



Homoeolog gene expression analysis reveals novel expression biases in upland hybrid cotton under intraspecific hybridization

Kashif Shahzad¹ · Xuexian Zhang^{1,2} · Meng Zhang^{1,2} · Liping Guo^{1,2} · Tingxiang Qi^{1,2} · Huini Tang^{1,2} · Hailin Wang¹ · Iqra Mubeen³ · Xiuqin Qiao¹ · Renhai Peng^{2,4} · Jianyong Wu^{1,2} · Chaozhu Xing^{1,2}

Published online: 30 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Hybridization is useful to enhance the yield potential of agronomic crops in the world. Cotton has genome doubling due to the allotetraploid process and hybridization in coordination with duplicated genome can produce more yield and adaptability. Therefore, the expression of homoeologous gene pairs between hybrids and inbred parents is vital to characterize the genetic source of heterosis in cotton. Investigation results of homoeolog gene pairs between two contrasting hybrids and their respective inbred parents identified 36853 homoeolog genes in hybrids. It was observed both high and low hybrids had similar trends in homoeolog gene expression patterns in each tissue under study. An average of 96% of homoeolog genes had no biased expression and their expressions were derived from the equal contribution of both parents. Besides, very few homoeolog genes (an average of 1%) showed no biased or novel expression in both hybrids. The functional analysis described secondary metabolic pathways had a majority of novel biased homoeolog genes in hybrids. These results contribute preliminary knowledge about how hybridization affects expression patterns of homoeolog gene pairs in upland cotton hybrids. Our study also highlights the functional genomics of metabolic genes to explore the genetic mechanism of heterosis in cotton.

Keywords Cotton · Heterosis · Homoeolog expression · RNA-Seq · Secondary metabolism

Introduction

Kashif Shahzad, Xuexian Zhang, and Meng Zhang contributed equally to this work.

- ✉ Renhai Peng
aydxprh@163.com
- ✉ Jianyong Wu
dr.wujianyong@live.cn
- ✉ Chaozhu Xing
Chaozhuxing@126.com

¹ State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Key Laboratory for Cotton Genetic Improvement, Ministry of Agriculture and Rural Affairs, 38 Huanghe Dadao, Anyang 455000, Henan, China

² Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang 455000, Henan, China

³ Department of Biochemistry, University of Agriculture, Faisalabad 38040, Pakistan

⁴ College of Biology and Food Engineering, Anyang Institute of Technology, Anyang 455000, Henan, China

Hybridization or polyploidy is a vital process, which produces exceptional phenotypes through the interaction of two distinct genomes, and has a substantial influence on the plant genome (Dufresne and Hebert 1994; Yoo et al. 2013). The novel phenotypes in plants can be a consequence of many modifications not only at genomic and epigenetic levels but also in gene expression patterns (Doyle et al. 2008; Kashkush et al. 2003). Several examples in cotton (Flagel and Wendel 2010), *Brassica napus* (Gaeta et al. 2009), and Spartina (Chelaifa et al. 2010) stated that hybridization (genome merger) modified more transcriptome than polyploidy. About 15% of flowering plants are polyploidy (Wood et al. 2009) and divided into categories of autopolyploid and allopolyploid. Autopolyploids evolved from a genome merger of the same or similar species but allopolyploids contain a genomic combination of two different species (Doyle et al. 2008). Allopolyploids have enhanced growth and wider adaptations relative to autopolyploids. Many important crops are allopolyploids together with upland cotton (*Gossypium hirsutum*). In allopolyploids, hybridization or polyploidy

is allied with various kinds of exchanges and interactions between homoeolog genes (duplicate genes) (Kovarik et al. 2008; Salmon et al. 2010; Szadkowski et al. n.d.). Further modifications are consisting of epigenetics (Doyle et al. 2008; Madlung et al. 2002), gene activation (Kashkush et al. 2003; Kraitshtein et al. 2010), and sequence removal (Ansour et al. 2009; Han et al. 2005). The result of different genomic interactions by hybridization or polyploidy produced several expression patterns of duplicate or genes with similar functions. For instance, these may be gene silencing, gene activation, novel expression, and altered parental contribution (Yoo et al. 2013; Li et al. 2014). However, the direction and magnitude of duplicate gene expression are varied with crop, tissue, and organ (Adams et al. 2003; Bottley et al. 2006; Bugs et al. 2010; Chaudhary et al. 2009).

Previous studies have identified four different expression forms of duplicated genes in plants, e.g., one duplicate gene copy may be lost its expression pattern as compared to other (De Smet et al. 2013), and duplicate gene copies may maintain original expression (McGrath et al. 2014), one duplicate gene copy may adopt new expression (Hughes et al. 2014; Liu and Adams 2010), and sometimes original expression is the equal contribution of duplicates (Kleinjan et al. 2008). Importantly, the duplicate gene pairs may show homoeolog expression bias in allopolyploid (Grover et al. 2012). Where one homoeolog displays preferential expression to the transcriptome as compared to the other. The phenomena of homoeolog expression have already been stated for allopolyploids like cotton (Flagel et al. 2008; Hovav et al. 2008; Samuel Yang et al. 2006), wheat (Bottley et al. 2006; Mochida et al. 2004), Arabidopsis (Chang et al. 2010; Wang et al. 2004), and brassica (Auger et al. 2009; Gaeta et al. 2007). The analysis results in cotton showed that hybridization caused more homoeolog expression changes and fewer changes in gene expressions (Yoo et al. 2013). In contrast, polyploidy formation is linked with more gene expression modifications and fewer changes in homoeolog expression. A study in cyprinid fish also reported hybridization brought more homoeolog expression changes than polyploidy (Ren et al. 2016). Homoeolog expression can be either balanced or unbalanced (Grover et al. 2012). When duplicate gene pairs display biased expression towards one parental genome is almost equal to the biased expression towards other parental genomes. It shows balanced homoeolog expression. On the other hand, an unbalanced homoeolog expression change states a situation where more duplicate gene pairs demonstrate expression twisted towards one parental genome. An unbalanced form of homoeolog expression is mostly reported in allopolyploids (Akhunova et al. 2010; Schnable and Freeling 2011; Wang et al. 2006).

Upland cotton is a prime fiber, an important source of edible oil, and cash crops in the world (Shahzad et al. 2019b). The demand for cotton fiber in the textile industry is

increasing. While supply is reducing due to low area of cultivation and staggered yield potential of cultivars. Heterosis is and will continue to be a basic breeding tool to produce superior genotypes (Birchler et al. 2003). Many researchers reported meaningful heterosis in seed cotton yield and fiber quality traits in upland cotton (Ali et al. 2016; Li et al. 2019). Furthermore, cotton genotypes produced through the utilization of heterosis exhibit more stability and adaptability to changing the environment (Shahzad et al. 2019a). The genetic mechanism of heterosis is a perturbed and so-called genetic model of dominance (Xiao et al. 1995), over dominance (Shahzad et al. 2020b), and epistasis (Yu et al. 1997) does not reveal it. However, progress has already been achieved in the recent few years with advanced molecular research. For instance, high-throughput transcriptome sequencing is a useful technology to get accurate knowledge about global gene expressions (Shahzad et al. 2020a; Wang and Gerstein 2009). This also creates a possibility to discriminate the expression of homologous genes (Wu et al. 2018). To date, many previous studies in agronomic crops analyzed patterns of homoeolog expression in allotetraploids and their diploid progenitors, e.g., cotton (Yoo et al. 2013) and brassica (Li et al. 2020). However, information regarding how hybridization effect homoeolog expression changes in newly formed upland cotton hybrids is limited. Thus, we conducted a study to investigate the effect of hybridization on homoeolog gene expression modifications on a genome-wide scale. The direction and extent of homoeolog gene expressions between contrasting F₁ hybrids and their reciprocal parents were analyzed by RNA-seq of leaf, root, and flower buds. The result will provide new insights into how homoeolog gene expressions contribute to growth heterosis in upland cotton and enhance our genetic knowledge for hybridization process.

Material and methods

Plant material and RNA sequencing

The plant material used in this study included two seed cotton yield contrasting F₁ hybrids of upland cotton and their three reciprocal inbred parents named A, B, and D. The high yielding hybrid (H) was a cross of inbred A and B, whereas a low hybrid was a combination of cross A and D. The seedling was raised in the lab for all these materials, root and leaf samples were picked at true leaf stage, and flower buds were picked at squaring stage from the field in three biological repeats. Total RNA was isolated using the Sigma Spectrum Plant Total RNA kit (Sigma-Aldrich, USA) as described by the manufacturing company. After quality checks and preparations, paired-end sequencing (300 ± 50 bp) on a second-generation Illumina Hiseq 4000 was performed for hybrid H

and L. These sequencing details can be seen at NCBI with accession number GSE144676 (Shahzad et al. 2020b) and GSE150052 (Shahzad et al. 2020a). The full-length Pacbio transcriptome sequencing (Unpublished) for parental A, B, and D was used for a better comparison of homoeolog gene expression patterns between parents and progeny.

Analysis of homoeolog expression changes and functional enrichment

To find homologous sequences between reciprocal parents of hybrid H and L, the full-length Pacbio sequences of the two parents were homoeologously aligned using the reciprocal BLAST (BLASTN) hit method with a screening index of e^{-20} (Ren et al. 2016). Two sequences were homologous, if each of them was the best hit of the other, and if the sequences were aligned over 300 bp. The total number of homoeolog gene pairs identified in parents A and B was 45216, while the parents A and D had 44920 homoeolog gene pairs. These homologous sequences were used as a reference for the already published second-generation Illumina sequencing H and L hybrid samples (Shahzad et al. 2020a, b). Each sample was subjected to mutation detection and specific SNPs. The hybrid offspring samples may be homozygous or heterozygous at the specific SNP sites of their respective parents. Later, we compared the SNPs of two parents to select the homoeologue-specific SNPs bins using custom python scripts (Li et al. 2020). Furthermore, genes with homoeologue-specific SNPs were compared on a case-by-case basis between the parents and then progeny to perform synchronous classification of homoeolog expression bias. The index to determine the significant difference was *p-value* < 0.05. Compared with the diploid progenitors, the homoeolog gene pairs between both parents of hybrids were divided into 12 main categories with three subcategories, including the parental condition, no bias, and novel bias in

progeny as already described (Yoo et al. 2013). Functional analyses including gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) were performed for all homoeolog expression genes that showed novel bias in different tissues of hybrids. GO terms and enriched pathways were retrieved from the cotton functional genomics database (<https://cottonfgd.org/analyz>) with *p-value* < 0.05.

Results

Homoeolog expression changes in different tissues of high hybrid

To observe the pattern of homoeolog expression in different tissues of high hybrid, a case-by-case gene comparison was performed for high hybrid (H) and its parents (maternal=A, paternal=B). The results are detailed in Table 1. Notably, a pattern of homoeolog expression was conserved in this hybrid for the greater number of genes in all tissues under investigation. Fewer homoeolog gene copies reflected a novel biased pattern. For example, 36853 total number of homoeolog gene pairs were identified in each tissue. We observed that 96% in the root, 97% in leaf, and 97% in flower buds had maintained parental expression patterns in hybrid. The total expression level of a homoeolog gene pair had no change or was equal to that of both parents. An average of 1% homoeolog gene pairs showed no bias in each hybrid tissue. According to the results of novel bias expression, 2.27% of homologous gene pairs had a novel bias in the root. Only 2.1% of homoeolog gene pairs in leaf exhibited novel bias expression. It was observed that 2.02% of homologous gene pairs had a novel bias in flower buds. The majority of novel biased homoeolog gene pairs exhibited maternal-like expression patterns in root and leaf (Fig. 1a–b), wherein parental-like novel expressions were observed in flower buds

Table 1 Homoeolog gene expression changes in different tissues of the high upland cotton hybrid

	Expression in parents	Expression in progeny	Root (%)	Leaf (%)	Flower buds (%)
A=B	A=B Parental condition	35471 (96.2)	35734 (97.1)	35779 (97.1)	
A<B	A<B Parental condition	4 (0.01)	0 (0.0)	1 (0.0)	
A>B	A>B Parental condition	5 (0.01)	7 (0.02)	4 (0.01)	
A>B	A=B No bias in progeny	190 (0.52)	279 (0.76)	210 (0.57)	
A<B	A=B No bias in progeny	345 (0.94)	66 (0.18)	107 (0.29)	
A=B	A>B Novel bias in progeny	416 (1.13)	375 (1.02)	391 (1.06)	
A<B	A>B Novel bias in progeny	20 (0.05)	10 (0.03)	12 (0.03)	
A>B	A<B Novel bias in progeny	10 (0.03)	18 (0.05)	10 (0.03)	
A=B	A<B Novel bias in progeny	392 (1.06)	364 (1.0)	339 (0.92)	
Overall A-biased in progeny		441 (1.20)	392 (1.06)	407 (1.10)	
Overall B-biased in progeny		406 (1.10)	382 (1.04)	350 (0.95)	
Total number of genes		36853	36853	36853	

Here in table, A is the maternal parent and B is the paternal parent

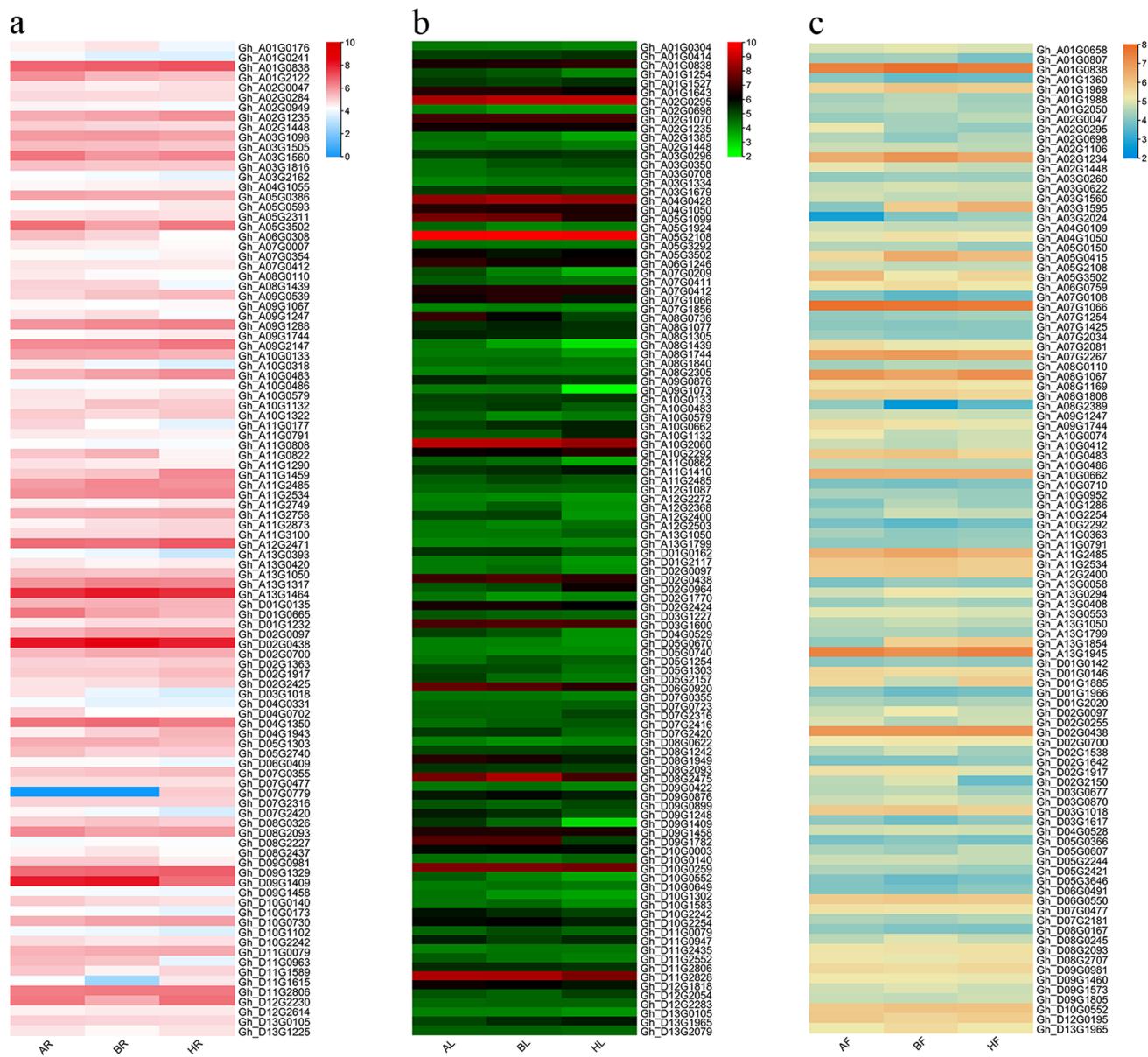


Fig. 1 Heatmaps for novel biased homoeolog expressed genes in each tissue of high hybrid. **a** Novel biased expressed genes in root. **b** Novel biased expressed genes in leaf. **c** Novel biased expressed genes in flower buds

(Fig. 1c). Further analysis indicated that 1.2% homoeolog gene pairs in the root, 1.06% in leaf, and 1.1% in flower buds demonstrated overall maternal parent biased expression, whereas overall homoeolog gene pairs with paternal parent biased expression were 1.1%, 1.04%, and 0.95% in the root, leaf, and flower buds, respectively. The overall statistics of homoeolog expression changes in different tissues of high hybrid seem to show a highly balanced biased expression of homoeolog gene pairs in each tissue. These results suggested that homoeolog gene pairs that displayed fewer novel patterns of expression may be responsible for the high level of heterosis in upland cotton.

Homoeolog expression changes in different tissues of low hybrid

The results of homoeolog expression analysis in different tissues of low hybrid (L) and its reciprocal parents (A= maternal, D= paternal) are represented in Table 2. Interestingly, the homoeolog expression patterns in hybrid were in the same direction for the majority of homoeolog gene pairs as observed in parents. For instance, 97.3% in the root, 97.2% in the leaf, and 97.6% in flower buds out of all homoeolog gene pairs (36853) reflected equal parental expression patterns in this hybrid. These result predicted that expression

Table 2 Homoeolog gene expression changes in different tissues of the low upland cotton hybrid

Expression in parents	Expression in progeny	Root (%)	Leaf (%)	Flower buds (%)
A=D	A=D Parental condition	35844 (97.3)	35827 (97.2)	35981 (97.6)
A<D	A<D Parental condition	0 (0.0)	0 (0.0)	1 (0.0)
A>D	A>D Parental condition	2 (0.01)	1 (0.0)	1 (0.0)
A>D	A=D No bias in progeny	202 (0.55)	233 (0.63)	107 (0.29)
A<D	A=D No bias in progeny	197 (0.53)	135 (0.37)	116 (0.31)
A=D	A>D Novel bias in progeny	290 (0.89)	352 (0.96)	294 (0.80)
A<D	A>D Novel bias in progeny	12 (0.03)	8 (0.02)	14 (0.04)
A>D	A<D Novel bias in progeny	6 (0.02)	8 (0.02)	8 (0.02)
A=D	A<D Novel bias in progeny	300 (0.81)	288 (0.78)	331 (0.90)
Overall A-biased in progeny		304 (0.83)	361 (0.98)	309 (0.84)
Overall D-biased in progeny		306 (0.84)	297 (0.81)	340 (0.92)
Total number of genes		36853	36853	36853

Here in table, A is the maternal parent and D is the paternal parent

biases pre-existing in parents was simply conserved in these tissues of hybrid. The group of homoeolog gene pairs with bias expression in parents and reverted to no bias expression in hybrid had an average of 1% in root and leaf. On the other hand, only 0.60% of homoeolog gene pairs in flower buds had this reversion. According to the results of novel bias expression in hybrid, 1.75, 1.78, and 1.76% of homoeolog gene pairs adopted novel bias in the root, leaf, and flower buds, respectively. The majority of homoeolog gene pairs exhibited novel biased expression similar to maternal parents in root and leaf (Fig. 2a–b). In contrast, flower buds had most of the homoeolog gene pairs that had shown novel biased expression with a higher paternal contribution in progeny (Fig. 2c). Overall, a group of maternal or paternal biased homoeolog gene pairs had very similar trends in each tissue. There was 0.83% maternal and 0.84% paternal biased homoeolog gene expression in the root, whereas the leaf had 0.98 and 0.81% maternal and paternal biased expression, respectively. Flower buds also showed a similar percentage of maternal and parental biased expression. Altogether, these results suggest that the majority of the homoeolog gene exhibited balanced or non-differential expression patterns in low hybrid. In addition, few homoeolog genes with novel expression profiles probably contribute to the low level of heterosis in upland cotton.

Distribution of novel expressed genes among different hybrids

This study further investigated the distribution of novel expressed genes among each tissue of hybrids and different hybrids. In the high hybrid, 374, 325, and 301 novel genes exhibited tissue-specific expression in root, leaf, and flower buds, respectively (Fig. 3). However, nonspecific or overlapped genes compared to tissue-specific expression did not account for the majority in these tissues. For example,

94 genes were overlapped between the comparison of root and leaf whereas 100 genes were overlapped between root and flower buds. The distribution results had shown that 298 novel genes displayed specific expression in root of low hybrid (Fig. 4). Similarly, the novel genes with specific expression were 324 in leaf and 308 in flower buds of low hybrid. There were 57 overlapped novel genes between root and leaf while flower bud and leaf had 76 overlapped novel genes. The novel gene distribution between high and low hybrids revealed the majority of novel genes were hybrid-specific and less overlapped among both hybrids (Supplementary Fig. S1). For instance, only 122 genes were overlapped among hybrids, while 877 genes were specifically expressed in high hybrid and 808 in low hybrids. These results anticipated that novel genes with hybrid-specific expression most likely caused contrasting levels of heterosis in hybrids.

Functional analysis of homoeolog genes exhibiting novel expression changes in hybrids

As to understand the possible functions of homoeolog genes that had novel biased expression under hybridization in cotton, we performed GO and KEGG enrichment analysis in each hybrid. GO enrichment analysis ($p\text{-value} < 0.05$) for high hybrid showed a majority of novel biased expressed homoeolog genes had functional annotation related to catalytic activity, nucleus, metabolic process, transmembrane transport, binding of nucleotide, nucleic acid, protein, ATP, and RNA (Fig. 5a). Genes associated with the binding of protein and ATP had the highest portion of novel biased expressed homoeolog genes in this hybrid. Most enriched GO terms for novel biased expressed homoeolog genes in low hybrid were involved in functional processes related to the nucleus, catalytic activity, binding of ATP, protein, and nucleotide (Fig. 5b). Interestingly, genes that performed the

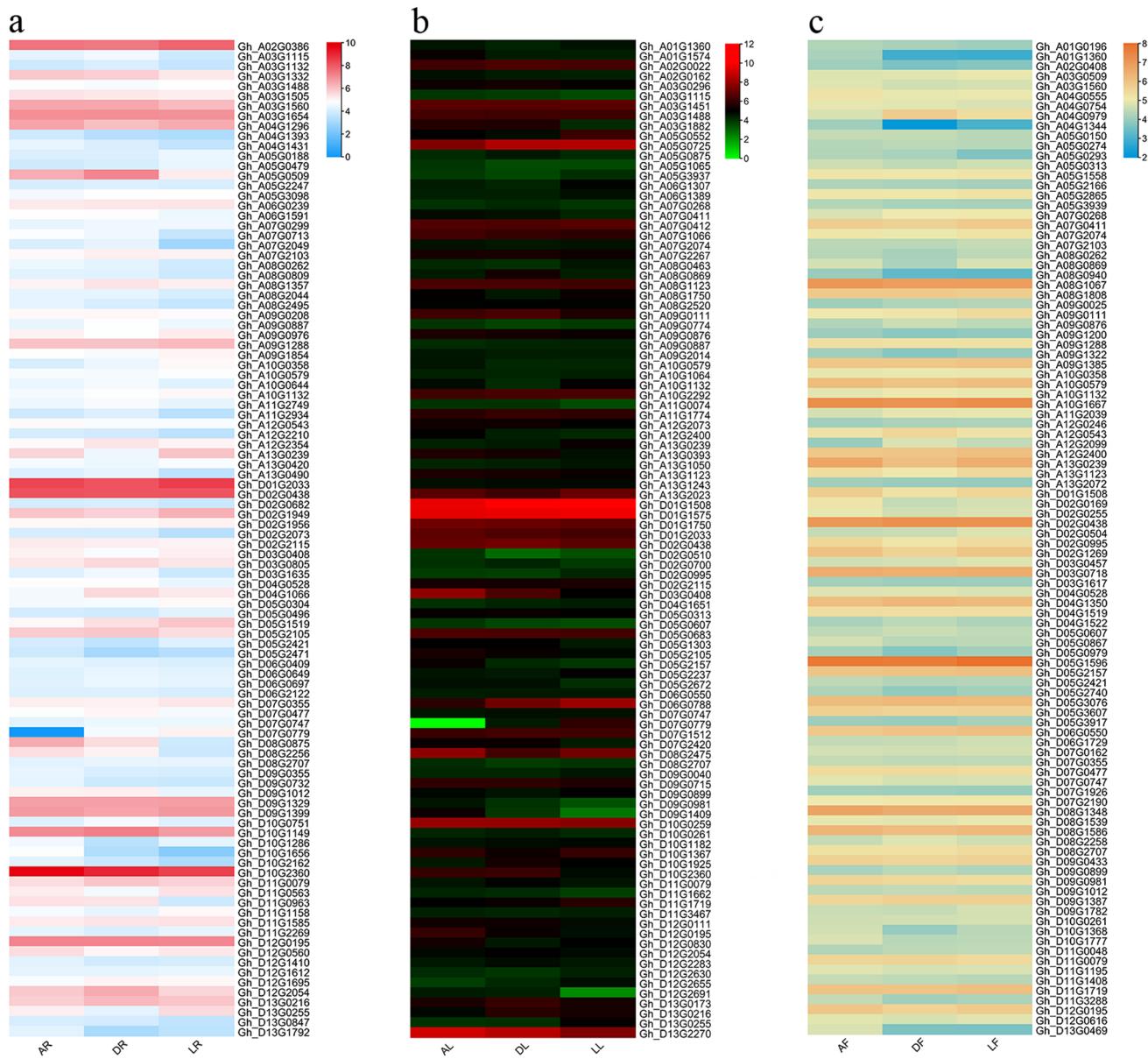


Fig. 2 Heatmaps for novel biased homoeolog expressed genes in each tissue of low hybrid. **a** Novel biased expressed genes in root. **b** Novel biased expressed genes in leaf. **c** Novel biased expressed genes in flower buds

binding function of protein and ATP also had the majority of novel biased expressed homoeolog genes in this hybrid.

KEGG analysis revealed that the majority of novel biased expressed homoeolog genes in high hybrid were enriched in pathways such as biosynthesis of amino acids, mRNA surveillance pathway, glycolysis/gluconeogenesis, biosynthesis of antibiotics, metabolism of carbon, glutathione, cysteine/methionine, biosynthesis of secondary metabolites, and microbial metabolism in diverse environments (Fig. 6a). Biosynthesis of secondary metabolites had more significant gene enrichment as compared to other pathways. It was observed that a major

portion of novel biased expressed homoeolog genes in low hybrid had enrichment in mRNA surveillance pathway, RNA degradation, oocyte meiosis, metabolism of pyruvate, carbon, sugars, and biosynthesis of secondary metabolites (Fig. 6b). In a low hybrid, biosynthesis of secondary metabolites also had a maximum number of novel biased expressed homoeolog genes as compared to other pathways. Remarkably, the biosynthesis of secondary metabolites had not only a higher percentage of novel biased expressed homoeolog genes but also was the overlapped pathway in both hybrids. Therefore, our functional analysis predicts that genes annotated in the biosynthesis

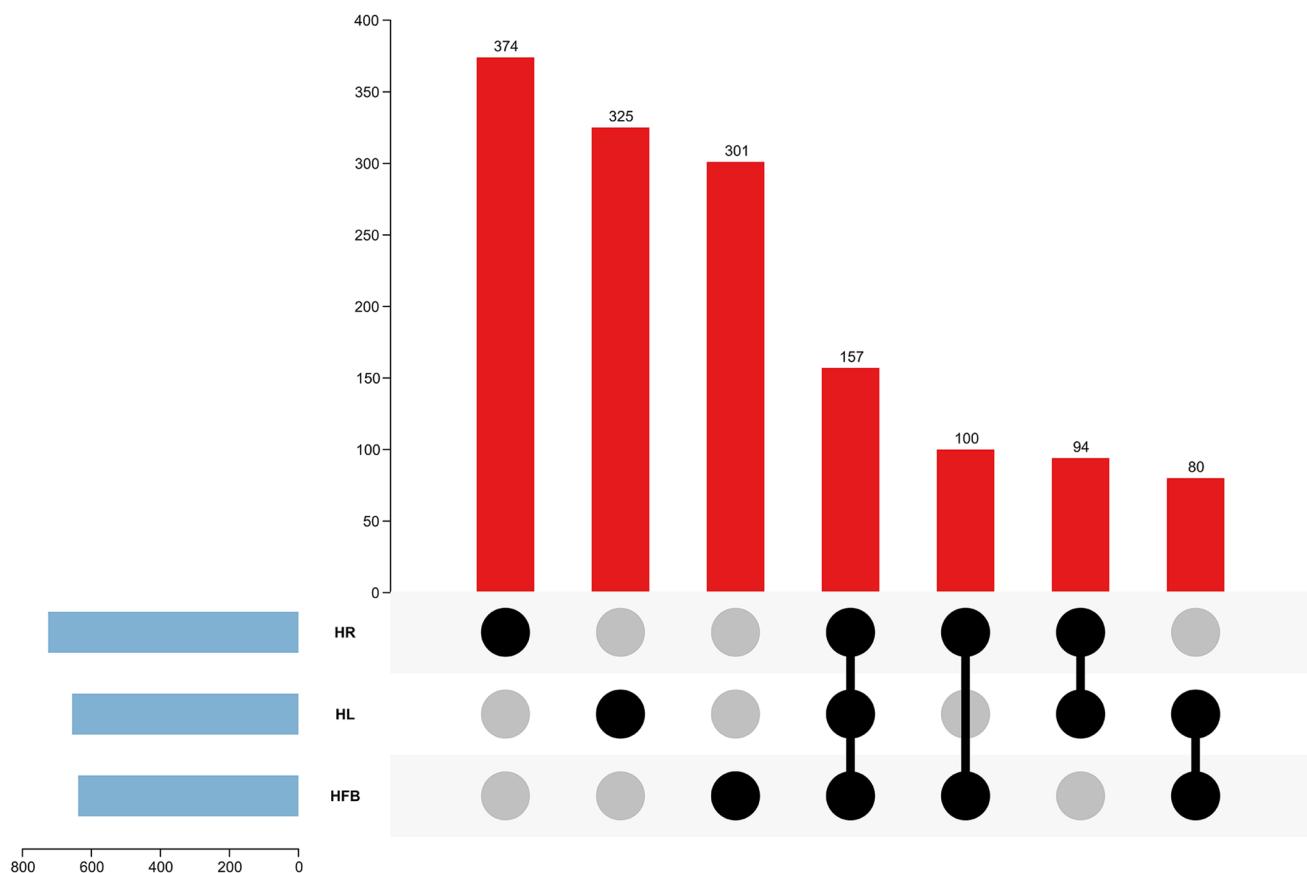


Fig. 3 Distribution of novel biased homoeolog expressed genes in high hybrid. Here R, L, and FB stand for root, leaf, and flower buds, respectively

of secondary metabolites had a crucial role in the regulation of cotton heterosis compared to other pathway genes.

Discussion

Cotton is the backbone of the textile industry because of its unique fiber characteristics. However, the area of cotton cultivation, as well as production, is decreasing in recent decades all over the world. Hybridization has effectively resolved the problem of decreased crop productivity. For example, maize and rice cultivars developed with the exploitation of heterosis have more harvest potential than conservative genotypes (Schnable and Springer 2013). Cotton belongs to the genus *Gossypium* and upland cotton (*G. hirsutum*) is stated as new world cotton. The hybridization in upland cotton whether it is between species (interspecific) or within species (intraspecific) produced meaningful heterosis in yield-related traits (Li et al. 2019). It was observed that intraspecific crosses and improvement in fiber quality traits require more breeding efforts in cotton. How hybridization increased yield potential in upland cotton needs plenty of

investigation at the genomic, transcriptomic, and epigenetic levels. Interestingly, the previous study has shown hybridization itself had sound effects on duplicate genes (homoeolog genes) in allopolyploids (Yoo et al. 2013; Ren et al. 2016). This study analyzed the effects of intraspecific hybridization on naturally occurring homoeolog genes in upland cotton through RNA-seq data of contrasting hybrids and full-length Pacbio sequences of their respective parents. In particular, gene expression data from three tissues (root, leaf, and flower buds) were used to explore changes in homoeolog expression and functional annotation of novel biased genes of hybrids.

Hybridization produced a parent legacy of homoeolog expressions in upland cotton

Hybridization is a vital phenomenon to produce genetic improvement in crops. It is a common method in which two genetically different genotypes are crossed to obtain vigor progeny in desired traits. More than two sets of chromosomes in allotetraploid upland cotton caused genetic, gene expression, and epigenetic changes. These effects may result in more

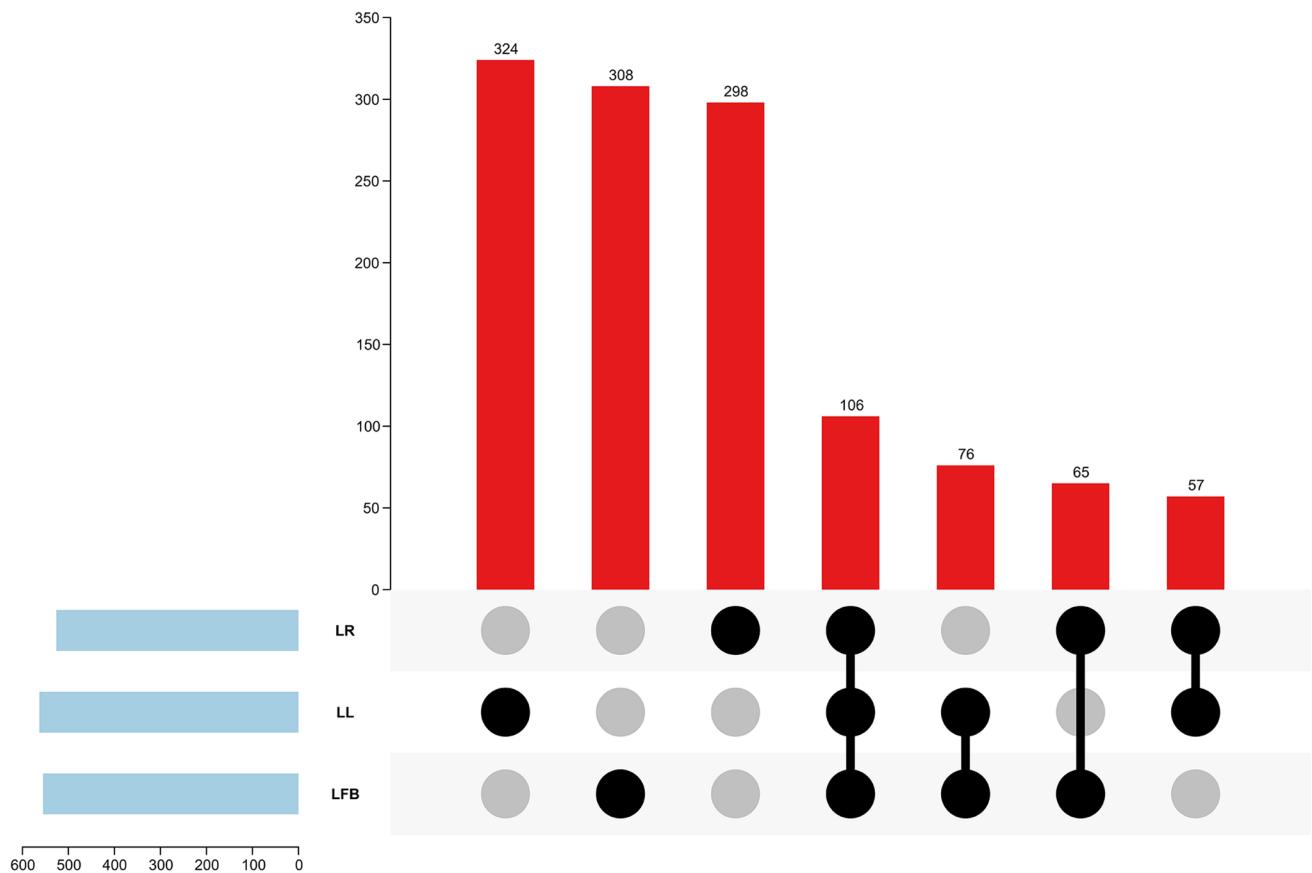


Fig. 4 Distribution of novel biased homoeolog expressed genes in low hybrid. Here R, L, and FB stand for root, leaf, and flower buds, respectively

vigorous growth or heterosis (Chen 2007). Previous studies in maize (Yao et al. 2013) and alfalfa (Bingham et al. 1994) suggested that allopolyploids exhibit more heterosis as compared to autopolyploids and tetraploids display increased heterosis as compared with diploids. It was found that intraspecific crossing in upland cotton produced progeny that often displays heterosis in early biomass growth and seed cotton yield (Shahzad et al. 2020a, b). However, there has been no study to analyze changes in homoeolog gene expression of inbred parents due to hybridization. The result of our study revealed that intraspecific hybridization in upland cotton produced highly balanced changes in the pattern of homoeolog expression. The pattern was conserved in hybrids for the greater number of genes. For example, both higher and lower hybrids had around 95% homoeolog gene pairs with similar expression pre-existing in parents in all tissues under investigation. Fewer homoeolog gene copies imitated novel homoeolog expression patterns in hybrids. Furthermore, fewest homoeolog gene pairs exhibited maternal or paternal overall biases expression. A previous study in allopolyploid cotton showed that most homoeolog gene pairs reflected parental expression conditions in hybrid and synthetic allopolyploids (Yoo et al. 2013). Similar results

were reported in hybrid fish and its progenitors (Ren et al. 2016; Ren et al. 2017). An assessment of homoeolog expressions in *Brassica napus* described that majority of homoeolog gene pairs maintained parental expression patterns in synthetic hybrids and natural allotetraploids (Zhang et al. 2016). The effect of polyploidization on homoeolog expression in hexaploid wheat revealed equal homoeolog expression of parental lines is often conserved in progeny (Li et al. 2014). Altogether, these results indicated that the effect of hybridization or genome doubling on the expression of homoeolog gene pairs in derived hybrids is often inherited from diploid parents. Though the relationship between heterosis and homoeolog gene pairs is still not fully understood, these findings suggest that fewer changes at the transcriptomic level could be associated with heterosis in upland cotton.

Novel homoeolog gene expressions probably regulate heterosis in cotton through biosynthesis of secondary metabolites

As already described in the majority of the results, homoeolog gene pairs had a parental legacy in hybrids.

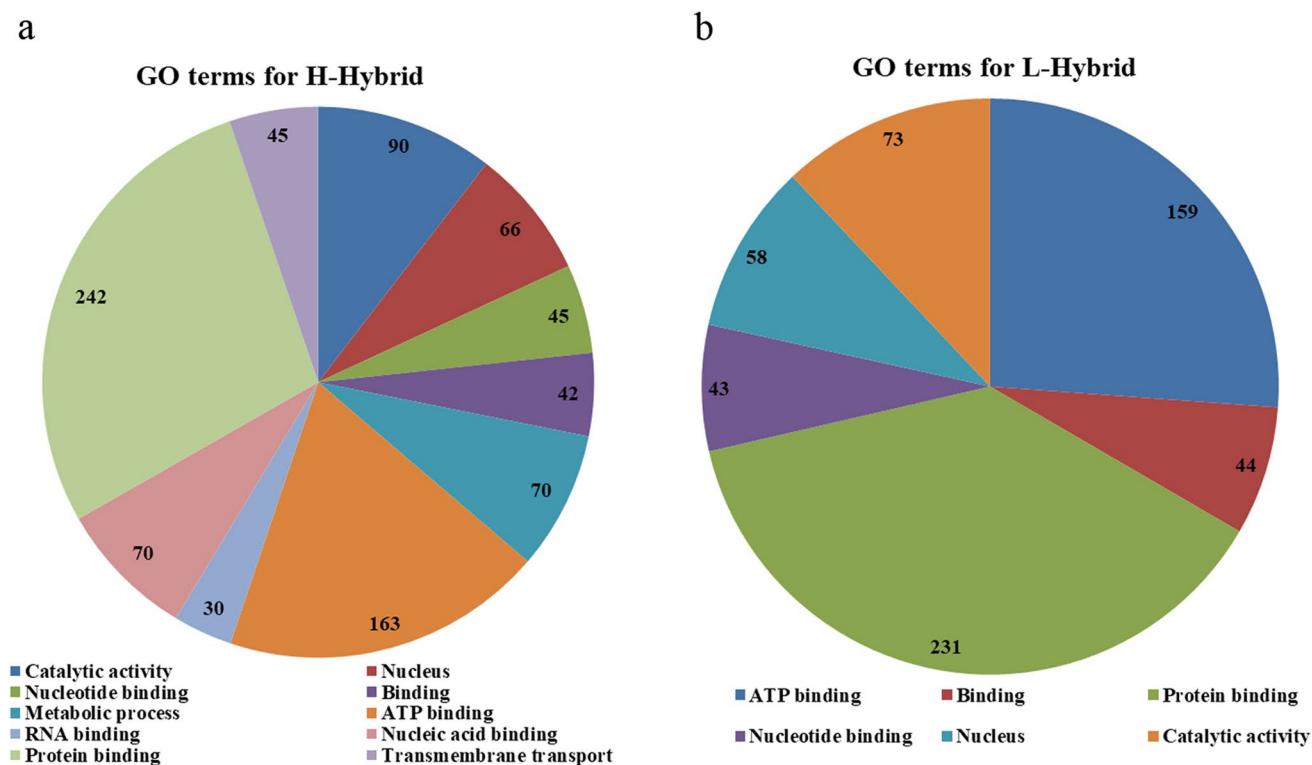


Fig. 5 Most enriched GO terms for novel biased homoeolog expressed genes in each hybrid. **a** GO terms for the high hybrid. Black numbers represent total genes for each term at a significance

of 0.05. **b** GO terms for the low hybrid. Black numbers represent total genes for each term at a significance of 0.05

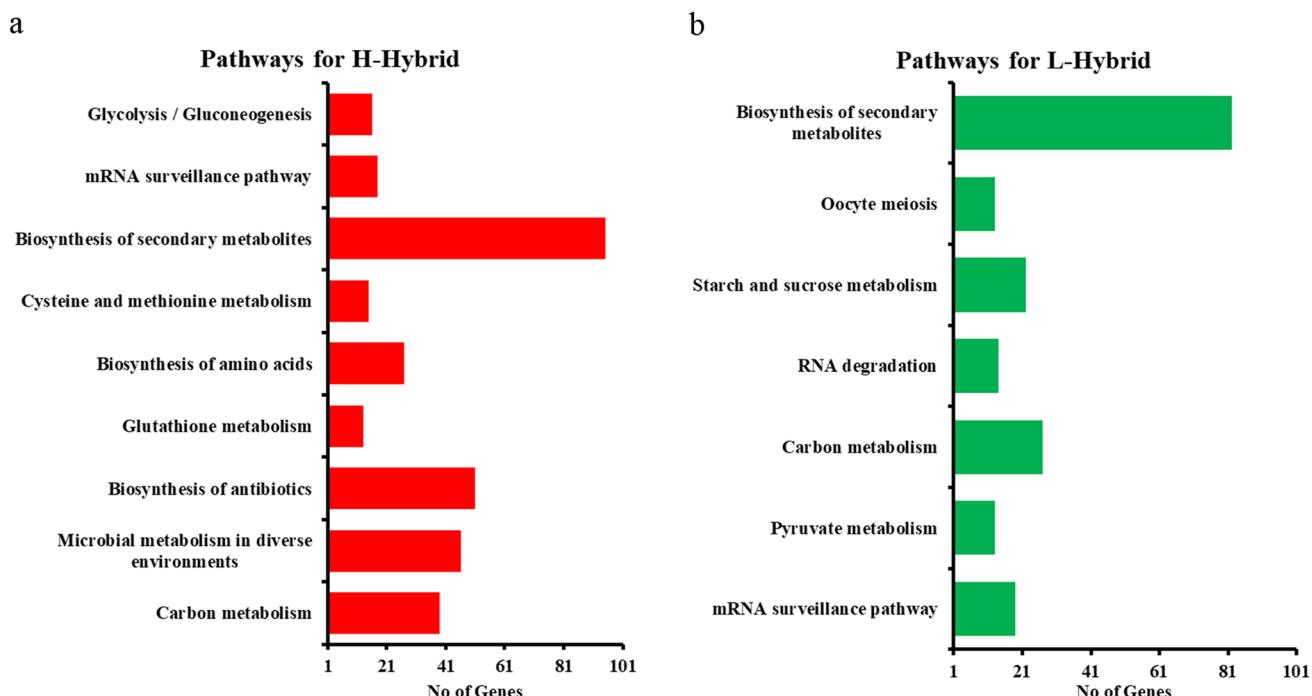


Fig. 6 Most enriched Pathways for novel biased homoeolog expressed genes in each hybrid. **a** Pathways for the high hybrid. **b** Pathways for the low hybrid

Theoretically, these gene pairs do not contribute to the superior or inferior performance of cotton hybrids. Homoeolog gene pairs with a novel or differential expression in hybrids could regulate heterosis. Functional analysis of these novel homoeolog gene pairs revealed that protein and ATP binding genes through the secondary metabolism pathway play an essential role to increase growth potential in cotton hybrids. Metabolites are an essential component for adaptation, stress management, and regulation of different biological processes in plants (Kleessen et al. 2014; Miller et al. 2015). Previous research showed that modification of secondary metabolism is useful for growth heterosis in oilseed crops such as *Brassica juncea* (Bajpai et al. 2019). Furthermore, these researchers reported an additive mode of inheritance of metabolite components in hybrids. Previous metabolic investigation in hybrids of maize (Wang et al. 2014), rice (Ma et al. 2011), and Arabidopsis (Korn et al. 2010) provided preliminary perceptions and enhanced knowledge about the molecular mechanism of plant heterosis. There is plenty of emphasis on genetic, gene expression, and epigenetic-related studies to explore the biological basis of heterosis in cotton. Yet, the relationship between heterosis and metabolic changes in hybrids has been poorly known in cotton. Our results underpin the importance of genomic inheritance of secondary metabolites between hybrids and inbred parents in cotton.

Conclusion

In this study, global homoeolog gene expression patterns between intraspecific contrasting cotton hybrids and their reciprocal inbred parents were analyzed using RNA sequencing. The results indicated that hybridization caused no expression biases of homoeolog gene pairs in hybrids. Hybrids showed no biased expression preference towards any parents. Furthermore, functional annotation of novel expressed homoeolog gene pairs explained that changed secondary metabolic gene expression between parents and hybrids may regulate heterosis in upland cotton. Together, our results provide a general understanding of homoeolog gene expression patterns and may help to explore the genetic basis of heterosis in cotton.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10142-022-00877-1>.

Acknowledgements The authors thank Yasir Muhammad for his valuable suggestions in data analysis, reviewers for their valuable critics, and field staff for their anonymous work to plant materials.

Author contribution Xing CZ and Wu JY designed and funded this research. Guo LP, Qi TX, Tang HN, Qiao XQ, and Wang HL prepared the plant population. Zhang XX, Mubeen I, and Zhang M carried out

RNA extraction, performed data analysis, and helped in the final figures. Shahzad K, Zhang M, and Peng RH drafted and revised the manuscript. All authors read and approved the final manuscript.

Funding This work was funded by Zhongyuan Academician Foundation (212101510001) and by the General Program of the National Natural Science Foundation of China (31871679). The funding organizations were not involved in the design of the study, analysis, and writing the manuscript.

Data availability All datasets analyzed in this study are presented in this article. All utilized RNA sequencing datasets are freely available with accession numbers GSE144676 and GSE150052 at NCBI.

Declarations

Ethics approval and consent to participate All the cotton materials used and analyzed were public property of the Institute of Cotton Research of Chinese Academy of Agricultural Sciences, China, and were available on-demand for non-commercial purposes.

Human and animal ethics 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 no competing interests.

References

- Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceed Nat Acad Sci* 100(8):4649–4654. <https://doi.org/10.1073/pnas.0630618100>
- Akhunova AR, Matniyazov RT, Liang H, Akhunov ED (2010) Homoeolog-specific transcriptional bias in allopolyploid wheat. *BMC Genom* 11(1):505. <https://doi.org/10.1186/1471-2164-11-505>
- Ali I, Shakeel A, Saeed A, Nazeer W, Zia Z, Ahmad S, Mahmood K, Malik W (2016) Combining ability analysis and heterotic studies for within-boll yield components and fibre quality in cotton. *J Animal Plant Sci* 26(1):156–162
- Anssour S, Krügel T, Sharbel T, Saluz H, Bonaventure G, Baldwin I (2009) Phenotypic, genetic and genomic consequences of natural and synthetic polyploidization of *Nicotiana attenuata* and *Nicotiana obtusifolia*. *Annals of Bot* 103(8):1207–1217. <https://doi.org/10.1093/aob/mcp058>
- Auger B, Baron C, Lucas M-O, Vautrin S, Bergès H, Chalhoub B, Fauret A, Renard M, Nesi N (2009) *Brassica* orthologs from BANYULS belong to a small multigene family, which is involved in procyandin accumulation in the seed. *Planta* 230(6):1167. <https://doi.org/10.1007/s00425-009-1017-0>
- Bajpai PK, Reichelt M, Augustine R, Gershenson J, Bisht NC (2019) Heterotic patterns of primary and secondary metabolites in the oilseed crop *Brassica juncea*. *Heredity* 123(3):318–336. <https://doi.org/10.1038/s41437-019-0213-3>
- Bingham E, Groose R, Woodfield D, Kidwell KK (1994) Complementary gene interactions in alfalfa are greater in autotetraploids than diploids. *Crop Sci* 34(4):823–829. <https://doi.org/10.2135/cropsci1994.0011183X003400040001x>
- Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. *Plant Cell* 15(10):2236–2239. <https://doi.org/10.1105/tpc.151030>

- Bottley A, Xia G, Koechner R (2006) Homoeologous gene silencing in hexaploid wheat. *Plant J* 47(6):897–906. <https://doi.org/10.1111/j.1365-313X.2006.02841.x>
- Buggs RJ, Elliott NM, Zhang L, Koh J, Viccini LF, Soltis DE, Soltis PS (2010) Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol* 186(1):175–183. <https://doi.org/10.1111/j.1469-8137.2010.03205.x>
- Chang PL, Dilkes BP, McMahon M, Comai L, Nuzhdin SV (2010) Homoeolog-specific retention and use in allotetraploid *Arabidopsis suecica* depends on parent of origin and network partners. *Genome Biol* 11(12):R125. <https://doi.org/10.1186/gb-2010-11-12-r125>
- Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genet* 182(2):503–517. <https://doi.org/10.1534/genetics.109.102608>
- Chelaifa H, Monnier A, Ainouche M (2010) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina townsendii* and *Spartina anglica* (Poaceae). *New Phytol* 186(1):161–174. <https://doi.org/10.1111/j.1469-8137.2010.03179.x>
- Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* 58:377–406. <https://doi.org/10.1146/annurev.arplant.58.032806.103835>
- De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proceed Nat Acad Sci* 110(8):2898–2903. <https://doi.org/10.1073/pnas.1300127110>
- Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. *Ann Review of Genet* 42:443–461. <https://doi.org/10.1146/annurev.genet.42.110807.091524>
- Dufresne F, Hebert PD (1994) Hybridization and origins of polyploidy. *Proceed Royal Society of London Series B: Bioll Sci* 258(1352):141–146
- Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol* 6(1):16. <https://doi.org/10.1186/1741-7007-6-16>
- Flagel LE, Wendel JF (2010) Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. *New Phytol.* 186(1):184–193. <https://doi.org/10.1111/j.1469-8137.2009.03107.x>
- Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. *Plant Cell* 19(11):3403–3417. <https://doi.org/10.1105/tpc.107.054346>
- Gaeta RT, Yoo S-Y, Pires J, Doerge RW, Chen ZJ, Osborn TC (2009) Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using *Arabidopsis* 70mer oligo microarrays. *PLoS One* 4(3):e4760. <https://doi.org/10.1371/journal.pone.0004760>
- Grover C, Gallagher J, Szadkowski E, Yoo M, Flagel L, Wendel J (2012) Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol* 196(4):966–971. <https://doi.org/10.1111/j.1469-8137.2012.04365.x>
- Han F, Fedak G, Guo W, Liu B (2005) Rapid and repeatable elimination of a parental genome-specific DNA repeat (pGc1R-1a) in newly synthesized wheat allopolyploids. *Genet* 170(3):1239–1245. <https://doi.org/10.1534/genetics.104.039263>
- Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendel JF (2008) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proceed Nat Acad Sci* 105(16):6191–6195. <https://doi.org/10.1073/pnas.0711569105>
- Hughes TE, Langdale JA, Kelly S (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Res* 24(8):1348–1355. <https://doi.org/10.1101/gr.172684.114>
- Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genet* 33(1):102–106. <https://doi.org/10.1038/ng1063>
- Kleessen S, Laitinen R, Fusari CM, Antonio C, Sulpice R, Fernie AR, Stitt M, Nikoloski Z (2014) Metabolic efficiency underpins performance trade-offs in growth of *Arabidopsis thaliana*. *Nat Commun* 5(1):1–10. <https://doi.org/10.1038/ncomms4537>
- Kleinjan DA, Banczewicz RM, Gautier P, Dahn R, Schonthaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V (2008) Subfunctionalization of duplicated zebrafish pax6 genes by cis-regulatory divergence. *PLoS Genet* 4(2):e29. <https://doi.org/10.1371/journal.pgen.0040029>
- Korn M, Gärtner T, Erban A, Kopka J, Selbig J, Hincha DK (2010) Predicting *Arabidopsis* freezing tolerance and heterosis in freezing tolerance from metabolite composition. *Mol Plant* 3(1):224–235. <https://doi.org/10.1093/mp/ssp105>
- Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR (2008) Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. *Annals of Bot* 101(6):815–823. <https://doi.org/10.1093/aob/mcn019>
- Kraitshtein Z, Yaakov B, Khasdan V, Kashkush K (2010) Genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. *Genet* 186(3):801–812. <https://doi.org/10.1534/genetics.110.120790>
- Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. *Plant Cell* 26(5):1878–1900. <https://doi.org/10.1105/tpc.114.124388>
- Li M, Wang R, Wu X, Wang J (2020) Homoeolog expression bias and expression level dominance (ELD) in four tissues of natural allotetraploid *Brassica napus*. *BMC Genom* 21:1–15. <https://doi.org/10.1186/s12864-020-6747-1>
- Li X, Shahzad K, Guo L, Qi T, Zhang X, Wang H, Tang H, Qiao X, Zhang J, Wu J (2019) Using yield quantitative trait locus targeted SSR markers to study the relationship between genetic distance and yield heterosis in upland cotton (*Gossypium hirsutum*). *Plant Breed* 138(1):105–113. <https://doi.org/10.1111/pbr.12668>
- Liu SL, Adams KL (2010) Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the *Brassicaceae*. *Mol Biol Evol* 27(12):2817–2828. <https://doi.org/10.1093/molbev/msq169>
- Ma Q, Hedden P, Zhang Q (2011) Heterosis in rice seedlings: its relationship to gibberellin content and expression of gibberellin metabolism and signaling genes. *Plant Physiol* 156(4):1905–1920. <https://doi.org/10.1104/pp.111.178046>
- Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J (2002) Comai L (2010) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. *Plant physiology* 129(2):733–746. <https://doi.org/10.1104/pp.003095>
- McGrath CL, Gout J-F, Johri P, Doak TG, Lynch M (2014) Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res* 24(10):1665–1675. <https://doi.org/10.1101/gr.173740.114>
- Miller M, Song Q, Shi X, Juenger TE, Chen ZJ (2015) Natural variation in timing of stress-responsive gene expression predicts heterosis

- in intraspecific hybrids of *Arabidopsis*. *Nat Commun* 6(1):1–13. <https://doi.org/10.1038/ncomms8453>
- Mochida K, Yamazaki Y, Ogiwara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genom* 270(5):371–377. <https://doi.org/10.1007/s00438-003-0939-7>
- Ren L, Cui J, Wang J, Tan H, Li W, Tang C, Qin Q, Liu S (2017) Analyzing homoeolog expression provides insights into the rediploidization event in gynogenetic hybrids of *Carassius auratus* red var. × *Cyprinus carpio*. *Sci Rep* 7(1):1–11. <https://doi.org/10.1038/s41598-017-14084-7>
- Ren L, Li W, Tao M, Qin Q, Luo J, Chai J, Tang C, Xiao J, Tang X, Lin G (2016) Homoeologue expression insights into the basis of growth heterosis at the intersection of ploidy and hybridity in Cyprinidae. *Sci Rep* 6(1):1–12. <https://doi.org/10.1038/srep27040>
- Salmon A, Flagel L, Ying B, Udall JA, Wendel JF (2010) Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol* 186(1):123–134. <https://doi.org/10.1111/j.1469-8137.2009.03093.x>
- Samuel Yang S, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J* 47(5):761–775. <https://doi.org/10.1111/j.1365-313X.2006.02829.x>
- Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS one.* 6(3):e17855. <https://doi.org/10.1371/journal.pone.0017855>
- Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. *Annual Rev Plant Biol* 64. <https://doi.org/10.1146/annurev-arplant-042110-103827>
- Shahzad K, Qi T, Guo L, Tang H, Zhang X, Wang H, Qiao X, Zhang M, Zhang B, Feng J (2019a) Adaptability and stability comparisons of inbred and hybrid cotton in yield and fiber quality traits. *Agronomy* 9(9):516. <https://doi.org/10.3390/agronomy9090516>
- Shahzad K, Xue L, Tingxiang Q, Liping G, Huini T, Zhang X, Hailin W, Zhang M, Zhang B, Xiuqin Q (2019b) Genetic analysis of yield and fiber quality traits in upland cotton (*Gossypium hirsutum* L.) cultivated in different ecological regions of China. *J Cotton Res* 2(1):14. <https://doi.org/10.1186/s42397-019-0031-4>
- Shahzad K, Zhang X, Guo L, Qi T, Bao L, Zhang M, Zhang B, Wang H, Tang H, Qiao X (2020a) Comparative transcriptome analysis between inbred and hybrids reveals molecular insights into yield heterosis of upland cotton. *BMC Plant Biol* 20:1–18. <https://doi.org/10.1186/s12870-020-02442-z>
- Shahzad K, Zhang X, Guo L, Qi T, Tang H, Zhang M, Zhang B, Wang H, Qiao X, Feng J (2020b) Comparative transcriptome analysis of inbred lines and contrasting hybrids reveals overdominance mediate early biomass vigor in hybrid cotton. *BMC Genom* 21(1):1–16. <https://doi.org/10.1186/s12864-020-6561-9>
- Szadkowski E, Eber F, Huteau V, Lode M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux M, Delourme R, King GJ The first meiosis of resynthesized *Brassica napus*, a genome blender. *New Phytol* 186(1):102–112. <https://doi.org/10.1111/j.1469-8137.2010.03182.x>
- Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge R, Comai L (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genet* 172(1):507–517. <https://doi.org/10.1534/genetics.105.047894>
- Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. *Genet* 167(4):1961–1973. <https://doi.org/10.1534/genetics.104.027896>
- Wang Z, Gerstein M, Snyder M: RNA-Seq (2009) A revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63. <https://doi.org/10.1038/nrg2484>
- Wang Z, Xue Z, Wang T (2014) Differential analysis of proteomes and metabolomes reveals additively balanced networking for metabolism in maize heterosis. *J Proteome Res* 13(9):3987–4001. <https://doi.org/10.1021/pr500337j>
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. *Proceed Nat Acad Sci* 106(33):13875–13879. <https://doi.org/10.1073/pnas.0811575106>
- Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genom* 19(1):1–13. <https://doi.org/10.1186/s12864-018-4966-5>
- Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genet* 140(2):745–754. <https://doi.org/10.1093/genetics/140.2.745>
- Yao H, Gray AD, Auger DL, Birchler JA (2013) Genomic dosage effects on heterosis in triploid maize. *Proceed Nat Acad Sci* 110(7):2665–2669. <https://doi.org/10.1073/pnas.1221966110>
- Yoo M, Szadkowski E, Wendel J (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* 110(2):171–180. <https://doi.org/10.1038/hdy.2012.94>
- Yu S, Li J, Xu C, Tan Y, Gao Y, Li X, Zhang Q, Maroof MS (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proceed Nat Acad Sci* 94(17):9226–9231. <https://doi.org/10.1073/pnas.94.17.9226>
- Zhang D, Pan Q, Tan C, Zhu B, Ge X, Shao Y, Li Z (2016) Genome-wide gene expressions respond differently to A-subgenome origins in *Brassica napus* synthetic hybrids and natural allotetraploid. *Front Plant Sci* 7:1508. <https://doi.org/10.3389/fpls.2016.01508>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.