. Becker *et al.*, 2010 Meta-analysis of cotton
- 798 fiber quality QTLs across diverse environments in a *Gossypium hirsutum* x *G. barbadense*
- 799 RIL population. *BMC Plant Biol.* 10: 132.
- 800 Lacape, J.-M., T.-B. Nguyen, B. Courtois, J.-L. Belot, M. Gibaud *et al.*, 2005 QTL Analysis of
- 801 Cotton Fiber Quality Using Multiple *Gossypium hirsutum* *Gossypium barbadense*
- 802 Backcross Generations. Published in *Crop Sci.* 45: 123–140.

- 803 Langfelder, P., and S. Horvath, 2008 WGCNA: an R package for weighted correlation network
804 analysis. BMC Bioinformatics 9: 559.
- 805 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence
806 Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.
- 807 Li, H.-B., Y.-M. Qin, Y. Pang, W.-Q. Song, W.-Q. Mei *et al.*, 2007 A cotton ascorbate
808 peroxidase is involved in hydrogen peroxide homeostasis during fibre cell development.
809 New Phytol. 175: 462–471.
- 810 Liu, W., Z. Zhang, W. Zhu, Z. Ren, L. Jia *et al.*, 2019 Evolutionary Conservation and
811 Divergence of Genes Encoding 3-Hydroxy-3-methylglutaryl Coenzyme A Synthase in the
812 Allotetraploid Cotton Species *Gossypium hirsutum*. Cells 8.:
- 813 Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion
814 for RNA-seq data with DESeq2. Genome Biol. 15: 550.
- 815 Lv, F., P. Li, R. Zhang, N. Li, and W. Guo, 2016 Functional divergence of GhCFE5 homoeologs
816 revealed in cotton fiber and Arabidopsis root cell development. Plant Cell Rep. 35: 867–881.
- 817 Mori, I. C., and J. I. Schroeder, 2004 Reactive oxygen species activation of plant Ca²⁺ channels.
818 A signaling mechanism in polar growth, hormone transduction, stress signaling, and
819 hypothetically mechanotransduction. Plant Physiol. 135: 702–708.
- 820 Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen *et al.*, 2003 Understanding
821 mechanisms of novel gene expression in polyploids. Trends Genet. 19: 141–147.
- 822 Page, J. T., A. R. Gingle, and J. A. Udall, 2013 PolyCat: a resource for genome categorization of
823 sequencing reads from allopolyploid organisms. G3 3: 517–525.
- 824 Page, J. T., Z. S. Liechty, M. D. Huynh, and J. A. Udall, 2014 BamBam: genome sequence
825 analysis tools for biologists. BMC Res. Notes 7: 829.

- 826 Paran, I., and E. van der Knaap, 2007 Genetic and molecular regulation of fruit and plant
827 domestication traits in tomato and pepper. *J. Exp. Bot.* 58: 3841–3852.
- 828 Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated
829 polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*
830 492: 423–427.
- 831 Pfeifer, M., K. G. Kugler, S. R. Sandve, B. Zhan, H. Rudi *et al.*, 2014 Genome interplay in the
832 grain transcriptome of hexaploid bread wheat. *Science* 345: 1250091.
- 833 Qin, H., W. Guo, Y.-M. Zhang, and T. Zhang, 2008 QTL mapping of yield and fiber traits based
834 on a four-way cross population in *Gossypium hirsutum* L. *Theor. Appl. Genet.* 117: 883–
835 894.
- 836 Rapp, R. A., C. H. Haigler, L. Flagel, R. H. Hovav, J. A. Udall *et al.*, 2010 Gene expression in
837 developing fibres of Upland cotton (*Gossypium hirsutum* L.) was massively altered by
838 domestication. *BMC Biol.* 8: 139.
- 839 R Core Team, 2017 *R: A language and environment for statistical computing*. R Foundation for
840 Statistical Computing., Vienna, Austria.
- 841 Rong, J., F. A. Feltus, V. N. Waghmare, G. J. Pierce, P. W. Chee *et al.*, 2007 Meta-analysis of
842 polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of
843 genes and gene clusters implicated in lint fiber development. *Genetics* 176: 2577–2588.
- 844 Said, J. I., Z. Lin, X. Zhang, M. Song, and J. Zhang, 2013 A comprehensive meta QTL analysis
845 for fiber quality, yield, yield related and morphological traits, drought tolerance, and disease
846 resistance in tetraploid cotton. *BMC Genomics* 14: 776.
- 847 Said, J. I., M. Song, H. Wang, Z. Lin, X. Zhang *et al.*, 2015 A comparative meta-analysis of
848 QTL between intraspecific *Gossypium hirsutum* and interspecific *G. hirsutum* × *G.*

- 849 barbadense populations. Mol. Genet. Genomics 290: 1003–1025.
- 850 Shan, C.-M., X.-X. Shangguan, B. Zhao, X.-F. Zhang, L.-M. Chao *et al.*, 2014 Control of cotton
851 fibre elongation by a homeodomain transcription factor GhHOX3. Nature Communications
852 5.:
- 853 Shi, Y.-H., S.-W. Zhu, X.-Z. Mao, J.-X. Feng, Y.-M. Qin *et al.*, 2006 Transcriptome profiling,
854 molecular biological, and physiological studies reveal a major role for ethylene in cotton
855 fiber cell elongation. Plant Cell 18: 651–664.
- 856 Soltis, P. S., X. Liu, D. B. Marchant, C. J. Visger, and D. E. Soltis, 2014a Polyploidy and
857 novelty: Gottlieb's legacy. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369.:
- 858 Soltis, D. E., C. J. Visger, and P. S. Soltis, 2014b The polyploidy revolution then...and now:
859 Stebbins revisited. Am. J. Bot. 101: 1057–1078.
- 860 Stiff, M. R., and C. H. Haigler, 2012 Recent advances in cotton fiber development, pp. 163–192
861 in *Flowering and fruiting in cotton*, edited by D. M. Oosterhuis and J. Tom Cothren.
- 862 Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert *et al.*, 2005 Gene set
863 enrichment analysis: a knowledge-based approach for interpreting genome-wide expression
864 profiles. Proc. Natl. Acad. Sci. U. S. A. 102: 15545–15550.
- 865 Swanson-Wagner, R., R. Briskine, R. Schaefer, M. B. Hufford, J. Ross-Ibarra *et al.*, 2012
866 Reshaping of the maize transcriptome by domestication. Proc. Natl. Acad. Sci. U. S. A. 109:
867 11878–11883.
- 868 Szymanski, D. B., and D. J. Cosgrove, 2009 Dynamic coordination of cytoskeletal and cell wall
869 systems during plant cell morphogenesis. Curr. Biol. 19: R800–11.
- 870 Taylor-Teeple, M., L. Lin, M. de Lucas, G. Turco, T. W. Toal *et al.*, 2015 An *Arabidopsis* gene
871 regulatory network for secondary cell wall synthesis. Nature 517: 571–575.

- 872 Tuttle, J. R., G. Nah, M. V. Duke, D. C. Alexander, X. Guan *et al.*, 2015 Metabolomic and
873 transcriptomic insights into how cotton fiber transitions to secondary wall synthesis,
874 represses lignification, and prolongs elongation. *BMC Genomics* 16: 477.
- 875 Ulloa, M., S. Saha, J. N. Jenkins, W. R. Meredith Jr, J. C. McCarty Jr *et al.*, 2005 Chromosomal
876 assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton
877 (*Gossypium hirsutum* L.) Joinmap. *J. Hered.* 96: 132–144.
- 878 Wang, M., L. Tu, M. Lin, Z. Lin, P. Wang *et al.*, 2017 Asymmetric subgenome selection and cis-
879 regulatory divergence during cotton domestication. *Nat. Genet.* 49: 579–587.
- 880 Wendel, J. F., 2015 The wondrous cycles of polyploidy in plants. *Am. J. Bot.* 102: 1753–1756.
- 881 Wendel, J. F., C. L. Brubaker, and A. E. Percival, 1992 Genetic diversity in *Gossypium hirsutum*
882 and the origin of Upland cotton. *Am. J. Bot.* 79: 1291–1310.
- 883 Wendel, J. F., C. L. Brubaker, and T. Seelanan, 2010 The Origin and Evolution of *Gossypium*,
884 pp. 1–18 in *Physiology of Cotton*, edited by J. M. Stewart, D. M. Oosterhuis, J. J. Heitholt,
885 and J. R. Mauney. Springer Netherlands, Dordrecht.
- 886 Wendel, J. F., and C. E. Grover, 2015 Taxonomy and Evolution of the Cotton Genus,
887 *Gossypium*, pp. 25–44 in *Cotton*, Agronomy Monograph, American Society of Agronomy,
888 Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.,
889 Madison, WI.
- 890 Wu, T. D., and S. Nacu, 2010 Fast and SNP-tolerant detection of complex variants and splicing
891 in short reads. *Bioinformatics* 26: 873–881.
- 892 Xu, Z., L. Escamilla-Trevio, L. Zeng, M. Lalgondar, D. Bevan *et al.*, 2004 Functional genomic
893 analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Molecular Biology* 55:
894 343–367.

- 895 Yoo, M.-J., X. Liu, J. C. Pires, P. S. Soltis, and D. E. Soltis, 2014 Nonadditive gene expression
896 in polyploids. *Annu. Rev. Genet.* 48: 485–517.
- 897 Yoo, M.-J., and J. F. Wendel, 2014 Comparative evolutionary and developmental dynamics of
898 the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet.* 10: e1004073.
- 899 Yu, J., S. Jung, C.-H. Cheng, S. P. Ficklin, T. Lee *et al.*, 2014 CottonGen: a genomics, genetics
900 and breeding database for cotton research. *Nucleic Acids Res.* 42: D1229–36.
- 901 Zhang, T., Y. Hu, W. Jiang, L. Fang, X. Guan *et al.*, 2015 Sequencing of allotetraploid cotton
902 (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat.*
903 *Biotechnol.* 33: 531–537.
- 904 Zhang, H., A. Li, Z. Zhang, Z. Huang, P. Lu *et al.*, 2016 Ethylene Response Factor TERF1,
905 Regulated by ETHYLENE-INSENSITIVE3-like Factors, Functions in Reactive Oxygen
906 Species (ROS) Scavenging in Tobacco (*Nicotiana tabacum* L.). *Sci. Rep.* 6: 29948.

907

908 **Figure Legends**

909

- 910 **Figure 1.** An abbreviated evolutionary history of cotton, including the model diploid progenitor
911 species and estimated times for divergence and domestication. Branch lengths are not drawn to
912 scale. Wild and domesticated *G. hirsutum* seeds are shown for each species listed, illustrating the
913 effects of divergence and domestication on the epidermal seed trichomes (“fibers”). A 10mm
914 scale bar is shown with each seed. Each of the six accessions examined in this study are
915 represented by a single seed with attached fibers. TX2094, TX2095, and TX665 represent wild
916 accessions, whereas CRB252, Maxxa, and TM-1 represent the improved, domesticated cultivars.

917

918 **Figure 2.** PCA of RNA-seq libraries. PC1, which accounts for more than three quarters of the
919 variance (76.1%), is closely related to developmental stage, which ranges from 5 to 20 days post-
920 anthesis (dpa; key to the right). The red arrow denotes a wild 10 dpa sample that grouped with
921 the 20 dpa samples; this library was excluded from the analyses and replaced by a previously
922 sequenced *G. hirsutum* var. *yucatanense* (TX2094) sample (circled).

923

924 **Figure 3.** Module preservation of aggregate gene coexpression network topology in A- and D-
925 subgenomes. The top two graphs show aggregate module preservation in the A-subgenome fiber
926 expression data, while the bottom two graphs show the aggregate module preservation in the D-
927 subgenome fiber expression data. Module numbers are shown on graphs; the same numbers
928 correspond to the same aggregate modules. Red dashed line marks the preservation threshold at
929 $Z_{\text{summary}} = 10$. For medianRank, a lower score indicates higher preservation relative to higher
930 scores.

931

932 **Figure 4.** Module eigengene expressions of select modules from the homoeologous fiber
933 network. For each module, the barplot presents the module eigengene expression levels centered
934 by means across developmental timepoints for wild and domesticated fiber. Error bars represent
935 the standard errors among three accessions for each genome group at each developmental stage.
936 Developmental stages: red, 5 dpa; green, 10 dpa; blue, 15 dpa; purple, 20 dpa.

937

938 **Figure 5.** Preservation of wild cotton fiber gene coexpression network modules in the
939 domesticated cotton fiber gene expression data. Left panel: modules with lower medianRank
940 scores are relatively more preserved than modules with higher medianRank scores. Right panel:

941 modules above the red dashed line at $Z_{\text{summary}} = 10$ are considered well preserved. Modules with
942 $Z_{\text{summary}} < 10$ were colored grey without labels for clarity. Blue dashed line, $Z_{\text{summary}} = 2$.
943

Table 1. Differential expression analysis of homoeologous fiber transcriptomes


multivariate, LRT tests		
<i>development</i>		29,671
<i>domestication</i>		10,218
interactive effect of <i>domestication</i> over <i>developmental</i> stages		
10 v 5 dpa		11
15 v 10 dpa		15
20 v 15 dpa		18
overall		85
univariate, pairwise comparisons		
		log2 Fold Change > 0
Wild vs Domesticated		log2 Fold Change < 0
at:		
5 dpa	517	288
10 dpa	563	291
15 dpa	419	273
20 dpa	762	745
Wild, between adjacent time points:		
10 v 5 dpa	1052	214
15 v 10 dpa	754	360
20 v 15 dpa	2531	2918
Domesticated, between adjacent time points:		
10 v 5 dpa	1225	407
15 v 10 dpa	695	571
20 v 15 dpa	3983	3322

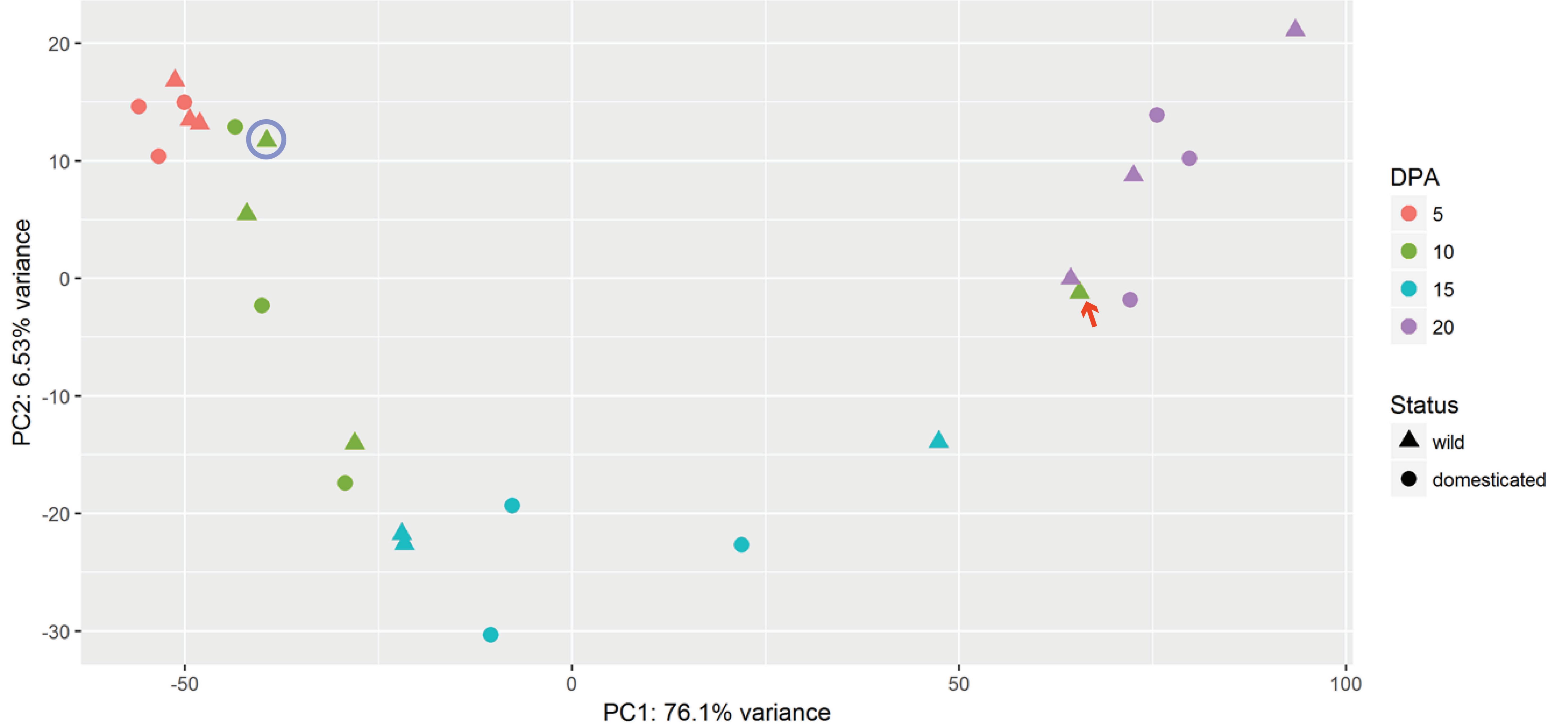
Table 2: Coexpression networks constructed in this study

Network Name	Accessions Included	Basis of Coexpression Relationships
Joint	All accessions	Sum of Homoeologous Gene Pair Expression
Homoeologous	All accessions	Individual Gene Expression
Wild	TX665, TX2094, TX2095	Individual Gene Expression
Domesticated	CRB252, Maxxa, TM-1	Individual Gene Expression

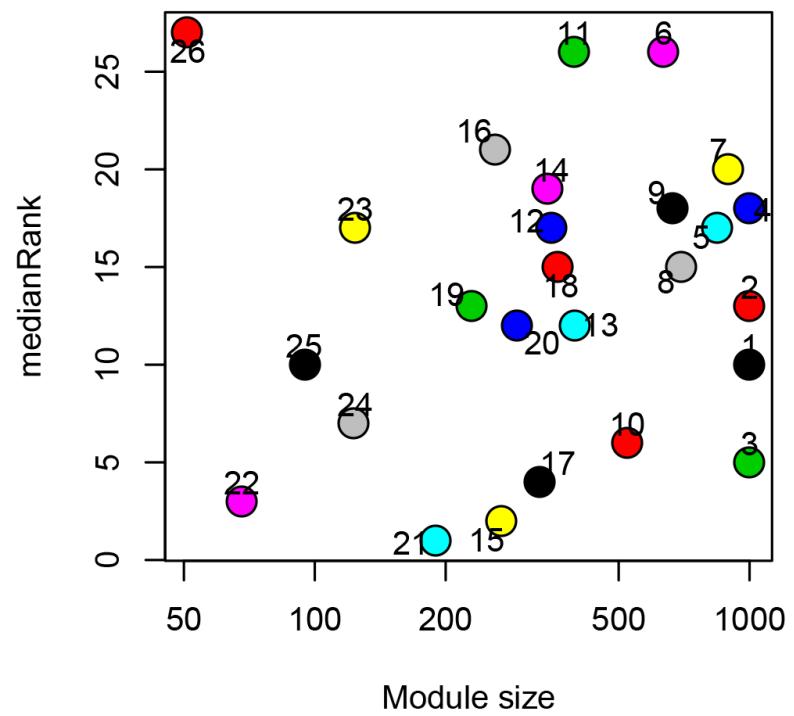
Table 3. Homoeolog expression bias by module for the homoeolog-pair network

Module ID	Gene Number			chi-square	Module level bias (P < 0.05)
	total	A-biased	D-biased		
0	615	280	335	0.03	D-bias
1	3747	1800	1947	0.02	D-bias
2	4591	2180	2411	0.00	D-bias
3	1072	498	574	0.02	D-bias
4	754	360	394	0.22	
5	665	339	326	0.61	
6	567	251	316	0.01	D-bias
7	497	220	277	0.01	D-bias
8	497	242	255	0.56	
9	481	229	252	0.29	
10	302	118	184	0.00	D-bias
11	252	117	135	0.26	
12	288	128	160	0.06	
13	251	124	127	0.85	
14	261	118	143	0.12	
15	240	216	24	0.00	A-bias
16	186	80	106	0.06	
17	122	61	61	1.00	
18	240	101	139	0.01	D-bias
19	132	61	71	0.38	
20	185	87	98	0.42	
21	33	13	20	0.22	
22	34	19	15	0.49	
23	77	39	38	0.91	
24	71	27	44	0.04	D-bias
25	65	24	41	0.03	D-bias
26	48	20	28	0.25	
Sum	16273	7752	8521	0	Unbalanced

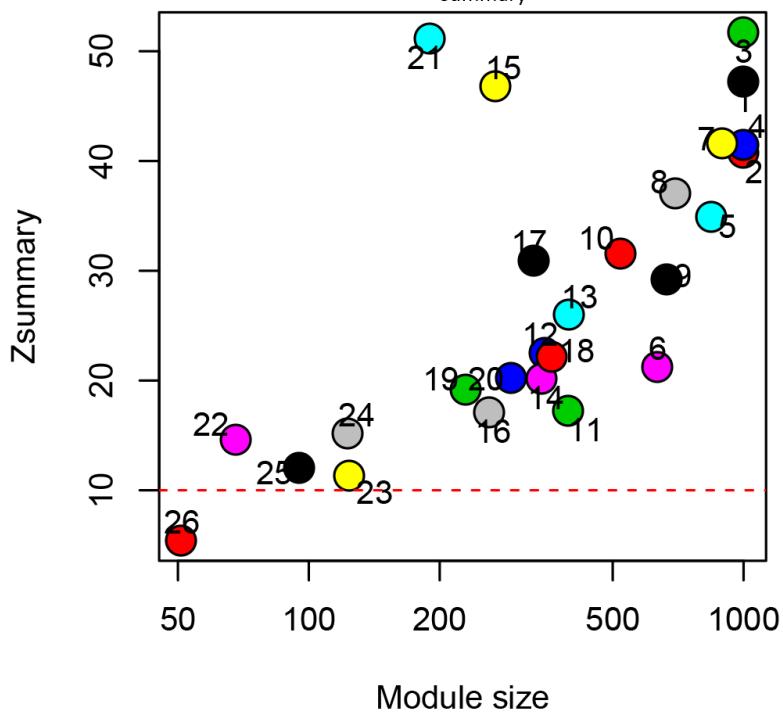




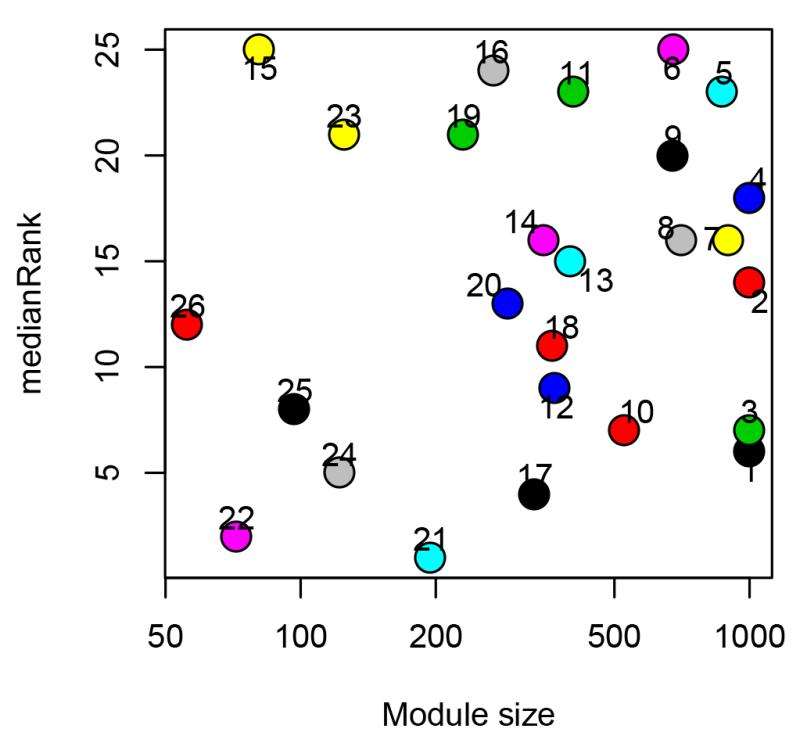
A-subgenome medianRank preservation



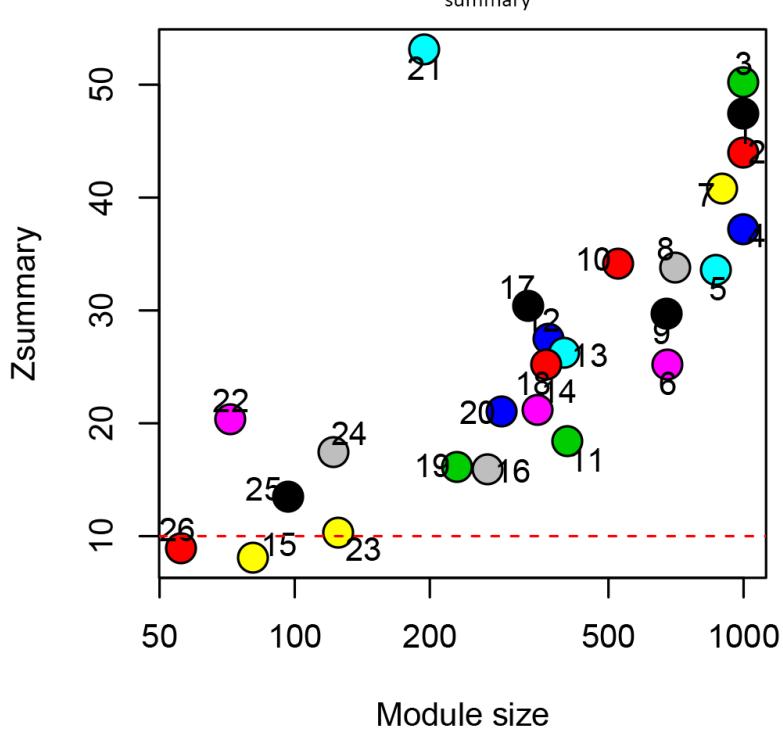
A-subgenome Z_{summary} Preservation



D-subgenome medianRank preservation

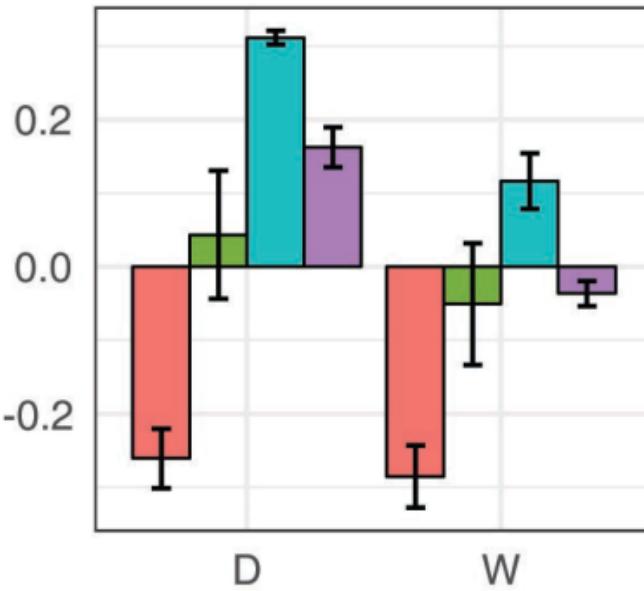


D-subgenome Z_{summary} Preservation



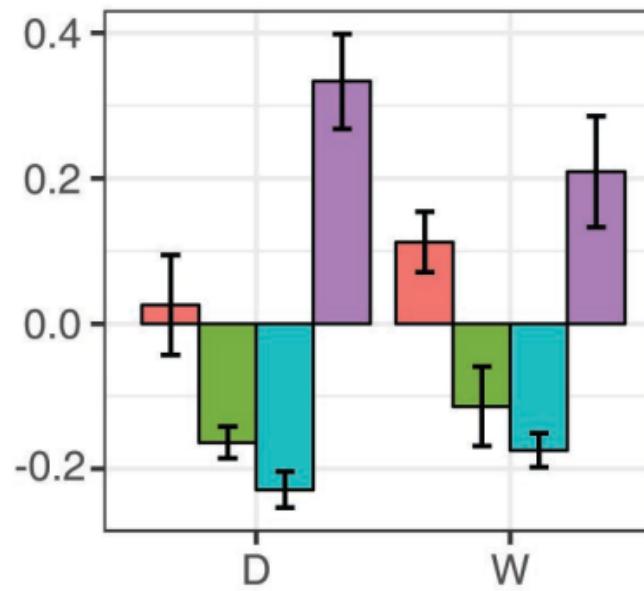
GO Terms

ME6



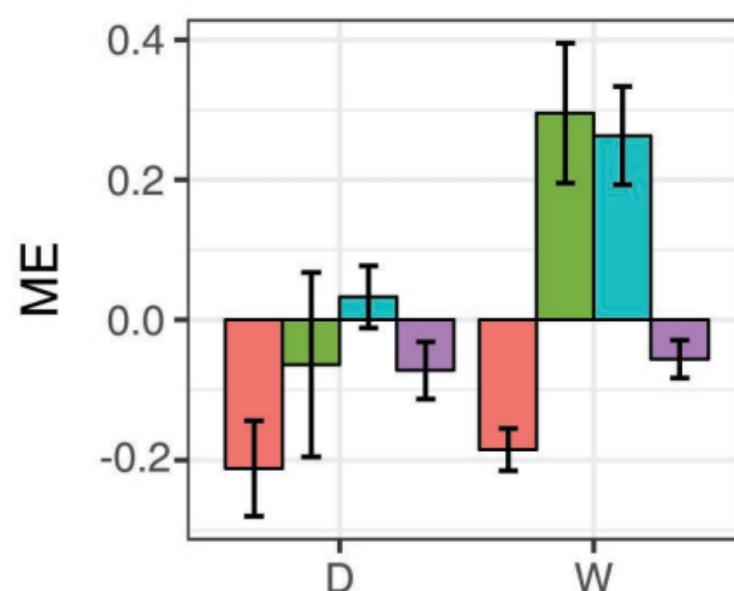
Cell wall modification
Sucrose metabolism
Meristem structure

ME8



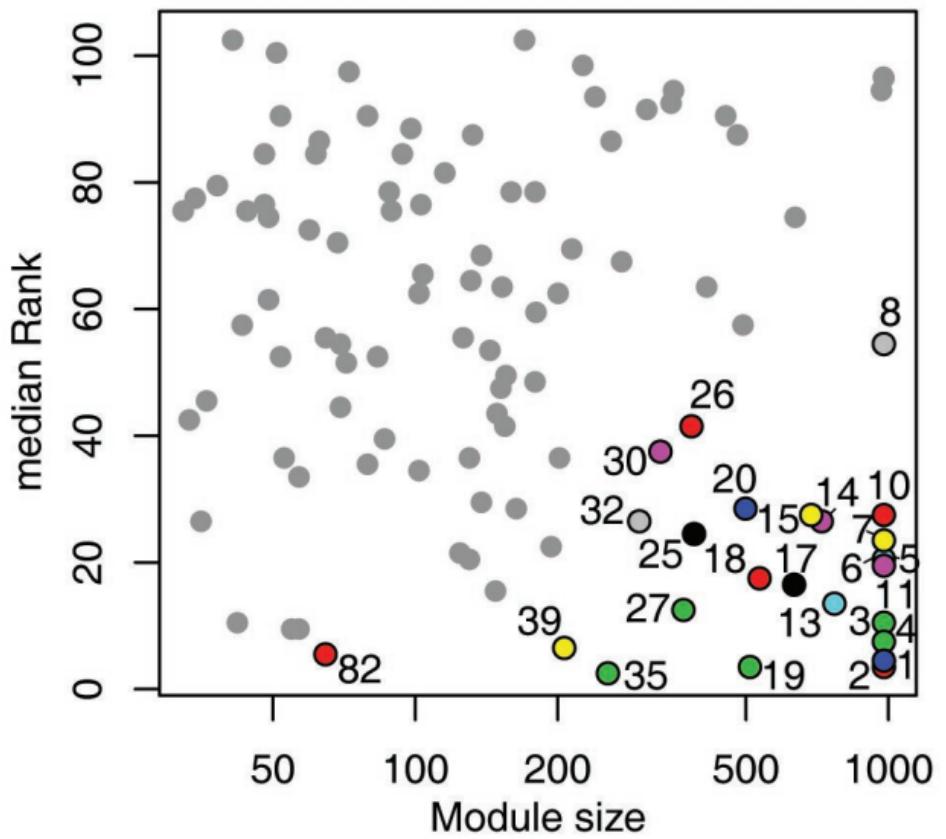
Cell macromolecule
catabolic processes
Cellulose biosynthesis

ME41



Cell wall modification
Sucrose metabolism
Meristem structure

**Domesticated
medianRank Preservation**



**Domesticated
Zsummary Preservation**

