Comparative transcriptome analysis between inbred and hybrids reveals molecular insights into yield heterosis of upland cotton

Kashif Shahzad

Chinese Academy of Agricultural Sciences Cotton Research Institute <https://orcid.org/0000-0001-5495-7105>

Xuexian Zhang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Liping Guo

Chinese Academy of Agricultural Sciences Cotton Research Institute

Tingxiang Qi

Chinese Academy of Agricultural Sciences Cotton Research Institute

Lisheng Bao

Jinhua Department of Economic Specialty Technology Promotion China

Meng Zhang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Bingbing Zhang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Hailin Wang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Huini Tang

Chinese Academy of Agricultural Sciences Cotton Research Institute

Xiuqin Qiao

Chinese Academy of Agricultural Sciences Cotton Research Institute

Juanjuan Feng

Chinese Academy of Agricultural Sciences Cotton Research Institute

Jianyong Wu (✉ dr.wujianyong@live.cn)

<https://orcid.org/0000-0003-1392-4790>

Chaozhu Xing

Chinese Academy of Agricultural Sciences Cotton Research Institute

---

Research article

Keywords: Cotton, Heterosis, RNA-Seq, Transcriptome, DEGs, Yield, Expression level dominance, QTLs, Co-expression

Posted Date: February 26th, 2020

DOI: <https://doi.org/10.21203/rs.2.24472/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on May 27th, 2020. See the published version at <https://doi.org/10.1186/s12870-020-02442-z>.

Abstract

Background: Utilization of heterosis has greatly enhanced the productivity of many crops worldwide. Understanding the potential molecular mechanism about how hybridization in cotton produces superior yield is critical for efficient plant breeding.

Results: With the whole-genome RNA sequencing, here, high, medium, and low hybrids varying in level of yield heterosis were screened based on different years and locations field experimentation. Phenotypically, high Department showed a mean of 14% more seed cotton yield than its better parent. A total of 63 samples comprised of different squaring stage tissues of three hybrids and four their inbred parents were used to perform transcriptomic analysis. A comparison of transcriptomic differences in each hybrid parent triad revealed a higher percentage of differentially expressed genes (DEGs) in each tissue. Expression level dominance analysis exposed the majority of hybrids DEGs followed parent like expressions. Functional annotations identified an array of DEGs involved in ATP and protein binding, membrane, cell wall, mitochondrion, and protein phosphorylation. Starch and sucrose metabolism and plant hormone signal transduction pathways were most enriched in each hybrid. Further, these two pathways had most mapped DEGs on known seed cotton yield QTLs. Integration of transcriptome, QTLs, and gene co-expression network analysis revealed genes *GhBZR1*, *GhASK8*, *At3g43860*, *GhGBSS1*, *GhAPL2*, *GhMPK4*, *GhPHO1*, *GhJAZ10*, and *GhCRR21* displayed a complex regulatory network of many interconnected genes. qRT-PCR of these DEGs was performed to ensure the accuracy of RNA-Seq data.

Conclusions: Through genome-wide comparative transcriptome analysis, the current study provides novel insights about phenomics and genomics of heterosis in upland cotton. Our results and data resources will be useful for dissecting the molecular mechanism of yield heterosis in cotton.

Background

Cotton is derived from the Arabic word 'quotr' [1], belongs to the genus is *Gossypium*, and has almost 50 species with diploid ($2n = 26$) and tetraploid ($2n = 52$) levels [2]. Among them, upland cotton (*G. hirsutum*) is allotetraploid, referred to as new world cotton, accounts for more than 90% of the world cotton production [3], and has an exceptional yield, early maturity, and moderately good fiber qualities. It is planted commercially for agricultural and industrial drives in the tropical and temperate region of the world on an area of 32–34 million hectares with an annual total production of 25.65 million metric tons [4]. Based on the average of the last three years, cotton is one of the few major commodities whose global production more or less matches the world mill consumption. Staggered yield potential of recent genotypes and climate change is major constraint. The breeders should mitigate these problems by developing varieties not only with superior yield and fiber quality but also with resistance against major pests, diseases and abiotic stresses such as adverse climatic conditions. The commercialization of hybrid cotton in China started around 1980 and the planting area increased in subsequent years with the development of hybrid Bt cotton [5, 6].

Utilization of heterosis proficiently increased the quantity and quality of crops. Heterosis is a phenomenon in which offspring produce more superior characters than their parents [7, 8]. In the last century, EM East [9] distinguished that crosses between different allotetraploids in the same genus of *Nicotiana Tabacum* formed exceptional heterosis. It was also experimentally observed by C Darwin [10] but Gorge. H Shull first time used the term heterosis in plant breeding [11]. After the introduction of hybrids, yield was raised effectively in many crops. Although cotton is an allopolyploid, it has more than two sets of basic chromosomes. Still, meaningful heterosis for different traits has been perceived in many filed experiments [12–15]. Cotton hybrid's main impacts include self-sufficiency, stability in production, improved fiber quality, generation of employment, foreign exchange earnings, and development of seed industry [5]. Hybrids in cotton are developed through utilization of heterosis in two ways: First, conventional method consists of emasculation and pollination with hands. Second is male-sterile system, which is an efficient method to reduce the cost of hand emasculation and ensure seed purity [16, 17].

The genetic basis of heterosis is perturbed and has been researched for almost a century using a variety of approaches, for instance, genetics [18, 19], molecular biology [20] omics [21], and physiological biochemistry [22]. Many researchers tried to explain crop heterosis with the so-called gene action hypothesis of dominance [23–25], overdominance [9, 26, 27], and epistasis [28, 29]. Genetic diversity between two parents and level of cotton heterosis is not simple and straightforward [30]. QTL mapping is advanced in 1990, makes some opportunities to understand individual QTLs, and interaction between them for heterosis. Genetic basis of heterosis is complicated, and involved dynamic dominance effect, epistasis and QTL by environmental interactions [31]. With the advancement in molecular research, many researchers concluded dominance, overdominance, and epistasis are basically conceptual and do not clarify the molecular mechanism and principals of heterosis [19, 32]. Meanwhile, it should be kept in mind that genetic models are equally important as the phenomenon of heterosis is a nonlinear effect from multiple heterozygous gene combinations. Moreover, yield traits are quantitative and many genetic characteristics function together to produce heterotic output, therefore a single genetic mechanism cannot explain genetic bases of heterosis in plants [33].

Considering the importance of F_1 heterosis in breeding, the genetic and molecular mechanism of heterosis has extensively been investigated in agronomic crops like rice [34] and maize [35]. However, advance research with high throughput sequencing has not performed yet in cotton. The one reason behind this is the late availability of the whole genome sequence of upland cotton to the researchers [36, 37]. Further, the allopolyploid cotton genome has a low level of molecular polymorphism as compared to other crops [38] and many genome duplications events occurred before and after polyploidization resulted in natural gene silencing, organ-specific and homoeologous biased gene expression in cotton [39]. Keeping in view the importance of heterosis and to accomplish the gap interlinked with molecular research in upland cotton, we designed a comprehensive study consists of field experiments and genome-wide transcriptome analysis between F_1 hybrids and their respective inbred parents. The following are the main objectives of this research: Selection of hybrids displayed better performance in

yield traits in diverse environments. Detection of DEGs, mode of gene expression, and biological pathways critical for yield heterosis. Identification of putative candidate genes and overview of their regulatory mechanism with co-expression network analysis.

Results

Identification of cotton hybrids exhibiting different level of heterosis

Our previous study of 11 inbred and 30 hybrids in six environments showed that cotton hybrids had better and stable performance compared with inbred parents in yield traits [40]. Here, analysis of heterosis revealed the majority of hybrids had positive mid parent heterosis (MPH) and better parent heterosis (BPH) for boll weight, seed cotton yield (SCY), lint yield and lint percentage (Table S1). Later on, three hybrids with different level of heterosis were identified and defined as high (H), medium (M) and low (L). Notably, H-hybrid (SJ48×Z98) produced 19.9% MPH and 14.1% BPH in SCY (Fig. 1). M-hybrid (SJ48×851) had MPH of 13.3% and BPH of 2% for SCY, whereas L-hybrid (SJ48×DT) showed the only MPH of 8.8% for SCY. These three hybrids had common maternal inbred parent but paternal inbred parents all differed. H-hybrid is cross of maternal line SJ48 (A) with paternal parent Z98 (B) and M-hybrid with paternal parent 851 (C), whereas L-hybrid is cross of same maternal line with paternal parent DT (D). The differences in the level of yield heterosis suggest that these three hybrids together with their inbred parents are suitable for studying comparative transcriptome analysis and regulatory mechanism of yield heterosis in cotton.

Transcriptome profiles of 63 RNA libraries of cotton hybrids and their inbred parents at squaring stage

To understand global transcriptome profiles of yield heterosis in cotton, three contrasting hybrids and their four inbred parents were used to perform RNA sequencing at the initiation of squaring stage. Leaf (Hereafter RL), flower buds (F), and one day post-anthesis (1DPA) ovule tissues with three biological replications were used for each genotype. In total, 63 libraries (7×3×3) were constructed for deep Illumina pair-end RNA sequencing. A total of ~43-50 million reads were generated per library (Table 1). The average percentage of the valid read was $\geq 98.5\%$. The value of Q30% was above 95% in this sequencing. Approximately, 94.9% of clean reads were mapped to the reference *G. hirsutum* genome [36]. Almost 90.5% reads were mapped to exon region in each sample and genotype. Wherein, intron and intergenic region mapping were $\leq 5\%$ (Fig. S1). The brief detail of total mapped, unique mapped, multi mapped, non-splice, and splice reads for each library can be seen in Table 1. The principal component analysis revealed two main clusters of all samples based on the two-first PCs (Fig. S2). RL replicates of most genotypes were found in one clustered with less variation, whereas tissues of F and 1DPA had distinct clusters. Pearson correlation test for all replicate and genotypes exposed that RL tissue had a strong correlation among each other (≥ 75) and weak with F and 1DPA (≤ 50) (Fig. S3). Correlation between F and IDPA samples was relatively strong (~55-60).

Global transcriptome changes for cotton hybrids and their inbred parents

The number of total expressed genes provides an overview of the transcriptomic landscape for all datasets. A total of ~60000 genes were expressed in each dataset (Fig. S4). Total expressed genes were much higher in flower buds compared to leaf and 1DPA ovule, while 1DPA ovule showed fewer total expressed genes compared with the other two tissues. Gene expression levels calculated as FPKM were used to analyze the dynamic changes of transcriptomes between hybrids and their respective parents. The expression levels significantly different at $p \leq 0.05$ and with $\log_2$ (fold change) >1 or <-1 designed as differentially expressed genes (DEGs). Here, the comparison of transcriptomes was performed for each hybrid parent triad and tissue. Total number of DEGs (Up + down) and their distribution among H-hybrid parent triad is shown in Fig. 2. Comparative analysis among H and its maternal parents (A) showed a total of 1121 DEGs in leaf (RL), 2073 in flower buds (F), and 1202 in 1DPA ovule, while compared with paternal parent (B), 1129, 1514, and 1805 DEGs were respectively identified in RL, F, and 1DPA (Fig. 2a). Combination of both parents A and B displayed higher number of DEGs (~2100 in all tissues) relative to H-hybrid. Distribution of DEGs found major portion were common in hybrid parent triad, whereas less were unique in each tissue (Fig. 2b, c, d). For example, 367 DEGs in RL, 775 in F, and 473 in 1DPA were unique in the combination of A with H. Unique DEGs in comparison of B and H were 358 in RL 577 in F, and 624 in 1DPA.

The results of DEGs in RL, F, and 1DPA transcriptomes of M-hybrid parent triad revealed that comparison of M with its maternal parent (A) had 1049, 1116, and 1207 DEGs respectively (Fig. 3a). Compared with paternal parent (C), this displayed 1518 DEGs in RL, 1471 in F, and 1196 in 1DPA. Combination of both parents A and C revealed a total of 1258, 1395, and 1576 differential gene expressions in RL, F, and 1DPA respectively. Distribution of unique and common DEGs identified more commonality in each tissue (Fig. 3b, c, d). Fewer DEGs were unique in A versus M such as 365 in RL, 405 in F, and 466 in 1DPA, while these were 626 in RL, 511 in F, and 376 in 1DPA in combination of C versus M. In context of commonality, both A and C compared with M had 362 overlap DEGs in RL, 375 in F, and 305 in 1DPA. Comparative analysis of DEGs for L-hybrid parent triad is summarized in Fig. 4. The result revealed that L versus maternal parent (A) had 1112 total DEGs in RL, 935 in F, and 3528 in 1DPA (Fig. 4a). The comparison of L and paternal parent (D) showed 1524, 790, and 4118 total number of DEGs in RL, F, 1DPA, respectively. Comparisons among parents A and D as compared to L displayed higher number of DEGs in RL (1626) and F (1415) and much less in 1DPA (2027). Distribution of DEGs revealed that majority of genes were overlapped but less were unique in each tissue (Fig. 4b, c, d). A versus L had 346, 299, and 1101 unique DEGs in RL, F, and 1DPA, respectively. Unique DEGs in D versus L were 649 in RL, 256 in F, and 1357 in 1DPA. In the context of commonality, comparison of L with both A and D showed 304, 181, and 1963 shared DEGs in RL, F, and 1DPA, respectively. At squaring stage, comparative transcriptome analysis in each hybrid parent triad determined that the percentage of genetic differential expressions in hybrids was similar to those of between parents.

Analysis of expression level dominance for hybrids

Expression level dominance is a phenomenon in which expression of hybrid DEGs is statistically similar to that of one parent, maybe additive or transgressive. To identify the expression magnitude and

directionality in interspecific F_1 hybrid cotton, DEGs were divided into 12 possible groups as described in material and methods following the conventions of [41] (Fig. 5a). Results showed that the magnitude of expression of most genes in hybrids was similar to that of one parent (Fig. 5). For instance, most of the genes in leaf and 1DPA ovule of H-hybrid displayed higher-maternal dominance expression and lower-maternal dominance expression respectively (Fig. 5b; Fig. S5a, c). Lower-paternal dominance expression had highest proportion in flower buds (Fig. 5b; Fig. S5b). Non-additive expressed genes in leaf and 1DPA ovule of M-hybrid mostly displayed higher-maternal dominance expression (Fig. 5c; Fig. S6a, c), wherein lower-maternal dominance expression was highest in flower buds (Fig. 5c; Fig. S6b). In the case of L-hybrid, most of leaf genes showed higher-maternal dominance expression (Fig. 5d; Fig. S7a), while lower-paternal dominance expression had highest portion in flower buds (Fig. 5d; Fig. S7b). In parent like expression groups, most of the genes in 1DPA ovule displayed higher maternal dominance expression in this hybrid (Fig. 5d; Fig. S7c). Accordance with analysis of expression level dominance, most DEGs of hybrids displayed parent expression level dominance (P-ELD) at squaring stage in different tissues that probably play role for yield heterosis of cotton.

Functional annotation and pathway enrichment analysis of DEGs with P-ELD

To identify the function of DEGs with P-ELD, a total of four gene sets (3-6 groups) were pooled to perform GO and KEGG enrichment analysis for every hybrid and tissue (Table S2, S3). GO enrichment analysis showed that majority of genes in leaf of H-hybrid had functional annotation related to biological process, integral component of membrane, ATP and protein binding (Fig. 6a). Most abundant functional gene groups in flower buds were plasma membrane, integral component of membrane and ATP binding, wherein nucleus, plasma membrane, and protein binding genes had higher significance in 1DPA ovule (Fig. 6a). Genes associated with chloroplast, extracellular region, and ATP binding had highest portion in leaf of M-hybrid (Fig. 6b). In flower buds, most of the genes were linked with biological process, molecular function and integral component of membrane. Most abundant functional gene groups in 1DPA ovule were biological process, chloroplast, and mitochondrion (Fig. 6b). Most enriched GO terms in leaf of L-hybrid were ATP binding, serine/threonine kinase activity (Fig. 6c). Functional annotation related to biological process, integral component of membrane, and chloroplast were abundant in flower bud, whereas gene involved in plasma membrane, regulation of transcription, and mitochondrion had highest portion in 1DPA ovule of L-hybrid (Fig. 6c).

KEGG enrichment analysis of P-ELD genes exposed that in leaf of H-hybrid starch and sucrose metabolism, endocytosis, and tryptophan metabolism had more significance (Fig. 7a). Majority of genes in flower buds were enriched in starch and sucrose metabolism, phagosome, and pentose and glucuronate interconversions. Whereas plant-pathogen interaction and plant hormone signal transduction were more representative pathways in 1DPA ovule (Fig. 7a). Leaf of M-hybrid showed more enrichment in metabolic pathways e.g. starch and sucrose, ascorbate and aldarate, and tryptophan (Fig. 7b). In flower buds, starch and sucrose metabolism, plant hormone signal transduction had more gene enrichment (Fig. 7b). Starch and sucrose metabolism and phosphatidylinositol signaling system was enriched and significant in 1DPA of this hybrid (Fig. 7b). Pathway categorization showed that L-hybrid had starch and

sucrose metabolism and ribosome significance in leaf (Fig. 7c), whereas starch and sucrose metabolism, plant hormone signal transduction, and ribosome were more representative pathways in flower buds (Fig. 7c). In 1DPA ovule, plant hormone signal transduction, plant-pathogen interaction, and starch and sucrose metabolism had more gene enrichment in this hybrid (Fig. 7c).

The common function of P-ELD genes in hybrids was integral component of membrane, biological process, mitochondrion, and ATP binding (Fig. S8a). Genes involved in common pathways such as starch and sucrose metabolism, plant hormone signal transduction, pentose and glucuronate interconversions, linoleic acid metabolism, and cutin, suberine and wax biosynthesis seems had a key role in hybrids growth during reproductive stage (Fig. S8b). Theoretically, genes from these pathways contribute to the yield of heterosis in hybrid cotton.

Mapping of key pathways DEGs on known seed cotton yield QTLs

QTL provide associations between genomics and phenomics for complex traits. Using QTLs can be an effective approach to understand the genetic complexities of yield heterosis. Here, we investigated the relationship between key pathway DEGs, seed cotton yield QTLs and heterosis. DEGs of high, medium, and low hybrid compared to their parents, and involved in key pathways such as starch and sucrose metabolism, plant hormone signal transduction, pentose and glucuronate interconversions, linoleic acid metabolism, and cutin, suberine and wax biosynthesis were used to map on already known 57 seed cotton yield (SCY)_QTLs (Table S4). Mapping results showed that 74 hybrid DEGs from these pathways were mapped on 43 QTLs (Fig. 8). Interestingly, most of the genes were mapped on QTLs regions that were reported more than once (Fig. 8). Out of these DEGs, 6 genes had differential expression in all hybrids compared to their parents, whereas 13 were common between any of two hybrids. 13 DEGs were specific to medium hybrid. However, 17 and 25 DEGs were specific to high and low hybrid respectively (Fig. S9)

Co-expression network analysis of high and low hybrids

WGCNA is an important tool for inferring the potential relationships between co-expressed genes in different samples or conditions. Construction of co-expression network helps to understand the functional linkage between gene groups and provides novel insight into the system-level understanding of a biological process. Here, co-expression network analysis was performed in high and low hybrids by using DEGs (Pooled from leaf, flower buds and 1DPA) identified in our RNA-seq data analysis. Highly connected genes were clustered into distinct modules. Cluster dendrogram displayed that interconnectivity and size of each module were quite different in high and low hybrids (Fig. S10a, b). Interestingly, 8862 genes distributed in 21 modules with the size ranging from 34 to 1972 genes was identified in high hybrid, wherein 7863 genes scattered in 33 modules with the module size ranging from 42 to 2880 was determined in low hybrid (Table S5). The correlation analysis between high and low hybrids modules was very strong (Fig. S10c). The average connectivity of genes in most of the modules was higher in high hybrids compared with low hybrid (Fig. S10d). Although turquoise and blue module

had the highest connected genes in both hybrids, but correlation coefficients among different genes were much higher in blue module.

Genes regulatory co-expression network for seed cotton yield heterosis

To construct the significant regulatory network for yield heterosis, first, we selected blue gene module in both hybrids. Secondly, the DEGs mapped on SCY-QTLs together with their interaction genes having weight ≥ 0.40 were screened (Table S6). Subsequently, our constructed gene co-expression network showed that nine SCY-QTLs genes had many interconnected genes involved in biological and molecular functions such as cell wall regulation or modification, carbohydrate metabolic processes, ion binding and transportation, biological regulation, methylation, endoplasmic reticulum, actin filament processing, catalytic activities, and protein binding or transportation (Fig. 9). We considered these nine genes as putative candidate genes for yield heterosis. Out of nine genes, two genes *Gh_A03G1024* (*BZR1*: Brassinazole-resistant 1) and *Gh_D08G1440* (*ASK8*: Shaggy-related protein kinase theta) involved in plant hormone signal transduction pathway showed differential expressions in flower buds of high hybrid. Another three starch and sucrose metabolism pathway also had differential expressions in flower buds of high hybrid. These genes were *Gh_A08G2210* (*At3g43860*: Endoglucanase 16), *Gh_A12G2183* (*GBSS1*: Granule-bound starch synthase 1, chloroplastic/amyloplastic), and *Gh_D07G1312* (*APL2*: Glucose-1-phosphate adenylyltransferase large subunit 2, chloroplastic). The remaining four had specific differential expression in 1DPA ovule of low hybrid. Three of them denoted as *Gh_D08G1467* (*MPK4*: Mitogen-activated protein kinase 4), *Gh_A03G0889* (*PHO1*: Phosphate transporter PHO1 homolog 3), and *Gh_A08G2199* (*JAZ10*: Jasmonate-zim-domain protein 10) were enriched in plant hormone signal pathway. The remained one gene *Gh_D05G0202* (*CRR21*: Pentatricopeptide repeat-containing protein At5g55740, chloroplastic) belonged to starch and sucrose metabolism pathway. Collectively, gene co-expression network analysis in hybrids not only helped to screen key genes but also provides novel insights into the regulatory mechanism of yield heterosis of cotton.

Quantitative real-time PCR validation of RNA-Seq data

Nine DEGs were selected to perform real-time qRT-PCR analysis. These genes showed hybrid and tissue-specific expression in RNA-seq, mapped on SCY-QTLs, and had the highest interconnected genes in the co-expression network. Primer was designed in gene-specific way to ensure accuracy of expression (Table S7). Analysis results showed *Gh_A03G1024*, and *Gh_D08G1440*, *Gh_A12G2183*, *Gh_A03G0889*, *Gh_D05G0202*, *Gh_A08G2199*, *Gh_D08G1467* and *Gh_D07G1312* displayed a significant change in each hybrid compared from their one parent or both parents (Fig. S11). However, gene *Gh_A08G2210* showed no statistical difference in hybrids. More interestingly, these genes apparently showed the expression similar to their maternal or paternal parent as determined in RNA-seq results.

Discussion

Hybridization leads to significant improvement in cotton yield

The utilization of heterosis has long been one of the main objectives of cotton breeders. Both intraspecific and interspecific cotton crosses showed meaningful heterosis for yield and yield-related traits [12, 15, 42–44]. Our field tests at six different environments revealed that more than 70% of intraspecific hybrids of upland cotton showed positive mid parent heterosis for boll weight, seed cotton yield (SCY), lint yield and lint percentage in each environment. Phenotypically, selected high hybrid had 20% mid and 14% better parent heterosis for SCY. However, medium hybrid showed 13% mid and 2% better parent heterosis for SCY. In contrast, low hybrid showed only mid parent heterosis of 8% for SCY. A study with five diverse *Gossypium barbadense* cotton genotypes and their reciprocal progenies reported that F_1 and F_2 out-yielded the parent group by 20–40% and 6–10%, respectively [45]. Previously, a mid-parent heterosis of 20% in F_1 and 10% in F_2 were found for lint yield [46]. However, others reported upland cultivars displayed about 34% higher mid parent heterosis than modern ones in lint yield [47]. The level of heterosis in any desired trait depends on hybridized parents and breeding design. Selection of genetically superior and physiological effective parents already improved maize yield up to 15% in the USA [32] and 10–20% rice in China [48].

Cotton hybrids exhibit parent expression level dominance at squaring stage

To better understand the genetics behind yield heterosis, transcriptome sequencing at the start of the reproductive stage was performed using three contrasting yield hybrids and their four inbred parents. Comparative analysis of DEGs in all tissues of each hybrid parent triad identified that the number of total DEGs in hybrids was similar to those of between parents. Few unique genomic differences existed in hybrids compared to parents at the gene expression level. Transcriptional reprogramming following hybridization even with few quantitative or qualitative differences in gene expression is expected to produce phenotypic differences in hybrids. Research in many agronomic crops has confirmed that hybridization has a dramatic effect on gene expression [49–52]. Previously, Comparative analysis between hybrids and their parents reported only 0.8–2.3% DEGs in *B. napus* [52] and 10.6% in super hybrid rice LYP9 [53]. Another study in rice noticed only 2.8% DEGs in hybrid [54]. These studies enabled us to understand only a small number of genetic differences can produce superior performance in hybrids.

Allopolyploids have been found to exhibit expression level dominance (ELD), a phenomenon in which gene expression of progeny is statistically similar to that of one parent [55]. Our analysis results identified that most hybrid DEGs showed non-additive gene expression at squaring stage especially high/low maternal or paternal like expression groups had highest portion. DEGs with additive or mid parent expressions were quite low. Despite this result, these hybrids and inbred lines have higher number of non-additive but overdominant genes at seedling [56]. Based on transcriptomes analysis, super hybrid rice showed 30–50% non-additive gene expression at the early developmental stages [53]. Many maize

scientists noticed higher non-additive gene expression in hybrids [57–59] while others detected higher additive gene expression [60, 61]. The predictions about additive genes involved in heterosis are considered because in many cases, the number of DEGs between parental lines was more than those between parental lines and F₁ hybrids. Plenty of recent studies have implemented ELD classification to check gene expression patterns in hybrids. For instance, 70–80% of the non-additive genes in *B. napus* F₁ hybrid displayed high parent expression level dominance during the early flowering development stage [52]. Most of hybrid DEGs in wheat followed the expression of maternal parent in seedling tissues and paternal parent in spike tissues [62]. Overdominance or transgressive expressions in heterotic hybrids compared to non-heterotic hybrids have also been reported in some crop hybrids [63, 64]. Based on the results of many studies, it is still unclear which expression class is important, but it seems both additive and non-additive gene expressions probably produced heterosis.

Possible roles of signaling and sugar metabolic pathways genes in yield heterosis

Our investigation found DEGs of hybrids with P-ELD were enriched plenty of diverse pathways. However, the plant hormone signaling and starch and sucrose metabolisms were the most keys and common pathways among the different tissues of hybrids. Probably the function of these pathways' genes was associated with the integral component of plasma membrane, biological processes, mitochondrion, protein phosphorylation, and ATP binding. Changes in the expression of signaling genes related to Auxin response factors (ARF), Leucine-rich repeat protein, Jasmonate-zim (JAZ) domain protein, MAP kinase, and Brassinosteroid signaling regulator (BZR) proteins were found in this study. Plant hormones are a structurally distinct group of key molecules that regulate plant growth and control feedback linked with both biotic and abiotic stresses. Phytohormone signaling pathway genes linked with brassinosteroids and gibberellic acid play critical role in plant height regulation of hybrid maize [65]. Further, transcriptome comparison in *B. napus* specified that IAA and SA response genes had differential expression in F₁ hybrids as compared to both parents which finally triggered hybrid vigor [52]. Auxin control transcription through IAA proteins and ARFs. IAA proteins bind with the ARF and repress transcription with the help of co-repressor called topless [66]. Transcription is either inhibited or activated depending upon the kind of an ARF which binds DNA [67]. Although plant hormone signaling and responses usually make major changes in transcription however, these are not as well categorized as the transcription responses in plants [68]. At the advent of abnormal conditions or stresses, plants generate diverse sugar and hormone signaling to maintain the balance of metabolic processes by changes at the transcript, protein, and metabolite level [69]. Earlier transcript analysis revealed sugar and auxin signaling regulate anther development during high-temperature stress in cotton [70].

To a particular interest, many sugar metabolic genes associated with Cellulose synthase (CESAs), Pectin lyase like (PEL), Sucrose synthase (Sus), Starch synthase, Alpha-beta-amylase, Pentatricopeptide repeat (PPR) and Tetratricopeptide repeat (TPR) protein had differential expression in hybrids. Cellulose and pectin are key components of plant cell walls and play role during cotton fiber cell development. Cellulose synthesis control fiber secondary wall thickening [71]. Reduced PEL gene transcripts caused a reduction in their enzymatic activity which ultimately reduced fiber elongation in cotton [72]. Mutants of cellulose

synthase 6 (CESA6) showed strong cellulose deficit as a consequence short hypocotyl phenotype was witnessed in Arabidopsis [73]. A genome-wide association study in *B. napus* determined CESA6 can be a promising candidate gene for stem lodging resistance [74]. Photosynthate in plants is transported in the form of sucrose and regulated by Sus enzymes through active participation in phloem unloading process. Sus activity associated genes perform a pivotal role in cotton fiber development by providing sucrose and cellulose to growing fiber [75, 76]. Additionally, Sus activity has strong connection with the final harvest index of starch storing organs in the plant.

Starch is a carbon storage polymer and acts as a major regulator of plant growth through balancing net carbon availability. The carbon fixed by plants is stored as starch in the light which is completely utilized in the night for respiration and growth [77]. Thus, any change that leads to a temporary state of carbon shortage, reduce the rate of plant growth by rebalancing the net carbon [78, 79] through enhancing the rate of starch synthesis and reducing the rate of starch breakdown [80, 81]. Gene expression study in hybrid rice has shown that several genes linked with circadian rhythm, carbon fixation, and starch and sucrose metabolism were located into yield QTL regions, expecting these putative candidate genes for yield heterosis [82]. More to this, besides photosynthesis, other metabolic pathways such as sucrose and starch possibly a key contribution to wheat heterosis [83]. Maize seed and potato tuber contain plenty of starch on dry weight basis, so their yield mainly depends on starch and sucrose metabolism [84, 85]. Starchless mutants cannot grow in a light/dark cycle due to an imbalance of carbon and starch which ultimately produced carbon shortage at night resulting in growth impedance for quite a few hours the next day [78].

Role of QTLs and co-expression network to excavate candidate genes for yield heterosis

Quantitative trait locus (QTLs) provides an association between genomics and phenomics of complex traits. In the last decade, several researchers have performed QTLs analysis to understand heterosis [86]. However, difficult to point out target genes in most cases due to the large region of the identified QTLs. To minimize this challenge, a promising approach is mapping DEGs on already known QTLs and then the construction of their co-expression regulatory network. Further, yield heterosis is a reflection of the effect of many interconnected genes rather than individual genes, thus understanding the interrelationship between genes facilitates to exploit candidate genes that are likely associated with a biological system. A similar approach already has been done in rice and *B. napus* [54, 74]. Our selected hybrids were superior or inferior to their parents in seed cotton yield, and investigation found 74 DEGs from key pathways were mapped on 43 overlapped seed cotton yield QTLs.

Our integrated co-expression network analysis of high and low hybrids using green module revealed nine seed cotton yield QTLs mapped genes, interconnected with many genes enriched in biological and molecular functions such as carbohydrate metabolic processes, catalytic activities, and protein binding or transportation. Our identified putative candidate genes were GhBZR1, GhASK8, At3g43860, GhGBSS1, GhAPL2, GhMPK4, GhPHO1, GhJAZ10, and GhCRR21. More interestingly, all these genes showed tissue-specific differential expression in reproductive tissue flower buds and 1DPA ovule of hybrids. Gene BZR1

is associated with brassinosteroid mediated signaling pathway and regulates cell elongation and division. BES1 transgenic plants in upland cotton showed reduced hypocotyl sensitivity to brassinazole, curled leaves, bent petioles and changed expression of BR constitutive genes [87]. The suppression of rice OsBZR1 caused straight leaves, dwarfism, reduced BR response and increased BR insensitivity. ASK8 is a Shaggy-related gene that enables extracellular signals to adjust transcription in differentiating cells, stress responses [88], and also participate in biological functions linked with BR response. The gene At3g43860 belongs to endoglucanase gene family and involved in cell wall organization and cellulose catabolism. In Arabidopsis, it was predicted that it may have a function in pollen and pollen tube growth [89].

Other putative candidate genes associated with starch biosynthesis such as GBSS1 and APL2 showed differential expression in flower buds of high hybrid. Knockdown or mutation of GBSS1 or APL2 effect reproductive growth as a result shrunken seeds were produced in rice [90]. Additionally, promising candidate signaling genes associated with JAZ10 and MPK4 protein showed differential expression in only 1DPA ovule of low hybrid. MPK4 modifies the expression of genes responding to biotic and abiotic stresses and also has an important role in pathogen defense [91]. More to this perform functions in the regulation of cytokine [92], salicylic acid, and jasmonic acid-mediated defense gene expression. JAZ10 gene acts as repressor of jasmonate responses and Jasmonoyl isoleucine specifically promotes COI1-TIFY10A/JAZ1 interaction as observed in Arabidopsis [93]. Gene PHO1 thought to be involved in the transportation of inorganic phosphate [94], here in our results showed downregulation in 1DPA ovule of low hybrid. Chloroplast CRR21 genes play the main role in chloroplast RNA editing and modifications showed downregulation in fiber ovule of low hybrid only. To be concise, the integration of transcriptomic, QTLs and co-expression network provide useful understanding about the genomics of seed cotton yield heterosis.

Conclusions

Notably, F_1 hybrids of upland showed meaningful mid and better heterosis in seed cotton yield. Comparative transcriptome study of contesting hybrids and their inbred parents at squaring stage revealed many DEGs with parent like gene expressions. In addition, many auxins, brassinosteroid, jasmonic acid hormone signaling and cellulose, sucrose, and starch synthase metabolic DEGs were mapped on known seed cotton yield QTLs. Co-expression network analysis identified nine promising candidate genes and discovered their complex biological network with interconnected genes that probably mediate seed cotton yield heterosis. Together all, our field experimentation quantified that hybridization in cotton is useful to improve yield. Comprehensive genome-wide transcriptome analysis gave new insights to understand the preliminary genetics of yield heterosis. However, further research at the gene functional level is desired to understand the perturbed biological system of yield heterosis in upland cotton.

Methods

Calculation of phenotypic heterosis for yield parameter

In 2015, our research group produced 30 intraspecific F_1 upland cotton hybrids by using 11 inbred parents. The brief detail about all plant material and two-year field experimentation at three locations can be seen in our previous study [95]. The phenotypic data from all field experimentation were used to determine the degree of heterosis in yield traits, as mid parent heterosis (MPH), was calculated using the equation: $MPH = 100 \times (F_1 - MP)/MP$, and that of better parent heterosis (BPH) was calculated using the equation: $BPH = 100 \times (F_1 - BP)/BP$, where F_1 is the performance value of each trait for the F_1 progeny, MP is the mean value of that trait for the parents, and BP is the value of that trait for the better parent

RNA extraction, Illumina sequencing, and data analysis

Based on the level of phenotypic heterosis in yield traits, high (denoted as H), medium (M), and low (L) hybrid and their four inbred parents (A, B, C, and D) were selected for transcriptomic analysis at initiation of squaring which is generally the start of reproductive stage. The detail about field experimentation is described in our published study [56]. For RNA samples, leaf, flower buds, and 1-day after post anthesis (DPA) ovule tissues in three biological replicates were collected for all seven materials from the field test of the year 2018. Young leaves about 2mm picked from different plants were pooled for the composite samples. Similarly, flower buds of 2mm, 3mm and 4mm size were also pooled. All harvesting samples were instantly placed in liquid nitrogen and stored at -80°C before use. A total of 63 samples were used to isolate total RNA using the Sigma Spectrum Plant Total RNA kit (Sigma-Aldrich, USA) according to the manufacturer's protocol. The total RNA quantity and purity were analyzed on Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA). After initial preparations, paired-end sequencing on an Illumina HiSeq 4000 was performed at the (LC-bio, China) following the vendor's recommended protocol. Cutadapt [96] and in-house Perl scripts were used for quality control. HISAT2 [97] was used to map reads to the upland cotton genome [36]. The mapped reads of each sample were assembled using StringTie [98]. After the final transcriptome was generated, StringTie together with Ballgown [99] was used to estimate the expression levels of all transcripts. Different expression analysis of mRNAs the StringTie was used to perform expression level for mRNAs by calculating FPKM. The differentially expressed mRNAs were selected with $\log_2$ (fold change) >1 or $\log_2$ (fold change) <-1 and with statistical significance (p -value < 0.05) by R package Ballgown. The expression level dominance analysis was performed as described by [55]. Briefly, to discover the distribution of expression statistics among the 12 possible groups of expression modes between two parents and their derived hybrids, we mapped the kernel density of expression for each class using the density estimator in the R software package. These were plotted on a standardized scale against the experimental mean to illustrate inter-specific comparisons. By this analysis, genes expression in hybrids may be additive (1 and 2), paternal-expression level dominance (3 and 4), maternal-expression level dominance (5 and 6), and transgressive expression lower or higher than either parent (7 to 12). GO enrichment and KEGG pathway enrichment analysis was performed on the DEGs using the Goatools software [100] and KOBAS software [101], with a corrected p -value ≤ 0.05 as the threshold and rich factor.

Mapping key DEGs on seed cotton yield QTLs

Seed cotton yield QTLs (LOD > 2) with genetic position and linked molecular markers were acquired from the cotton QTL database (www.cottonqtl.db.org). The physical positions of QTLs, gene loci, and coordinates were attained from the upland cotton genome [36]. Depend on the physical positions of both gene loci and QTL. DEGs from highly enriched and key pathways were then mapped to QTLs. Map chart software was used to display the physical location of genes and QTLs on concerned chromosomes.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was performed individually in high and low hybrids by using squaring stage DEGs. These DEGs were identified in comparative RNA-seq data analysis between these hybrids and their respective parents and pool from leaf, flower buds and 1DPA ovule tissues using each biological replicate as an individual dataset. To make sure that the network was biologically appropriate, the mean connectivity of the network and the scale-free topology model fit were estimated over a range of the soft threshold power β . The topological overlap-based difference measure was used to cluster all the coding sequences hierarchically as described by [102]. Further, the resulting gene dendrogram was used for module detection using the dynamic tree cut method (min Module Size = 100 and merge Cut Height = 0.25). In the WGCNA, if any two genes were connected, the topology overlap measure provided in WGCNA was used to determine edge weight. The weights ranged from 0 to 1 reflected the relationship between the two genes. Connectivity was defined as the sum of the weights across all the edges of a node. Genes that shared a consistent expression pattern were clustered into different modules depended on their correlation. Finally, the module having the highest connectivity from each hybrid was selected to construct a network. The co-expression network was built using only putative candidate genes (Genes enriched both in seed cotton yield QTLs and key pathways) and their interconnected genes having weights > 0.40. Cytoscape_v 3.7.1 software [103] was used to display the final co-expression network.

Quantitative real-time PCR (qRT-PCR) analysis

For, quantitative real-time RT-PCR, we used the same total RNA to generate the first standard cDNA that was used for RNA sequencing. First-strand cDNA was generated from 1 μ g total RNA from individual replications using a PrimerScript RT Reagent kit (Perfect Real Time, TaKaRa, Japan). Quantitative real-time RT-PCR was performed using SYBR® Premix Ex Taq™ (Perfect Real Time, TaKaRa, Japan) according to the manufacturer's instructions. Primers for qPCR were designed using the Primer Express software (Applied Biosystems, Foster City, CA, USA), synthesized commercially (Tianyi Huiyuan Biotechnology, Beijing, China). qRT-PCR and data analysis were performed as previously described [16, 17].

Abbreviations

DEGs

Differentially expressed genes; MPH:Mid parent heterosis; P-ELD:Parent expression level dominance; QTLs:Quantitative trait locus; WGCNA:Weighted gene co-expression network analysis

Declarations

Ethics approval and consent to participate

All the cotton lines used and analyzed were public and available for non-commercial purpose. This article did not contain any studies with human participants or animals performed by any of the authors.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Program of China (2016YFD0101400). The funding body had no role in the design, analysis, interpretation of data, and in writing the manuscript.

Availability of data and materials

The datasets generated or analyzed in this study are freely available online.

Authors' contributions

CZX and JYW designed the experiments. LPG, TXQ, HT, XQQ, LSB, and HLW constructed hybrid plant population and field evaluation. KS, XXZ MZ, JJF, and BBZ performed data analysis. KS, JYW, and CZX contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

The authors are thankful to Muhammad Yasir, Haris Ahmad, Zhiden Zuo, Ting Li, and Yongfeng Zhang (ICR-CAAS, China) for their assistance in this work.

References

1. Lee JA, Fang DD: **Cotton as a world crop: origin, history, and current status.** *Cotton* 2015(agronmonogr57):1-24.

2. Cronn RC, Small RL, Haselkorn T, Wendel JF: **Rapid diversification of the cotton genus (*Gossypium: Malvaceae*) revealed by analysis of sixteen nuclear and chloroplast genes.** *American journal of botany* 2002, **89**(4):707-725.
3. Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD: **Toward sequencing cotton (*Gossypium*) genomes.** *Plant physiology* 2007, **145**(4):1303-1310.
4. Abdurakhmonov IY, Ayubov MS, Ubaydullaeva KA, Buriev ZT, Shermatov SE, Ruziboev HS, Shapulatov UM, Saha S, Ulloa M, Yu JZ: **RNA interference for functional genomics and improvement of cotton (*Gossypium* sp.).** *Frontiers in plant science* 2016, **7**:202.
5. Xing C, Jing S, Xing Y: **Review and Prospect on Cotton Heterosis Utilization and Study in China [J].** *Cotton Science* 2007, **5**.
6. Dong H, Li W, Tang W, Zhang D: **Development of hybrid Bt cotton in China—a successful integration of transgenic technology and conventional techniques.** *Current science* 2004, **86**(6):778-782.
7. Birchler JA, Auger DL, Riddle NC: **In search of the molecular basis of heterosis.** *The Plant cell* 2003, **15**(10):2236-2239.
8. Chen ZJ: **Genomic and epigenetic insights into the molecular bases of heterosis.** *Nature Reviews Genetics* 2013, **14**(7):471.
9. East EM: **Heterosis.** *Genetics* 1936, **21**(4):375.
10. Darwin C: **The effects of cross and self fertilisation in the vegetable kingdom:** D. Appleton; 1877.
11. Shull GH: **The composition of a field of maize.** *Journal of Heredity* 1908(1):296-301.
12. Rauf S, Khan TM, Nazir S: **Combining ability and heterosis in *Gossypium hirsutum* L.** *Int J Agri Biol* 2005, **7**(1):109-113.
13. Hussain A, Rafique M, Mahmood T, Khan M, Khan M: **Hybrids vigour in relation to development in cotton (*G. hirsutum* L.).** *J Agric Res* 1990, **28**:205-217.
14. Ali I, Shakeel A, Saeed A, Nazeer W, Zia Z, Ahmad S, Mahmood K, Malik W: **Combining ability analysis and heterotic studies for within-boll yield components and fibre quality in cotton.** *The Journal of Animal and Plant Sciences* 2016, **26**(1):156-162.
15. Khan NU, Hassan G, Kumbhar MB, Kang S, Khan I, Parveen A: **Heterosis and inbreeding depression and mean performance in segregation generations in upland cotton.** *Editorial Advisory Board e* 2005, **17**(4):531-546.
16. Zhang B, Zhang X, Liu G, Guo L, Qi T, Zhang M, Li X, Wang H, Tang H, Qiao X: **A combined small RNA and transcriptome sequencing analysis reveal regulatory roles of miRNAs during anther development of Upland cotton carrying cytoplasmic male sterile *Gossypium harknessii* (D2) cytoplasm.** *BMC plant biology* 2018, **18**(1):242.
17. Wu J, Zhang M, Zhang B, Zhang X, Guo L, Qi T, Wang H, Zhang J, Xing C: **Genome-wide comparative transcriptome analysis of CMS-D2 and its maintainer and restorer lines in upland cotton.** *BMC genomics* 2017, **18**(1):454.

18. Yang M, Wang X, Ren D, Huang H, Xu M, He G, Deng XW: **Genomic architecture of biomass heterosis in Arabidopsis**. *Proceedings of the National Academy of Sciences* 2017, **114**(30):8101-8106.
19. Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q: **Genetic composition of yield heterosis in an elite rice hybrid**. *Proceedings of the National Academy of Sciences* 2012, **109**(39):15847-15852.
20. Zhu X, Zhang Y, Guo W, Zhang T-Z: **Relationships between differential gene expression and heterosis in cotton hybrids developed from the foundation parent CRI-12 and its pedigree-derived lines**. *Plant science* 2011, **180**(2):221-227.
21. Westhues M, Schrag TA, Heuer C, Thaller G, Utz HF, Schipprack W, Thiemann A, Seifert F, Ehret A, Schlereth A: **Omics-based hybrid prediction in maize**. *Theoretical and applied genetics* 2017, **130**(9):1927-1939.
22. Feng G, Luo H, Zhang Y, Gou L, Yao Y, Lin Y, Zhang W: **Relationship between plant canopy characteristics and photosynthetic productivity in diverse cultivars of cotton (*Gossypium hirsutum* L.)**. *The Crop Journal* 2016, **4**(6):499-508.
23. Xiao J, Li J, Yuan L, Tanksley SD: **Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers**. *Genetics* 1995, **140**(2):745-754.
24. Stuber CW, Lincoln SE, Wolff D, Helentjaris T, Lander E: **Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers**. *Genetics* 1992, **132**(3):823-839.
25. Davenport CB: **Degeneration, albinism and inbreeding**. *Science* 1908, **28**(718):454-455.
26. Li Z-K, Luo L, Mei H, Wang D, Shu Q, Tabien R, Zhong D, Ying C, Stansel J, Khush G: **Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield**. *Genetics* 2001, **158**(4):1737-1753.
27. Tian M, Nie Q, Li Z, Zhang J, Liu Y, Long Y, Wang Z, Wang G, Liu R: **Transcriptomic analysis reveals overdominance playing a critical role in nicotine heterosis in *Nicotiana tabacum* L.** *BMC plant biology* 2018, **18**(1):48.
28. Yu S, Li J, Xu C, Tan Y, Gao Y, Li X, Zhang Q, Maroof MS: **Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid**. *Proceedings of the National Academy of Sciences* 1997, **94**(17):9226-9231.
29. Williams W: **Heterosis and the genetics of complex characters**. *Nature* 1959, **184**(4685):527.
30. Li X, Shahzad K, Guo L, Qi T, Zhang X, Wang H, Tang H, Qiao X, Zhang J, Wu J: **Using yield quantitative trait locus targeted SSR markers to study the relationship between genetic distance and yield heterosis in upland cotton (*Gossypium hirsutum*)**. *Plant Breeding* 2019, **138**(1):105-113.
31. Shang L, Wang Y, Cai S, Wang X, Li Y, Abduweli A, Hua J: **Partial dominance, overdominance, epistasis and QTL by environment interactions contribute to heterosis in two upland cotton hybrids. G3: Genes, Genomes, Genetics** 2016, **6**(3):499-507.
32. Duvick DN: **Biotechnology in the 1930s: the development of hybrid maize**. *Nature Reviews Genetics* 2001, **2**(1):69.

33. Schnable PS, Springer NM: **Progress toward understanding heterosis in crop plants.** *Annual review of plant biology* 2013, **64**:71-88.
34. He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F: **Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids.** *The Plant Cell* 2010, **22**(1):17-33.
35. Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP: **Repeat associated small RNAs vary among parents and following hybridization in maize.** *Proceedings of the National Academy of Sciences* 2012, **109**(26):10444-10449.
36. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM: **Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement.** *Nature biotechnology* 2015, **33**(5):531.
37. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J: **Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution.** *Nature biotechnology* 2015, **33**(5):524.
38. Abdurakhmonov I, Kohel R, Yu J, Pepper A, Abdullaev A, Kushanov F, Salakhutdinov I, Buriev Z, Saha S, Scheffler B: **Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm.** *Genomics* 2008, **92**(6):478-487.
39. Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF: **Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*).** *Genetics* 2009, **182**(2):503-517.
40. Shahzad K, Qi T, Guo L, Tang H, Zhang X, Wang H, Qiao X, Zhang M, Zhang B, Feng JJA: **Adaptability and Stability Comparisons of Inbred and Hybrid Cotton in Yield and Fiber Quality Traits.** 2019, **9**(9):516.
41. Yoo M, Szadkowski E, Wendel JJH: **Homoeolog expression bias and expression level dominance in allopolyploid cotton.** 2013, **110**(2):171.
42. Mukhtar MS, Khan TM: **Estimation of heterosis for yield and yield components in cotton (*Gossypium hirsutum* L.).** *Pakistan J Biol Sci* 2000, **3**:1412-1413.
43. Zhang J, Wu M, Yu J, Li X, Pei W: **Breeding potential of introgression lines developed from interspecific crossing between upland cotton (*Gossypium hirsutum*) and *Gossypium barbadense*: heterosis, combining ability and genetic effects.** *PLoS One* 2016, **11**(1):e0143646.
44. Zhang J, Abdelraheem A: **Combining ability, heterosis, and genetic distance among nine elite American Pima cotton genotypes (*Gossypium barbadense*).** *Euphytica* 2017, **213**(11):240.
45. Zhang J, Abdelraheem A, Wu J: **Heterosis, combining ability and genetic effect, and relationship with genetic distance based on a diallel of hybrids from five diverse *Gossypium barbadense* cotton genotypes.** *Euphytica* 2017, **213**(9):208.
46. Meredith WR: **Yield and fiber-quality potential for second-generation cotton hybrids.** *Crop science* 1990, **30**(5):1045-1048.

47. Campbell B, Bowman D, Weaver D: **Heterotic effects in topcrosses of modern and obsolete cotton cultivars.** *Crop science* 2008, **48**(2):593-600.
48. Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY: **Progress in research and development on hybrid rice: a super-domesticated in China.** *Annals of botany* 2007, **100**(5):959-966.
49. Song G, Guo Z, Liu Z, Cheng Q, Qu X, Chen R, Jiang D, Liu C, Wang W, Sun Y: **Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F₁ hybrids.** *BMC plant biology* 2013, **13**(1):221.
50. Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J: **mRNA and small RNA transcriptomes reveal insights into dynamic homeolog regulation of allopolyploid heterosis in nascent hexaploid wheat.** *The Plant Cell* 2014, **26**(5):1878-1900.
51. Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM: **Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis.** *BMC plant biology* 2008, **8**(1):33.
52. Shen Y, Sun S, Hua S, Shen E, Ye CY, Cai D, Timko MP, Zhu QH, Fan L: **Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs.** *The Plant Journal* 2017, **91**(5):874-893.
53. Wei G, Tao Y, Liu G, Chen C, Luo R, Xia H, Gan Q, Zeng H, Lu Z, Han Y: **A transcriptomic analysis of superhybrid rice LYP9 and its parents.** *Proceedings of the National Academy of Sciences* 2009, **106**(19):7695-7701.
54. Song G-S, Zhai H-L, Peng Y-G, Zhang L, Wei G, Chen X-Y, Xiao Y-G, Wang L, Chen Y-J, Wu B: **Comparative transcriptional profiling and preliminary study on heterosis mechanism of super-hybrid rice.** *Molecular plant* 2010, **3**(6):1012-1025.
55. Rapp RA, Udall JA, Wendel JF: **Genomic expression dominance in allopolyploids.** *BMC biology* 2009, **7**(1):18.
56. Shahzad K, Zhang X, Guo L, Qi T, Tang H, Zhang M, Zhang B, Wang H, Qiao X, Feng J *et al*: **Comparative transcriptome analysis of inbred lines and contrasting hybrids reveals overdominance mediate early biomass vigor in hybrid cotton.** *BMC Genomics* 2020, **21**(1):140.
57. Auger DL, Gray AD, Ream TS, Kato A, Coe EH, Birchler JA: **Nonadditive gene expression in diploid and triploid hybrids of maize.** *Genetics* 2005, **169**(1):389-397.
58. Użarowska A, Keller B, Piepho H-P, Schwarz G, Ingvaridsen C, Wenzel G, Lübberstedt T: **Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height.** *Plant molecular biology* 2007, **63**(1):21-34.
59. Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho H-P, Hochholdinger F: **Comparison of maize (*Zea mays* L.) F₁-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends.** *Genetics* 2008, **179**(3):1275-1283.
60. Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS: **All possible modes of gene action are observed in a global comparison of gene expression in a maize F₁ hybrid and its inbred parents.** *Proceedings of the National Academy of Sciences* 2006, **103**(18):6805-6810.

61. Meyer S, Pospisil H, Scholten S: **Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern.** *Plant molecular biology* 2007, **63**(3):381-391.
62. Liu Y, Gao S, Tang Y, Gong J, Zhang X, Wang Y, Zhang L, Sun R, Zhang Q, Chen Z: **Transcriptome analysis of wheat seedling and spike tissues in the hybrid Jingmai 8 uncovered genes involved in heterosis.** *Planta* 2018, **247**(6):1307-1321.
63. Hedgecock D, Lin J-Z, DeCola S, Haudenschield CD, Meyer E, Manahan DT, Bowen B: **Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*).** *Proceedings of the National Academy of Sciences* 2007, **104**(7):2313-2318.
64. Li D, Zeng R, Li Y, Zhao M, Chao J, Li Y, Wang K, Zhu L, Tian W-M, Liang C: **Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F₁ hybrids.** *Scientific reports* 2016, **6**:24984.
65. Wang H, Gu L, Zhang X, Liu M, Jiang H, Cai R, Zhao Y, Cheng B: **Global transcriptome and weighted gene co-expression network analyses reveal hybrid-specific modules and candidate genes related to plant height development in maize.** *Plant molecular biology* 2018, **98**(3):187-203.
66. Szemenyei H, Hannon M, Long JA: **TOPELESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis.** *Science* 2008, **319**(5868):1384-1386.
67. Guilfoyle TJ, Hagen G: **Auxin response factors.** *Current opinion in plant biology* 2007, **10**(5):453-460.
68. Santner A, Calderon-Villalobos LIA, Estelle M: **Plant hormones are versatile chemical regulators of plant growth.** *Nature chemical biology* 2009, **5**(5):301.
69. Suzuki N, Koussevitzky S, Mittler R, Miller G: **ROS and redox signalling in the response of plants to abiotic stress.** *Plant, Cell & Environment* 2012, **35**(2):259-270.
70. Min L, Li Y, Hu Q, Zhu L, Gao W, Wu Y, Ding Y, Liu S, Yang X, Zhang X: **Sugar and auxin signaling pathways respond to high-temperature stress during anther development as revealed by transcript profiling analysis in cotton.** *Plant physiology* 2014, **164**(3):1293-1308.
71. Haigler CH, Betancur L, Stiff MR, Tuttle JRJFips: **Cotton fiber: a powerful single-cell model for cell wall and cellulose research.** 2012, **3**:104.
72. Wang H, Guo Y, Lv F, Zhu H, Wu S, Jiang Y, Li F, Zhou B, Guo W, Zhang TJPmb: **The essential role of GhPEL gene, encoding a pectate lyase, in cell wall loosening by depolymerization of the de-esterified pectin during fiber elongation in cotton.** 2010, **72**(4-5):397-406.
73. Hématy K, Sado P-E, Van Tuinen A, Rochange S, Desnos T, Balzergue S, Pelletier S, Renou J-P, Höfte H: **A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis.** *Curr Biol* 2007, **17**(11):922-931.
74. Li H, Cheng X, Zhang L, Hu J, Zhang F, Chen B, Xu K, Gao G, Li H, Li L: **An integration of genome-wide association study and gene co-expression network analysis identifies candidate genes of stem lodging-related traits in *Brassica napus*.** *Frontiers in plant science* 2018, **9**:796.

75. Haigler CH, Ivanova-Datcheva M, Hogan PS, Salnikov VV, Hwang S, Martin K, Delmer DP: **Carbon partitioning to cellulose synthesis.** In: *Plant Cell Walls*. Springer; 2001: 29-51.
76. Ruan Y-L, Llewellyn DJ, Furbank RT: **Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development.** *The Plant Cell* 2003, **15**(4):952-964.
77. Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC: **Starch as a major integrator in the regulation of plant growth.** *Proceedings of the National Academy of Sciences* 2009, **106**(25):10348-10353.
78. Gibon Y, Bläsing OE, Palacios-Rojas N, Pankovic D, Hendriks JH, Fisahn J, Höhne M, Günther M, Stitt M: **Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the following light period.** *The Plant Journal* 2004, **39**(6):847-862.
79. Geiger DR, Servaites JC, Fuchs MA: **Role of starch in carbon translocation and partitioning at the plant level.** *Functional Plant Biology* 2000, **27**(6):571-582.
80. Smith AM, Stitt M: **Coordination of carbon supply and plant growth.** *Plant, cell & environment* 2007, **30**(9):1126-1149.
81. Stitt M, Gibon Y, Lunn JE, Piques M: **Multilevel genomics analysis of carbon signalling during low carbon availability: coordinating the supply and utilisation of carbon in a fluctuating environment.** *Functional Plant Biology* 2007, **34**(6):526-549.
82. Li D, Huang Z, Song S, Xin Y, Mao D, Lv Q, Zhou M, Tian D, Tang M, Wu Q: **Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase.** *Proceedings of the National Academy of Sciences* 2016, **113**(41):E6026-E6035.
83. Yao Y, Ni Z, Zhang Y, Chen Y, Ding Y, Han Z, Liu Z, Sun Q: **Identification of differentially expressed genes in leaf and root between wheat hybrid and its parental inbreds using PCR-based cDNA subtraction.** *Plant molecular biology* 2005, **58**(3):367-384.
84. Liu N, Xue Y, Guo Z, Li W, Tang J: **Genome-wide association study identifies candidate genes for starch content regulation in maize kernels.** *Frontiers in plant science* 2016, **7**:1046.
85. Fernie AR, Willmitzer L, Trethaway RN: **Sucrose to starch: a transition in molecular plant physiology.** *Trends in plant science* 2002, **7**(1):35-41.
86. Garcia AAF, Wang S, Melchinger AE, Zeng Z-B: **Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice.** *Genetics* 2008, **180**(3):1707-1724.
87. Liu Z, Qanmber G, Lu L, Qin W, Liu J, Li J, Ma S, Yang Z, Yang Z: **Genome-wide analysis of BES1 genes in *Gossypium* revealed their evolutionary conserved roles in brassinosteroid signaling.** *Science China Life Sciences* 2018, **61**(12):1566-1582.
88. Tavares R, Vidal J, Van Lammeren A, Kreis M: **AtSK0, a plant homologue of SGG/GSK-3 marks developing tissues in *Arabidopsis thaliana*.** *Plant molecular biology* 2002, **50**(2):261-271.
89. Xie X-J, Huang J-J, Gao H-H, Guo G-Q: **Expression patterns of two *Arabidopsis* endo- β -1, 4-glucanase genes (At3g43860, At4g39000) in reproductive development.** *Molecular Biology* 2011, **45**(3):458-465.

90. Pérez L, Soto E, Farré G, Juanos J, Villorbina G, Bassie L, Medina V, Serrato AJ, Sahrawy M, Rojas JA: **CRISPR/Cas9 mutations in the rice Waxy/GBSSI gene induce allele-specific and zygosity-dependent feedback effects on endosperm starch biosynthesis.** *Plant cell reports* 2019, **38**(3):417-433.
91. Andreasson E, Jenkins T, Brodersen P, Thorgrimsen S, Petersen NH, Zhu S, Qiu JL, Micheelsen P, Rocher A, Petersen M: **The MAP kinase substrate MKS1 is a regulator of plant defense responses.** *The EMBO journal* 2005, **24**(14):2579-2589.
92. Kosetsu K, Matsunaga S, Nakagami H, Colcombet J, Sasabe M, Soyano T, Takahashi Y, Hirt H, Machida Y: **The MAP kinase MPK4 is required for cytokinesis in *Arabidopsis thaliana*.** *The Plant Cell* 2010, **22**(11):3778-3790.
93. Chini A, Fonseca S, Fernandez G, Adie B, Chico J, Lorenzo O, Garcia-Casado G, López-Vidriero I, Lozano F, Ponce M: **The JAZ family of repressors is the missing link in jasmonate signalling.** *Nature* 2007, **448**(7154):666.
94. Wang Y, Ribot C, Rezzonico E, Poirier Y: **Structure and expression profile of the Arabidopsis PHO1 gene family indicates a broad role in inorganic phosphate homeostasis.** *Plant physiology* 2004, **135**(1):400-411.
95. Shahzad K, Li X, Qi T, Guo L, Tang H, Zhang X, Wang H, Zhang M, Zhang B, Qiao X *et al*: **Genetic analysis of yield and fiber quality traits in upland cotton (*Gossypium hirsutum* L.) cultivated in different ecological regions of China.** *Journal of Cotton Research* 2019, **2**(1):14.
96. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet journal* 2011, **17**(1):10-12.
97. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nat Methods* 2015, **12**(4):357.
98. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nature biotechnology* 2015, **33**(3):290.
99. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT: **Ballgown bridges the gap between transcriptome assembly and expression analysis.** *Nature biotechnology* 2015, **33**(3):243.
100. Klopfenstein D, Zhang L, Pedersen BS, Ramírez F, Vesztröcy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M: **GOATOOLS: A Python library for Gene Ontology analyses.** *Scientific reports* 2018, **8**(1):10872.
101. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L: **KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.** *Nucleic acids research* 2011, **39**(suppl_2):W316-W322.
102. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Statistical applications in genetics and molecular biology* 2005, **4**(1).
103. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome research* 2003, **13**(11):2498-2504.

Table 1

Table 1. Summary of RNA sequencing and mapping for all 63 samples

Sample	Total (base)	Raw	Valid (base)	Valid (reads)	ratio	Q30%	Mapped (%)	Unique Mapped (%)	Multi (%)	Mapped	Non-splice (%)	Splice (%)
ARL1	5.94G		5.87G	98.8		94.4	95.2	62.9	32.3		52.1	31.0
ARL2	5.35G		5.30G	99.0		94.5	95.3	61.8	33.5		51.7	30.7
ARL3	7.79G		7.71G	99.0		94.6	95.6	62.0	33.6		51.2	31.2
BRL1	5.48G		5.43G	99.1		94.6	95.5	61.5	34.0		51.0	31.2
BRL2	5.33G		5.27G	98.9		95.0	95.4	61.3	34.1		51.2	30.8
BRL3	7.88G		7.77G	98.6		96.2	95.6	61.0	34.6		51.0	31.2
CRL1	6.00G		5.94G	99.0		93.7	95.0	59.1	36.0		51.2	31.0
CRL2	5.36G		5.31G	99.1		93.8	95.0	59.0	36.0		51.4	30.7
CRL3	6.80G		6.72G	98.9		93.5	94.9	59.1	35.9		50.7	31.2
DRL1	6.00G		5.93G	98.9		94.4	95.2	59.8	35.4		51.2	31.3
DRL2	6.84G		6.74G	98.5		95.5	95.4	60.9	34.6		50.9	31.8
DRL3	5.52G		5.45G	98.8		94.4	95.2	59.5	35.7		50.5	31.6
HRL1	5.77G		5.70G	98.9		93.6	94.7	58.3	36.4		50.2	31.4
HRL2	5.95G		5.87G	98.8		94.7	95.4	59.9	35.6		50.4	31.9
HRL3	6.88G		6.79G	98.8		93.4	94.9	58.8	36.1		50.5	31.0
MRL1	5.72G		5.58G	97.5		93.7	94.3	59.5	34.8		50.6	31.0
MRL2	7.92G		7.83G	98.8		93.9	95.4	60.3	35.1		51.2	31.0
MRL3	6.62G		6.52G	98.5		94.9	95.5	60.3	35.2		51.4	30.9
LRL1	7.93G		7.84G	98.9		93.8	95.1	60.0	35.1		51.1	30.9
LRL2	7.31G		7.24G	98.9		92.8	94.7	58.5	36.2		51.6	30.6
LRL3	5.79G		5.72G	98.8		93.3	95.1	59.4	35.7		51.6	30.9
AF1	6.30G		6.20G	98.5		94.8	95.3	60.5	34.8		52.2	30.8
AF2	7.75G		7.64G	98.6		94.0	94.5	62.0	32.6		51.5	30.9
AF3	6.02G		5.95G	98.9		94.0	95.0	61.8	33.3		52.4	30.6
BF1	7.41G		7.28G	98.2		94.2	95.3	63.2	32.1		52.0	31.0
BF2	6.08G		5.99G	98.5		94.2	95.3	63.2	32.1		52.3	31.0
BF3	7.91G		7.79G	98.5		94.5	94.0	62.3	31.7		51.2	30.8
CF1	7.02G		6.90G	98.3		94.4	88.9	58.6	30.3		48.8	28.4
CF2	5.91G		5.83G	98.6		94.1	95.1	62.5	32.7		51.6	31.5
CF3	5.96G		5.88G	98.6		94.0	95.1	62.6	32.5		51.9	31.2
DF1	7.60G		7.53G	99.1		94.0	95.3	62.3	32.9		51.9	31.0
DF2	7.07G		7.00G	99.1		93.8	94.7	61.6	33.1		52.5	30.5
DF3	7.56G		7.49G	99.0		94.4	95.6	63.7	31.8		50.9	32.8
HF1	5.99G		5.93G	99.0		93.4	92.3	60.5	31.7		51.0	29.8
HF2	7.17G		7.10G	99.0		94.1	94.7	62.4	32.3		51.7	31.5
HF3	5.59G		5.53G	99.0		94.4	95.5	63.0	32.4		51.9	32.1
MF1	7.44G		7.35G	98.8		94.3	95.6	63.9	31.7		52.4	31.7
MF2	5.98G		5.92G	98.9		94.2	92.3	59.8	32.4		50.4	30.8
MF3	5.80G		5.75G	99.1		94.1	95.2	62.1	33.1		52.4	31.5
LF1	5.76G		5.70G	99.1		94.3	95.3	62.3	33.0		51.5	32.0
LF2	6.14G		6.09G	99.2		93.8	93.4	59.9	33.5		51.3	30.5
LF3	6.45G		6.40G	99.1		93.9	94.5	61.2	33.2		51.8	31.3
DPAA1	7.80G		7.72G	99.0		94.0	94.1	61.2	32.9		50.6	31.6
DPAA2	7.76G		7.68G	99.0		94.0	95.2	62.1	33.1		50.9	32.3
DPAA3	6.01G		5.96G	99.1		94.3	94.5	61.3	33.3		50.9	32.0
DPAB1	7.71G		7.63G	98.9		93.8	93.4	60.8	32.6		50.5	31.4
DPAB2	5.76G		5.70G	99.1		93.7	92.8	60.1	32.8		50.7	31.1
DPAB3	5.52G		5.40G	97.9		97.9	95.1	66.5	28.6		52.0	31.9
DPAC1	5.56G		5.47G	98.2		98.7	95.9	66.9	29.0		52.6	31.8
DPAC2	6.70G		6.58G	98.2		98.7	96.2	66.7	29.5		52.2	32.3

DPAC3	5.88G	5.78G	98.3	98.7	96.7	67.4	29.3	52.4	32.6
DPAD1	6.67G	6.52G	97.7	98.7	92.8	65.0	27.8	50.0	31.4
DPAD2	5.53G	5.22G	94.5	97.8	93.1	65.0	28.1	50.0	31.6
DPAD3	6.69G	6.56G	98.1	98.7	95.5	66.7	28.8	50.9	32.7
DPAH1	7.67G	7.55G	98.4	98.6	96.5	67.2	29.3	52.4	33.0
DPAH2	7.81G	7.67G	98.2	98.7	96.5	67.2	29.4	51.4	33.5
DPAH3	7.66G	7.54G	98.5	98.5	93.7	64.5	29.2	50.2	32.2
DPAM1	5.77G	5.67G	98.3	98.5	96.2	66.5	29.7	52.4	32.2
DPAM2	6.76G	6.64G	98.2	98.8	96.6	67.8	28.8	52.6	32.5
DPAM3	6.22G	6.11G	98.1	98.3	95.6	65.7	29.8	52.1	31.8
DPAL1	7.62G	7.48G	98.2	98.8	96.5	67.1	29.4	52.3	32.6
DPAL2	7.88G	7.68G	97.4	98.6	96.5	66.7	29.7	52.4	31.1
DPAL3	5.61G	5.47G	97.6	98.5	95.1	65.1	30.0	51.3	30.4

Figures

Figure 1

Level of mean seed cotton yield heterosis in high, medium, and low hybrids. SJ48-1xZ98-15 SJ48-1x851-2 and SJ48-1xDT-8 correspond to high, medium, and low hybrids, respectively. Mean seed cotton yield was calculated from two years and three location field experimentation.

Figure 2

Total DEGs and their distribution in high-hybrid parent triad at squaring. (a) shows total number of DEGs in the hybrid parent triad. (b), (c), and (d) respectively represent distribution of unique and common DEGs in leaf (RL), flower buds (F) and 1DPA ovule in high-hybrid parent triad. Maternal parent is denoted with A, paternal parent with B, and high hybrid with H in each figure.

Figure 3

Total DEGs and their distribution in medium-hybrid parent triad at squaring. (a) shows total number of DEGs in hybrid parent triad. (b), (c), and (d) respectively represent distribution of unique and common DEGs in leaf (RL), flower buds (F) and 1DPA ovule in medium-hybrid parent triad. Maternal parent is denoted with A, paternal parent with C, and medium hybrid with M in each figure.

Figure 4

Total DEGs and their distribution in low-hybrid parent triad at squaring. (a) shows total number of DEGs in hybrid parent triad. (b), (c), and (d) respectively represent distribution of unique and common DEGs in leaf, flower buds and 1DPA ovule in low-hybrid parent triad. Maternal parent is denoted with A, paternal parent with D, and low hybrid with L in each figure.

Figure 5

The gene expression groups in each F1 hybrids at squaring. (a) Expression patterns of 12 groups, M: paternal parent, F1: Hybrid, and F: maternal parent. (b), (c), and (d) indicate total number of genes in each group of high (H), medium (M), and low (L) hybrids, respectively.

Figure 6

Most enriched GO terms for DEGs with parents like expression in hybrids at squaring. (a), (b) and (c) shows GO terms with total number of genes in high (H), medium (M) and low (L) hybrid, respectively. Here, most enriched GO terms with $p < 0.05$ are only presented in each figure.

Figure 7

Most enriched pathways for DEGs with parents like expression in hybrids at squaring. (a), (b), and (c) represent pathway analysis in leaf, flower buds and 1DPA ovule, respectively.

Figure 8

Distribution of hybrids DEGs with parents like expression on known seed cotton yield QTLs. Green, blue, and pink colors respectively indicate high, medium, and low hybrid-specific DEGs. Red and black colors show common DEGs among all and any of two hybrids, respectively. QTLs region reported more than once is in green color.

Figure 9

The co-expression network of known seed cotton yields QTLs mapped nine genes. Red and blue nodes respectively represent high and low hybrid-specific genes. Gene to gene correlation weight is above 0.40 in this network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3.xlsx](#)
- [Fig.S8.png](#)
- [TableS4.docx](#)
- [TableS1.docx](#)
- [TableS7.docx](#)
- [TableS5.docx](#)
- [TableS6.xlsx](#)
- [TableS2.xlsx](#)
- [Fig.S11.png](#)
- [Fig.S3Pearsoncorrelationbetweensample.png](#)
- [Fig.S6.png](#)
- [Fig.S4Totalnumberofexpresssedgenes.png](#)
- [Fig.S2Principlecomponantanalysis.png](#)
- [Fig.S9.png](#)
- [Fig.S1Statisticsofmappingregions.png](#)
- [Fig.S5.png](#)
- [Fig.S7.png](#)
- [Fig.S10.png](#)