the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) **104,** 1285–1300 doi: 10.1111/tpj.14999

Genome-wide dissection of hybridization for fiber quality- and yield-related traits in upland cotton

Xiaoli Geng[1,2], Gaofei Sun[3,*], Yujie Qu[1], Zareen Sarfraz[1], Yinhua Jia[1,2], Shoupu He[1,2], Zhaoe Pan[1], Junling Sun[1], Muhammad S. Iqbal[4], Qinglian Wang[5], Hongde Qin[6], Jinhai Liu[7], Hui Liu[8], Jun Yang[9], Zhiying Ma[10], Dongyong Xu[11], Jinlong Yang[7], Jinbiao Zhang[12], Zhikun Li[10], Zhongmin Cai[7], Xuelin Zhang[13], Xin Zhang[5], Guanyin Zhou[7], Lin Li[12], Haiyong Zhu[1], Liru Wang[1], Baoyin Pang[1] and Xiongming Du[1,2,*]

[1]*State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China,*
[2]*Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University, Zhengzhou 455001, China,*
[3]*Anyang Institute of Technology, Anyang 455000, China,*
[4]*Cotton Research Station, Ayub Agricultural Research Institute, Faisalabad 38000, Pakistan,*
[5]*Henan Institute of Science and Technology, Xinxiang 453003, China,*
[6]*Cash Crop Institute, Hubei Academy of Agricultural Sciences, Wuhan 430000, China,*
[7]*Zhongmian Cotton Seed Industry Technology Co., Ltd, Zhengzhou 455001, China,*
[8]*Jing Hua Seed Industry Technologies Inc, Jingzhou 434000, China,*
[9]*Cotton Research Institute of Jiangxi Province, Jiujiang 332000, China,*
[10]*Key Laboratory of Crop Germplasm Resources of Hebei, Hebei Agricultural University, Baoding 071000, China,*
[11]*Guoxin Rural Technical Service Association, Hejian 062450, China,*
[12]*Zhongli Company of Shandong, Dongying 257000, China, and*
[13]*Hunan Cotton Research Institute, Changde 415000, China*

Received 4 September 2019; revised 14 July 2020; accepted 3 September 2020; published online 29 September 2020.
For correspondence (e-mail dujeffrey8848@hotmail.com; sungaofei@sina.com).

SUMMARY

An evaluation of combining ability can facilitate the selection of suitable parents and superior F_1 hybrids for hybrid cotton breeding, although the molecular genetic basis of combining ability has not been fully characterized. In the present study, 282 female parents were crossed with four male parents in accordance with the North Carolina II mating scheme to generate 1128 hybrids. The parental lines were genotyped based on restriction site-associated DNA sequencing and 306 814 filtered single nucleotide polymorphisms were used for genome-wide association analysis involving the phenotypes, general combining ability (GCA) values, and specific combining ability values of eight fiber quality- and yield-related traits. The main results were: (i) all parents could be clustered into five subgroups based on population structure analyses and the GCA performance of the female parents had significant differences between subgroups; (ii) 20 accessions with a top 5% GCA value for more than one trait were identified as elite parents for hybrid cotton breeding; (iii) 120 significant single nucleotide polymorphisms, clustered into 66 quantitative trait loci, such as the previously reported *Gh_A07G1769* and *GhHOX3* genes, were found to be significantly associated with GCA; and (iv) identified quantitative trait loci for GCA had a cumulative effect on GCA of the accessions. Overall, our results suggest that pyramiding the favorable loci for GCA may improve the efficiency of hybrid cotton breeding.

Keywords: fiber quality, fiber yield, combining ability, genome-wide association study, single nucleotide polymorphism, upland cotton.

Highlight

We have identified 120 significant single nucleotide polymorphisms and 66 quantitative trait loci through genome-wide association analysis involving general combining ability for eight fiber quality- and yield-related traits.

INTRODUCTION

Upland cotton (*Gossypium hirsutum* L.) is an important natural fiber crop, accounting for approximately 95% of cotton production worldwide. Previous studies have revealed that hybrid cotton has great potential regarding yield and quality (Meredith and Bridge, 1972; Galanopoulou-Sendouca and Roupakias, 1999; Wu *et al.*, 2004). Although heterosis has been used successfully by breeders in hybrid cotton production, its molecular genetic basis is still unclear. Subsequent to George H. Shull rediscovering heterosis in 1908, scientists have proposed many hypothetical genetic mechanisms, including dominance, overdominance, and epistasis, although no single mechanism can adequately explain all aspects of the heterosis (Shull, 1908; Bruce, 1910; Jones, 1917; East, 1936; Richey, 1942; Powers, 1944; Crow, 1948; Jinks and Jones, 1958).

The general combining ability (GCA) of a line and the specific combining ability (SCA) of one hybrid combination were identified by Sprague and Tatum (1942). The GCA of a line is the average performance of hybrid combinations and is a very important factor for the selection of appropriate parents. Analysis of GCA also helps to identify promising cross combinations for hybrid breeding (Zhao *et al.*, 2016; Giraud *et al.*, 2017; Larièpe *et al.*, 2017; Zhou *et al.*, 2017; Werner *et al.*, 2018). Meanwhile, SCA is used to designate those cases in which certain combinations perform relatively better or worse than would be expected based on the GCA of the lines involved. SCA has been employed in the selection of specific combinations in hybrid breeding. Previous studies have demonstrated that GCA generally consists of additive and additive-by-additive effects, and SCA involves dominant and epistatic effects (Reif *et al.*, 2007). Therefore, exploration of the genetic mechanisms underlying GCA and SCA holds practical importance in hybrid cotton breeding.

In recent years, there have been several studies involving the genetic mapping of heterotic loci with molecular markers (Liu *et al.*, 2012; Guo *et al.*, 2013; Liang *et al.*, 2015; Shang *et al.*, 2015, 2016a,b,c; Wen *et al.*, 2015). However, as a result of the low genetic diversity of the mapping populations and low marker density, relatively few genetic loci associated with heterosis have been identified through the quantitative trait locus (QTL) mapping of cotton. Rapid developments in genome sequencing technology have resulted in the application of single nucleotide polymorphism (SNP) markers, which are characterized by low mutation rates, considerable abundance, and high accuracy for association analyses over traditional molecular markers. Additionally, genome-wide association studies (GWAS), which can reveal natural allelic variations, have been widely employed to explore the genetic loci and candidate genes responsible for agronomic traits in diverse plant species (Huang *et al.*, 2010, 2011; Kump *et al.*, 2011; Meijón *et al.*, 2013). In cotton, GWAS have been extensively used to dissect the genetic mechanism underlying flowering time, fiber quality, and yield traits (Islam *et al.*, 2016; Li *et al.*, 2016; Su *et al.*, 2016a,b; Fang *et al.*, 2017; Huang *et al.*, 2017; Shen *et al.*, 2017; Sun *et al.*, 2017; Wang *et al.*, 2017; Du *et al.*, 2018; Ma *et al.*, 2018). Nevertheless, GWAS based on SNP markers involving large cross populations for heterosis in cotton have not been reported. In the present study, to determine the genetic basis of the GCA and SCA in cotton, we constructed one population by crossing 282 female parents with four male parents and analyzed the GCA and SCA for boll weight (BW), lint percentage (LP), and six fiber quality traits. We performed GWAS by integrating the genotypic data of the female parents obtained by restriction site-associated DNA sequencing (RAD-seq) and deduced F_1 genotypes with the phenotypic data, GCA, and SCA values. We further analyzed the cumulative effect of favorable haplotypes in female parents. Thus, the methods used in the present study represent a large-scale approach for the evaluation of the effects of GCA and SCA for upland cotton parents and hybrid crosses. The detected selective SNPs of the GCA and SCA may ultimately be used to determine the biological and genetic factors related to combining ability.

RESULTS

Characterization and distribution of SNPs in the upland cotton genome

The 282 female and four male parents were genotyped by RAD-seq. In total, 306 814 filtered SNPs were detected based on a missing-data rate < 20% and a minor allele frequency > 5%. There was an average of 11 549 SNPs on each chromosome, with 189 261 SNPs in At subgenome and 111 003 SNPs in Dt subgenome. These SNPs were unevenly distributed throughout the upland cotton genome. Chromosomes A08 and D04 had the most (26 665) and fewest (5651) SNPs, respectively, and the average SNP density was 1 per 7.84 kb (Table S1). Additionally, the polymorphism information content values ranged from 0.342 to 0.393, whereas the gene diversity values ranged from 0.400 to 0.456 among chromosomes.

Population structure

To assess the genetic differences between parental lines, a neighbor-joining tree was constructed according to Nei's standard genetic distance. Phylogenetic analysis revealed that the 286 parental lines could be clustered into five subgroups, namely Group I, Group II, Group III-1, Group III-2, and Group III-3, which contained 56, 56, 34, 74, and 66 accessions, respectively (Figure 1a and Table S2). Based on first three axes of principal component analysis, Group I and Group II were distinguished from other accessions, which was consistent with the results of the neighbor-

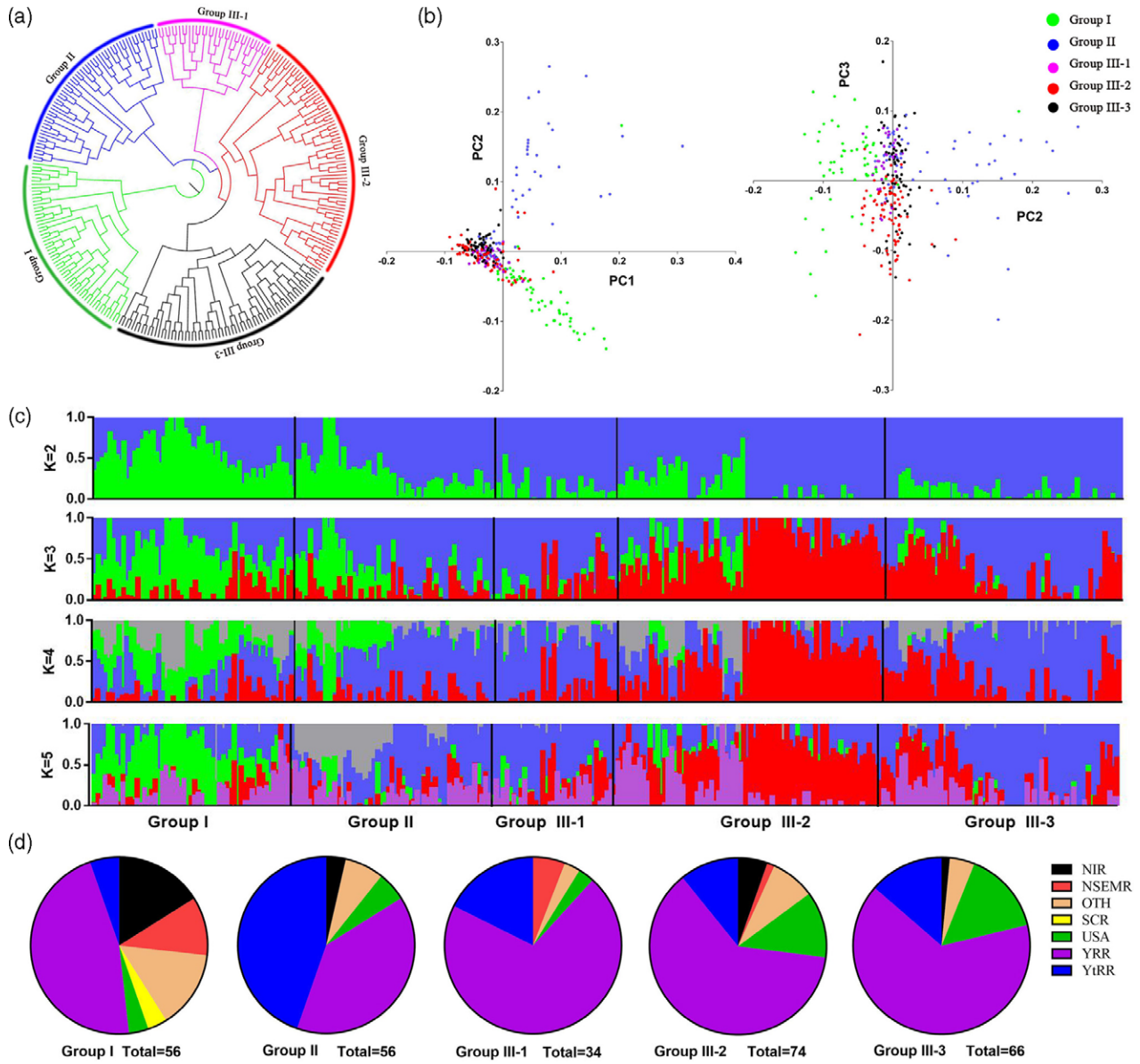


Figure 1. The population structure and geographic origin of parental lines. (a) A neighbor-joining tree of all parent lines. Different groups are represented by different colors. (b) Plots of the first three principal components of 286 parental lines using single nucleotide polymorphisms. (c) Population structure of 286 parental lines based on STRUCTURE from $k = 2$ to $k = 5$. (d) Geographic origin of the parental lines classified into five groups. Different geographic origins are represented by different colors. NIR, the Northwestern Inland Region; NSEMR, the Northern Specific Early Maturation Region; OTH, other countries; SCR, the Southern China Region; YRR, the Yellow River Region; YtRR, the Yangtze River Region.

joining analysis (Figure 1b and Figure S1). Population structure analysis revealed that the parental lines could be classified into five subgroups (Figure 1c). Next, we analyzed the geographic origins of the five subgroups. We determined that, in Group I, most of the accessions were from the Yellow River Region (YRR) (26; 46.4%), although there were also some accessions from the Northwestern Inland Region (NIR) (9; 16.1%) and the Northern Specific Early Maturation Region (NSEMR) (6; 10.7%). In Group II, most of the accessions were from the Yangtze River Region (YtRR) (25; 44.6%) and YRR (22; 39.3%). In Group III-1, Group III-2, and Group III-3, most of the accessions were from the YRR (Group III-1: 34; 70.6%; Group III-2: 74; 62.2%; and Group III-2: 66; 65.2%, respectively) (Figure 1d). The kinship (*K*) matrix is one of the important factors for GWAS. The mean pairwise relative kinship coefficient was 0.467, ranging from 0 to 1.92. In addition, kinship values < 0.5 accounted for 67.51% of all pairwise kinship coefficients (Figure S2). This result suggested that the majority of accessions were unrelated in the present study.

General and specific combining ability performance

Descriptive statistics for eight fiber quality- and yield-related traits of the female parent, F_1 hybrids, and GCA values are presented in Table S3. Significant variations ($P < 0.001$) were identified among the males, females, and males × females for all eight traits analyzed (Table 1). Four of the traits analyzed, including BW, fiber length (FL), LP, and micronaire (MIC), had the higher broad-sense heritability (0.51–0.75), indicating that these traits were mainly controlled by genotype. However, fiber strength (FS), fiber uniformity (FU), fiber elongation (FE), and spinning consistency index (SCI) had lower broad-sense heritability (0.34–0.49), suggesting that environment greatly effects the performance of these traits.

The GCA performance of 282 female parents that divided into five subgroups is presented in Figure 2. Among the five subgroups, Group III-2 had the highest GCA values for BW and MIC; Group III-3 had the highest GCA values for FE and SCI; and Group I had the lowest GCA values for FL and LP. Consequently, the GCA values of Group III-3 and Group III-2 were greater than those of Group I. There were no significant differences among the five subgroups regarding the GCA values for FS and FU.

The GCA performance of 282 female parents from six cotton-growing regions evaluated for eight fiber quality- and yield-related traits (the SCR region, which only has two accessions, was eliminated) is presented in Figure S3. Female parents cultivated in the YtRR and YRR showed the highest GCA values for BW and LP. Female parents in the NIR showed the highest GCA values for FU and FS, whereas female parents in YRR exhibited the highest GCA values for MIC. The GCA values for FL, FE, and SCI were

Table 1 Variance and genetic analysis of the North Carolina II population

	Mean squares									
Trait	Males	Females	Males × females	Environments	Hybrids × environments	σ^2_m	σ^2_f	σ^2_{mf}	h^2	H^2
FS	148.44****	19.51****	4.96****	1134.80****	4.85****	0.13	0.91	0.64	0.34	0.54
SCI	13659.43****	885.39****	378.72****	36934.58****	159.86****	11.77	31.67	18.39	0.34	0.49
FU	180.91****	37.26****	35.98****	118.72****	1.49****	0.13	0.08	7.14	0.23	0.78
MIC	26.38****	0.67****	0.12****	55.36****	0.21****	0.02	0.03	0.01	0.46	0.52
FE	2.93****	0.26****	0.16****	37.24****	0.06****	0.01	0.01	0.01	0.18	0.49
FL	239.81****	11.34****	4.40****	618.93****	1.79****	0.21	0.43	0.73	0.33	0.71
BW	119.97****	1.31***	0.21**	68.28****	0.46****	0.11	0.07	0.01	0.56	0.59
LP	2465.78****	158.64****	34.03****	9039.87****	8.26****	2.16	7.79	2.70	0.60	0.76

σ_m^2, additive genetic variance of male parents, σ_f^2, additive genetic variance of female parents, σ_{mf}^2, non-additive genetic variance of male parent × female parents, h^2, narrow-sense heritability, H^2, broad-sense heritability; FS, fiber strength; SCI, spinning consistency index; FU, fiber uniformity; MIC, micronaire; FE, fiber elongation; FL, fiber length; BW, boll weight; LP, lint percentage.
**$P < 0.01$,
***$P < 0.001$ and,
****$P < 0.0001$ significant, respectively.

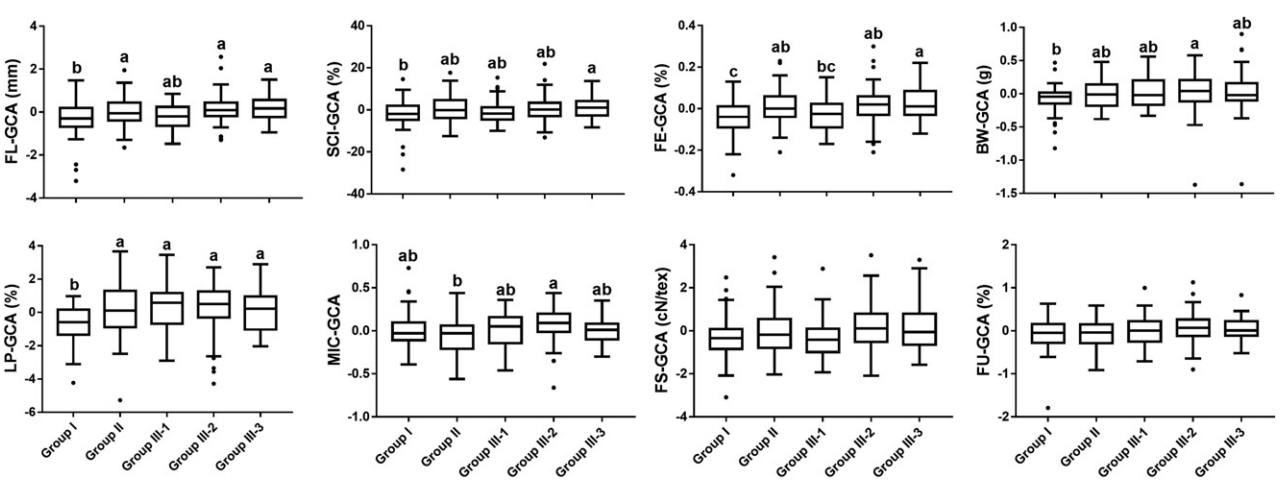


Figure 2. Comparison of the general combining ability (GCA) of female parents divided into different groups. The mean GCA value was compared using one-way analysis of variance followed by a Tukey's multiple comparisons test. Different letters indicate a significant difference among groups ($P < 0.05$).

not significantly different among the analyzed cotton-growing regions.

To help breeders select elite parents for hybrid breeding, we identified the accessions with GCA values within the top and bottom 5% for each analyzed trait (Table S4). Our data revealed that 20 and 19 accessions had a GCA value within the top and bottom 5%, respectively, for more than one trait. Additionally, 12 accessions had both top 5% and bottom 5% GCA values for more than one trait. Moreover, 33 accessions had a top 5% GCA value for only one trait and 28 accessions had a bottom 5% GCA value for only one trait. Therefore, these results suggest that the 20 accessions with a top 5% GCA value for more than one trait are appropriate parents for hybrid breeding. Especially, ZhongR014121, Su9108R03, SGK9708 (yuan), ZhongZi4480, and Hongtao had top 5% GCA values for both fiber yield- and quality-related traits.

SCA is a very important indicator during the selection of superior parents for hybrid cotton breeding. In total, 277 crosses had a preferred SCA for both fiber yield traits (BW and LP) and 88 crosses had a preferred SCA for both fiber quality traits. Finally, we selected only 19 F_1 hybrids with positive SCA values for both fiber yield and quality-related traits, except for MIC (Table S5). Interestingly, we found that, in these 19 F_1 hybrids, each female parent can produce a superior F_1 with just one male parent. Our results demonstrated that fiber yield and quality traits have negative correlations indicating that SCA is a complex trait.

Genomic dissection of the GCA differences among Group I and III-2, as well as the remaining groups

The accessions in Group I and III-2 had substantially different GCA values for BW, LP, FL, MIC, FE, and SCI (Figure 2). To dissect the underlying genomic mechanism, we compared the population fixation statistics (*Fst*) of Group I and III-2, as well as the remaining groups (Figure S4). A highly divergent genomic region between Group I and the remaining groups was detected on chromosome A06 (77.2–115.4 Mb) and this region contained 394 genes (*Gh_A06G138600*–*Gh_A06G177900*). Moreover, two highly divergent genomic regions between Group III-2 and the remaining groups were detected on chromosomes A02 (95.7–98.4 Mb) and A07 (33.6–33.7 Mb). These two regions comprised 95 (*Gh_A02G158400*–*Gh_A02G167800*) and five genes (*Gh_A07G155200*–*Gh_A07G155600*), respectively. Details regarding the *Fst* values exceeding the threshold (top 5% *Fst* values) are provided in Table S6.

GWAS of the phenotype and general combining ability of fiber-related traits

To characterize the genetic basis of the GCA in our population, we conducted single-locus and multi-locus GWAS of the female parent phenotypes and the GCA in four different environments.

The single-locus GWAS for the female parent phenotypes identified 740 significant SNPs, and 133 common SNPs can also be identified by multi-locus GWAS. These associated SNPs were distributed in 422 QTLs. The results of the GWAS of the female parent phenotype, including the significant SNPs and QTL regions, are summarized in Tables S7–S10. The FL and BW traits had more significant SNPs than the other traits (234 and 102, respectively), whereas the BW and FS had the highest proportion of common SNPs (50.0 and 26.9%, respectively).

From the single-locus GWAS of GCA, 120 significant SNPs were detected by EMMAX (Kang *et al.*, 2010), and 24 SNPs were identified by both single-locus and multi-locus methods. These associated SNPs which located in one linkage disequilibrium (LD) ($r^2 > 0.6$) region were therefore assigned to the same QTL, resulting in 66 unique QTLs. Furthermore, 29 QTLs were also detected by an association analysis involving the female parent phenotypic data (Table 2). Detailed information regarding the results of the GWAS of GCA, including the significant SNPs and QTLs, is provided in Tables S11–S13. Most of the significant SNPs for the important fiber quality- and yield-related traits, namely FS, SCI, and BW, were located on chromosomes A07, and A10, respectively.

Fiber strength. The FS_GCA was associated with the most SNPs (42), as identified by EMMAX (Tables S11 and S12). These 42 associated SNPs were distributed in 18 QTL regions, including 11 QTLs that were also identified by GWAS with the female parent phenotype. These 11 QTLs were located on chromosomes A02, A07, A08, A09, A10, and A13 (Table 2).

We identified 22 and 16 significant SNPs on chromosome A07 for FS_GCA and SCI_GCA, respectively (Figure 3a). We selected eight SNPs to investigate the allelic variation. Most of the accessions (164; 84.5%) carried one homozygous haplotype (GAGTCGAC) and had the lowest FS_GCA and SCI_GCA values (Figure 3b). Only one accession (Chuan R128) carrying another homozygous haplotype (AGTCTAGT) had the highest FS_GCA and SCI_GCA values. The remaining accessions carrying the heterozygous haplotype had a moderate GCA value. The LD heatmap revealed a high level of LD between these SNP markers (90.44–90.66 Mb) (Figure 3c). Seven candidate genes (*Gh_A07G218600*–*Gh_A07G219200*) were located in this region, and we analyzed their expression patterns based on published transcriptomic data (Figure 3d) (Zhang *et al.*, 2015). Based on the cotton gene expression patterns and the functional annotation of Arabidopsis homologs, we identified *Gh_A07G218800* as a candidate gene for FS_GCA. This gene was identified previously as a candidate gene for FS (Sun *et al.*, 2017; Ma *et al.*, 2018).

Table 2 The information of 66 QTLs identified for the GCA value of the female parents

Trait	QTL Name	LD block (bp)	Gene region	Number of significant SNPs in LD	Overlapped QTLs	References
BW_GCA	*qGhBW-A02-1*	36314216-37307418	Gh_A02G110800-Gh_A02G111000	1		
BW_GCA	*qGhBW-A02-2*	38974873-39878942	Gh_A02G111600-Gh_A02G111700	1		
BW_GCA	*qGhBW-A04-1*	43834912-43891532	Gh_A04G075100-Gh_A04G075200	1		
BW_GCA	*qGhBW-A05-1*	22580669-22717672	Gh_A05G210100-Gh_A05G210800	2	*qLY-chr5-2, qBW-chr5-2*	Liang *et al.* (2015)
BW_GCA	*qGhBW-A07-1*	91226889-91265421	Gh_A07G222500	1		
BW_GCA	***qGhBW-A10-1***	**112765425-112981991**	**Gh_A10G238400-Gh_A10G239400**	**6**		
BW_GCA	***qGhBW-A10-2***	**113066225-113254784**	**Gh_A10G239800-Gh_A10G240500**	**2**		
BW_GCA	***qGhBW-A10-3***	**113288542-113687829**	**Gh_A10G240800-Gh_A10G243000**	**10**		
BW_GCA	*qGhBW-A12-1*	98142203-98605752	Gh_A12G224700-Gh_A12G229900	1		
BW_GCA	*qGhBW-A13-1*	1649728-2024525	Gh_A13G015900-Gh_A13G020100	1		
BW_GCA	*qGhBW-A13-2*	11659839-11697379	Gh_A13G060300-Gh_A13G060400	1		
BW_GCA	*qGhBW-A13-3*	11749494-12039086	Gh_A13G060500-Gh_A13G061400	2		
BW_GCA	*qGhBW-A13-4*	24795225-25064859	Gh_A13G085000-Gh_A13G085100	1		
BW_GCA	*qGhBW-D01-1*		Gh_D01G146300-Gh_D01G146400	1		
BW_GCA	*qGhBW-D06-1*	52115527-52191643	Gh_D06G164900	1		
BW_GCA	*qGhBW-D08-1*	42750286-43285065	Gh_D08G127600-Gh_D08G129000	1		
BW_GCA	*qGhBW-D09-1*	12177810-12561027	Gh_D09G042600-Gh_D09G043200	1		
BW_GCA	*qGhBW-D12-1*	7086327-7226072	Gh_D12G048000-Gh_D12G048500	1		
FE_GCA	*qGhFE-A07-1*	34385684-34763876	Gh_A07G156900-Gh_A07G157600	1		
FE_GCA	*qGhFE-D09-1*	34098554-34506944	Gh_D09G093800-Gh_D09G097600	1		
FL_GCA	*qGhFL-A01-2*	112180491-112330857	Gh_A01G228000 -Gh_A01G228700	1	*qSY-Chr1-3, qLY-Chr1-4,*	Shang *et al.* (2015, 2016a)
FL_GCA	***qGhFL-A03-1***	**65110366-65476635**	**Gh_A03G124000-Gh_A03G124500**	**1**		
FL_GCA	***qGhFL-A08-1***	**92536181-92856211**	**Gh_A08G136000-GhA08G136500**	**1**		
FL_GCA	***qGhFL-D01-1***	**16004212-16312096**	**Gh_D01G108700-Gh_D01G110300**	**1**		
FL_GCA	***qGhFL-D05-1***	**36382943-36841968**	**Gh_D05G312200-Gh_D05G313300**	**1**	***GhHOX3***	Shan *et al.* (2014)
FL_GCA	***qGhFL-D13-1***	**60613810-61023461**	**Gh_D13G235000-Gh_D13G236400**	**2**		
FS_GCA	*qGhFS-A01-1*	62640835-63115960	Gh_A01G154600-Gh_A01G155000	1		
FS_GCA	***qGhFS-A02-1***	**68792315-69870329**	**Gh_A02G135300-Gh_A02G135500**	**1**		
FS_GCA	*qGhFS-A07-1*	35732932-35968544	Gh_A07G159100-Gh_A07G159500	1		
FS_GCA	***qGhFS-A07-2***	**88713289-88892162**	**Gh_A07G213700-Gh_A07G214200**	**1**		
FS_GCA	***qGhFS-A07-3***	**90156393-90392991**	**Gh_A07G217400-Gh_A07G218100**	**6**		
FS_GCA	***qGhFS-A07-4***	**90437372-90674997**	**Gh_A07G218600-Gh_A07G219200**	**14**	**i39753Gh, i02033Gh, i02034Gh, i02035Gh, i02037Gh, i49171Gh**	Sun *et al.* (2017)
FS_GCA	***qGhFS-A08-1***	**84041559-84110801**	**Gh_A08G126000-Gh_A08G126100**	**1**		
FS_GCA	***qGhFS-A08-2***	**108870379-109236639**	**Gh_A08G174600-Gh_A08G175600**	**6**		
FS_GCA	***qGhFS-A09-1***	**8317620-8541181**	**Gh_A09G032700-Gh_A09G032900**	**1**		
FS_GCA	***qGhFS-A09-2***	**61983739-62093616**	**Gh_A09G104700-Gh_A09G104900**	**2**		
FS_GCA	***qGhFS-A09-3***	**63398382-63602138**	**Gh_A09G111400-Gh_A09G111600**	**1**		
FS_GCA	***qGhFS-A10-1***	**12819189-13174438**	**Gh_A10G071000-Gh_A10G072000**	**1**		
FS_GCA	*qGhFS-A01-2*	107527868-107580384	Gh_A10G208300-Gh_A10G208700	1		
FS_GCA	*qGhFS-A01-3*	114127431-114160738	Gh_A10G247300-Gh_A10G247500	1		
FS_GCA	*qGhFS-A12-1*	57972334-58421653	Gh_A12G101800-Gh_A12G101900	1		
FS_GCA	***qGhFS-A13-1***	**87409145-87518532**	**Gh_A13G142300-Gh_A13G142500**	**1**		
FS_GCA	*qGhFS-D05-1*	1529063-1594392	Gh_D05G016900-Gh_D05G017500	1		
FS_GCA	*qGhFS-D09-1*	34098554-34506944	Gh_D09G093800-Gh_D09G097600	1		
FU_GCA	*qGhFU-D09-1*	14677970-15383743	Gh_D09G046700-Gh_D09G047300	1		
LP_GCA	***qGhLP-A02-1***	**100804120-100904217**	**Gh_A02G173200-Gh_A02G168000**	**1**		
LP_GCA	*qGhLP-A05-1*	31182585-31270123	Gh_A05G260600-Gh_A05G260900	1		
LP_GCA	*qGhLP-A06-1*	86248762-86351579	Gh_A06G144600-Gh_A06G144700	1		

(continued)

Table 2. (continued)

Trait	QTL Name	LD block (bp)	Gene region	Number of significant SNPs in LD	Overlapped QTLs	References
LP_GCA	*qGhLP-A10-1*	27272043-27703520	Gh_A10G100600-Gh_A10G101100	1		
MIC_GCA	*qGhMIC-A03-1*	7882492-7982203	Gh_A03G052300-Gh_A03G052700	1		
MIC_GCA	***qGhMIC-A05-1***	**11352803-11368489**	**Gh_A05G106800-Gh_A05G106900**	**1**	***GhWRKY40***	Wang *et al.* (2014)
MIC_GCA	***qGhMIC-A05-2***	**58313941-58641273**	**Gh_A05G311900**	**1**		
MIC_GCA	***qGhMIC-A10-1***	**113066225-113254784**	**Gh_A10G239800-Gh_A10G240500**	**1**		
MIC_GCA	***qGhMIC-A10-2***	**113288542-113687829**	**Gh_A10G240800-Gh_A10G243000**	**1**		
MIC_GCA	*qGhMIC-D03-1*	51432219-51778895	Gh_D03G178800-Gh_D03G180800	1		
MIC_GCA	*qGhMIC-D06-1*	44821938-44895871	Gh_D06G147100-Gh_D06G147300	1		
SCI_GCA	***qGhSCI-A01-1***	**48004360-48437775**	**Gh_A01G145500-Gh_A01G145600**	**1**		
SCI_GCA	*qGhSCI-A05-1*	101514803-101897212	Gh_A05G378600-Gh_A05G380900	1		
SCI_GCA	***qGhSCI-A07-1***	**88713289-88892162**	**Gh_A07G213700-Gh_A07G214200**	**1**		
SCI_GCA	***qGhSCI-A07-2***	**90156393-90392991**	**Gh_A07G217400-Gh_A07G218100**	**3**		
SCI_GCA	***qGhSCI-A07-3***	**90437672-90674997**	**Gh_A07G218500-Gh_A07G219200**	**12**		
SCI_GCA	***qGhSCI-A10-1***	**12819189-13174438**	**Gh_A10G071000-Gh_A10G072000**	**1**		
SCI_GCA	*qGhSCI-A12-1*	57972334-58421653	Gh_A12G101800-Gh_A12G101900	1		
SCI_GCA	*qGhSCI-D05-1*	1529063-1594392	Gh_D05G016900-Gh_D05G017500	1		
SCI_GCA	*qGhSCI-D09-1*	34098554-34506944	Gh_D09G093800-Gh_D09G097600	1		
SCI_GCA	*qGhSCI-D11-1*	19280791-19592325	Gh_D11G182200-Gh_D11G183600	1		

Bold indicates the 29 QTLs that were identified both for the GCA and the phenotype trait. QTL, quantitative trait locus; SNP, single nucleotide polymorphism; GCA, general combining ability; BW, boll weight; FE, fiber elongation; FL, fiber length; FS, fiber strength; FU, fiber uniformity; LP, lint percentage; MIC, micronaire; SCI, spinning consistency index.

Another QTL for FS_GCA was *qGhFS-A08-2*, which contained six associated SNPs (Figure S5a and Table S12). An investigation of the haplotype block structure around these SNPs revealed that this haplotype block was from 108.87 to 109.24 Mb and contained 21 SNPs and seven genes (Figure S5b). The female parents included six haplotypes with three SNPs. All accessions with haplotype TGC were known as high-quality upland cotton cultivars, and the average FS_GCA of haplotype TGC was 2.31, which was significantly greater than the corresponding values for the other haplotypes (Figure S5c). Among those genes, *Gh_A08G174600*, which encodes pinoresinol reductase 1, was highly expressed in fibers at 20 and 25 days post-anthesis (DPA) (Figure S5d and Table S12). The Arabidopsis homolog of this gene is *AtPRR1*, which encodes a pinoresinol reductase involved in the lignin biosynthesis pathway during secondary cell wall biosynthesis (Nakatsubo *et al.*, 2008; Zhao *et al.*, 2015).

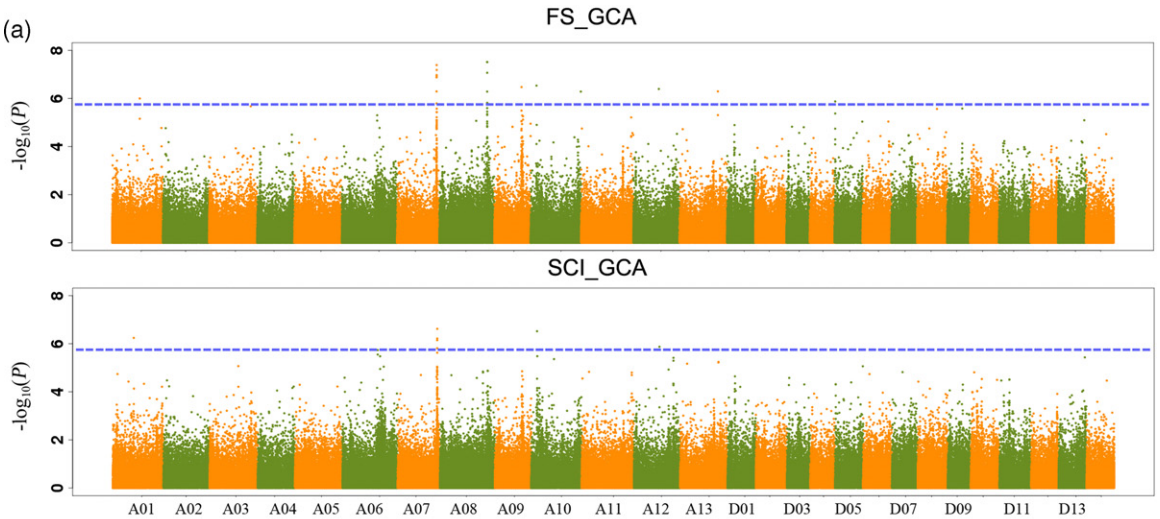
Fiber length. For the FL_GCA, we identified six significant SNPs that were located on chromosomes A01, A03, A08, D01, D05, and D13. One of these QTLs, *qGhFL-D05-1*, contained 12 genes (Table 2). The *Gh_D05G313300* gene encodes the homeobox-leucine zipper protein HOX3, which controls cotton fiber elongation (Shan *et al.*, 2014).

Another QTL, *qGhFL-D13-1*, contained 35 SNPs, and the associated haplotype block (60.61–61.02 Mb) consisted of 15 genes (*Gh_D13G235000*–*Gh_D13G236400*) (Figure S6 and Table S12). These 15 genes encode 2-oxoglutarate and Fe (II)-dependent oxygenase superfamily proteins, and one of these genes, *Gh_D13G236000*, was highly expressed in the ovules and fibers at 20 and 25 DPA (Table S12). A previous study identified three 2-oxoglutarate-dependent dioxygenase genes [*AOP1* (At4g03070), *AOP2* (At4g03060), and *AOP3* (At4g03050)] in the *GS-AOP* locus (Kliebenstein *et al.*, 2001). The *AOP2* and *AOP3* genes, which encode proteins that catalyze the conversion of methylsulfinylalkyl glucosinolates to either alkenyl or hydroxypropyl glucosinolate, are apparently the result of a gene duplication event. The *AOP1* gene has not been functionally characterized.

Spinning consistency index. In total, 23 SNPs were detected associated with SCI_GCA, and 16 SNPs were located on chromosome A07 (Table S12). Thirteen common SNPs on chromosome A07 were related to FS_GCA and SCI_GCA across the multiple environments.

Boll weight. Regarding the BW_GCA, we identified 35 significant SNPs and, among them, 18 SNPs were located on chromosome A10 and all associated SNPs were distributed in 18 QTLs (Table 2). One of the QTL for BW_GCA was *qGhBW-A10-3*, which contained 10 associated SNPs located on *Gh_A10G241400*, with nine of them being nonsynonymous SNPs (Figure 4a,b and Table S12). All

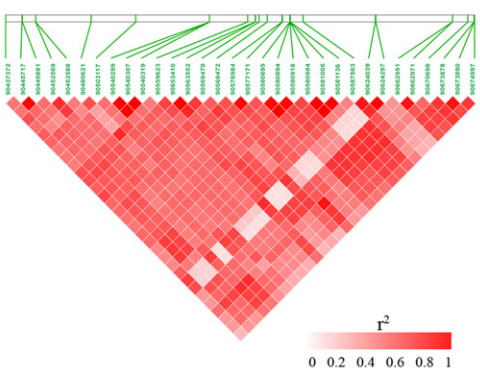
(a)



(b)

	Hap1	Hap2	Hap3	Hap4	Hap5	Hap6	Hap7	Hap8
A07_90563410	G	R	G	R	R	R	R	A
A07_90563552	A	R	A	R	R	R	R	G
A07_90569472	G	G	K	K	K	K	T	T
A07_90576984	T	Y	Y	Y	C	Y	Y	C
A07_90577171	C	Y	Y	Y	T	Y	Y	T
A07_90580984	G	R	R	R	R	A	A	A
A07_90581005	A	R	R	R	R	G	G	G
A07_90581136	C	Y	Y	Y	Y	T	T	T
Material number	164	4	4	10	3	5	3	1
FS_GCA	-0.20	-0.18	1.48	0.62	1.87	1.13	1.30	3.17
SCI_GCA	-1.24	-0.24	8.13	3.12	9.69	7.48	8.54	16.55

(c)



(d)

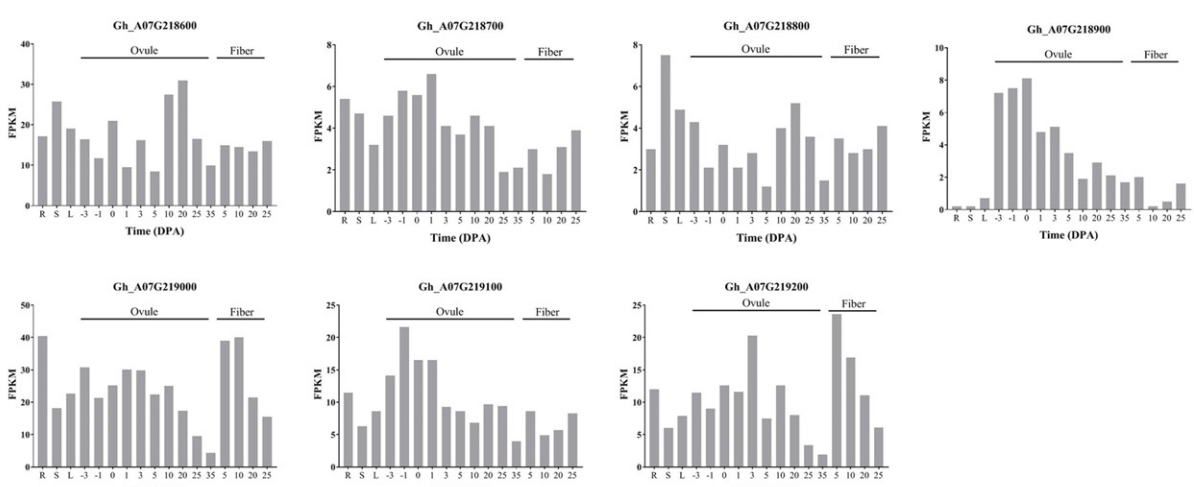


Figure 3. The associated single nucleotide polymorphisms (SNPs) and candidate genes for FS_GCA and SCI_GCA on chromosome A07. (a) Manhattan plots for the results of the genome-wide association studies of FS_GCA and SCI_GCA. The significance threshold is indicated by the blue dashed line. (b) Haplotypes observed in maternal accessions with eight SNPs and the difference of the general combining ability (GCA) value of fiber strength (FS) and spinning consistency index (SCI) among eight haplotypes. (c) Linkage disequilibrium (LD) pattern surrounding the peak on chromosome A07. (d) Transcriptomic patterns of associated genes located in the LD block of (B), based on the number of FPKM (fragments per kilobase of transcript per million mapped reads). DPA, day post-anthesis; R, S, and L represent root, stem, and leaf, respectively.

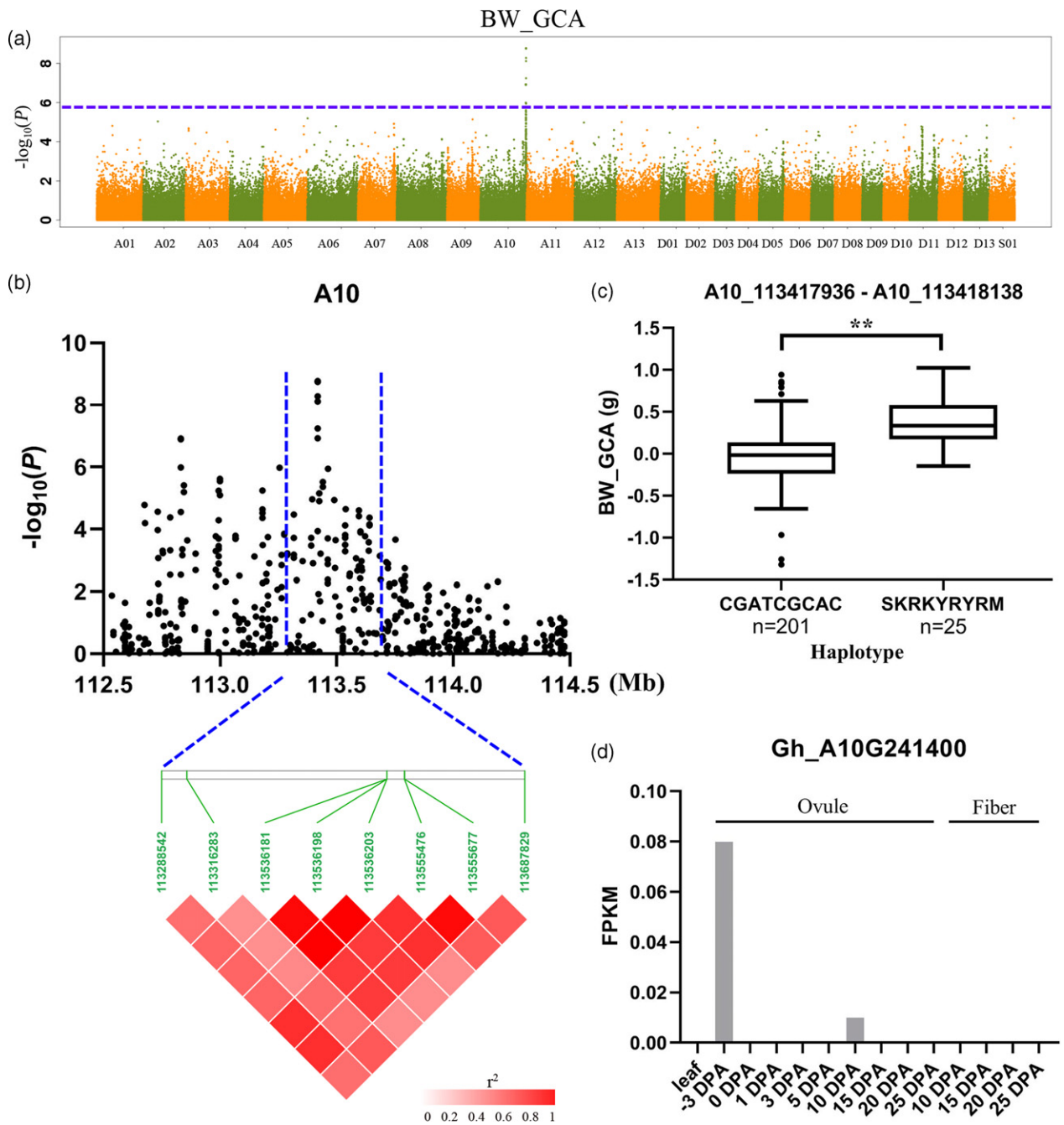


Figure 4. Identification of the candidate gene for BW_GCA on chromosome A10. (a) Manhattan plots for the results of the genome-wide association studies of BW_GCA. (b) Linkage disequilibrium (LD) heat map surrounding the single nucleotide polymorphisms (SNPs) estimated on chromosome A10. (c) Performance of BW_GCA for two haplotypes of associated SNPs in female parents (**$P < 0.01$, two-tailed *t*-test). (d) Transcriptomic pattern of the candidate gene located in the LD block based on the number of FPKM (fragments per kilobase of transcript per million mapped reads). DPA, day post-anthesis; R, S, and L represent root, stem, and leaf, respectively.

accessions with haplotype 'SKRKYRYRM' were known as high-yielding upland cotton cultivars. The average BW_GCA of haplotype 'SKRKYRYRM' was 0.36, which was significantly greater than the corresponding values for the other haplotypes (Figure 4c). *Gh_A10G241400*, which encodes disease resistance protein, was highly expressed in −3 and 10 DPA ovules and may be involved in fiber initiation and elongation (Figure 4d).

Lint percentage. For the LP_GCA, four significant SNPs were detected and distributed in four QTLs located on chromosomes A02, A05, A06, and A10 (Table 2 and

Figure S7). The candidate gene for *qGhLP-A05-1* was *Gh_A05G260800*, which encodes an Agamous-like MADS-box protein (AGL11) and was highly expressed in the ovules and fibers at various growth stages, although it was expressed at lower levels in the roots, stems, and leaves. This observation suggested that this gene may influence fiber initiation and elongation.

Micronaire. We identified seven significant SNPs and only one non-synonymous SNP (A10_113421252) for MIC_GCA (Table 2, Figure S7, and Table S12). For one QTL, *qGhMIC-A05-1*, the associated haplotype block (11.35–11.37 Mb) comprised two candidate genes, of which *Gh_A05G106800* encodes *GhWRKY40*. This gene was highly expressed in fibers at 25 DPA. A previous study found that *GhWRKY40* was induced by salicylic acid, methyl jasmonate, and ethylene and is involved in wound- and pathogen-induced responses (Wang *et al*., 2014).

Fiber elongation. For the FE_GCA, two significant SNPs were identified, including one SNP on the promoter of *Gh_A07G157100* (Table 2 and Figure S7). This gene was highly expressed in 20 and 35 DPA ovules, and may contribute to fiber elongation.

Fiber uniformity. For the FU_GCA, we identified only one significant SNP on chromosome D09. This SNP was located in the LD block from 14.68 to 15.38 Mb and contained seven genes. One of these genes, *Gh_D09G046700*, was highly expressed in −3, 0, 10 and 25 DPA ovules and may contribute to fiber uniformity (Table 2 and Figure S7).

Identification of SNPs associated with specific combining ability

For eight analyzed traits, 62 SNPs were identified in four F_1 populations by the single-locus GWAS method and 12 SNPs were also identified by the multi-locus GWAS method. Among these 62 SNPs, 11, 10, 13, and 28 SNPs were detected in the F_1 populations of A, C, D, and E, respectively. Tables S14 and S15 show the SNPs detected in different F_1 populations. Among the 11 SNPs detected in the F_1 populations A, four, two, one, two, and two SNPs were associated with BW, FE, LP, MIC, and SCI, respectively. Among the 10 SNPs detected in the F_1 populations C, two, three, one, and four SNPs were associated with BW, FL, LP, FS, and MIC, respectively. Among the 13 SNPs detected in the F_1 populations D, two, one, four, five, and one SNPs were associated with BW, FE, FU, LP, and MIC, respectively. Additionally, among the 28 SNPs detected in the F_1 populations E, 16 SNPs were associated with FE and the other 12 SNPs were associated with FL, FS, FU, LP, MIC, and SCI. Only two SNPs (A05_22626996 and A05_22627012) could simultaneously be detected for both of GCA and SCA of BW, which indicates that the genetic basis of GCA and SCA is different.

Pleiotropic effects of GCA loci

In the present study, we detected nine pleiotropic regions, including six pleiotropic regions for FS_GCA and SCI_GCA, two pleiotropic regions for BW_GCA and MIC_GCA, and one pleiotropic region for FE_GCA, FS_GCA and SCI_GCA (Table S16). Chromosome A07 occupied the largest number (3) of pleiotropic regions.

The favorable haplotypes of FS_GCA, SCI_GCA, and BW_GCA have a cumulative effect in accessions

From the results of the single-locus GWAS of GCA values, as outlined above, we identified 42, 23, and 35 significant SNPs for FS_GCA, SCI_GCA and BW_GCA, respectively. All of these SNPs have been classified into 32, 18, and 20 haplotypes and, subsequently, we identified the favorable haplotypes of these traits. To further understand the cumulative effect of favorable haplotypes, 282 female parents were grouped into four or five groups according to the GCA values of FS, SCI, and BW. We found that favorable haplotypes (FHs) accounted for a very large proportion of the accessions with higher FS_GCA values than those with lower FS_GCA values (Figure 5). Similarly, for SCI_GCA and BW_GCA, female parents carrying more FHs showed significnatly higher GCA values compared to those carrying fewer FHs. These results suggest that the genetic control of the GCA of FS, SCI, and BW exhibits a large cumulative effect in cotton.

DISCUSSION

GCA differences and the divergent genomic region between groups

The GCA values of varieties are important for selecting suitable parents, classifying heterotic groups, and breeding hybrids. A strong relationship between GCA effects and population structure was identified previously based on a maize association mapping study involving testcross data of 288 inbred lines and three testers (Larièpe *et al*., 2017). In the present study, we revealed significant differences in the GCA values for most of the cotton fiber quality- and yield-related traits between Group I, Group III-2, and the remaining groups. To clarify the underlying genomic mechanism, we compared the *Fst* values of these two groups with those of the remaining groups. Group I had high *Fst* values on chromosome A06 (77.2–115.4 Mb). We combined this finding with the results of our GWAS and identified one highly associated SNP (A06_ 86315716) for LP_GCA. Highly divergent genomic regions between Group III-2 and the remaining groups were identified on chromosomes A02 (95.7–98.4 Mb) and A07 (33.6–33.7 Mb). The candidate genes in these genomic regions should be

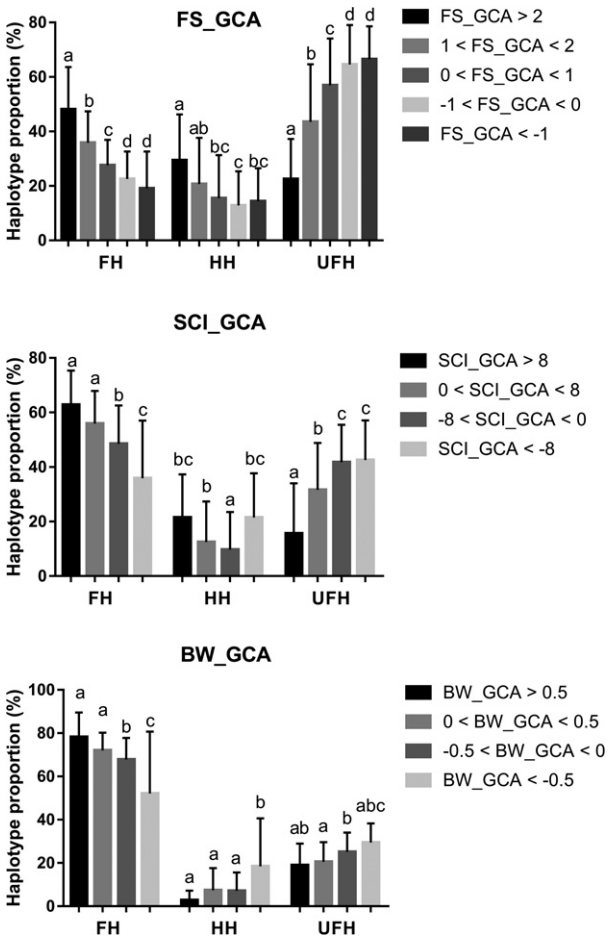


Figure 5. Haplotype proportions in FS_GCA, SCI_GCA and BW_GCA. FH, favorable haplotype; UFH, unfavorable haplotype; HH, heterozygous haplotype. The mean haplotype proportion value was compared using one-way analysis of variance followed by a Tukey's multiple comparisons test. Different letters indicate a significant difference among groups ($P < 0.05$).

identified and their molecular functions should also be characterized.

Comparison with QTLs detected in previous studies

Previous studies on the QTL mapping of heterosis in cotton involved immortalized F_2 populations, chromosome segment introgression lines, or backcross recombination lines. In the present study, 66 QTLs for eight fiber quality- and yield-related traits were detected, of which two QTLs were identified as heterotic loci in previous studies when mapping for mid-parent heterosis (Guo *et al.*, 2013; Liang *et al.*, 2015; Shang *et al.*, 2015). The stable QTLs identified across different populations may be relevant for marker-assisted selection (MAS). Additionally, four of presently identified 66 QTLs were also reported in previous studies regarding the QTL mapping of fiber quality- and yield-related traits (Qin *et al.*, 2009; Lacape *et al.*, 2010; Wang *et al.*, 2013; Zhang *et al.*, 2013; Shan *et al.*, 2014; Sun *et al.*, 2017; Ma *et al.*, 2018). One of the QTLs for FS_GCA (*qGhFS-A07-4*) localized to a previously reported QTL region. Sun *et al.* (2017) identified one QTL region on chromosome A07 (71.99–72.25 Mb) for FS, and Ma *et al.* (2018) identified *Gh_A07G1769* (*Gh_A07G218800*) as a candidate gene (Sun *et al.*, 2017; Ma *et al.*, 2018). These studies found that this region is associated with FS, although it was not identified as a pleiotropic region for FS_GCA and SCI_GCA, in contrast to our results. This pleiotropic QTL may be useful for MAS. One of the QTLs for FL_GCA, *qGhFL-D05-1*, contained one candidate gene, *Gh_D05G313300* (*GhHOX3*). This gene encodes a homeobox-leucine zipper protein, which controls cotton fiber elongation (Shan *et al.*, 2014). However, none of the previous studies on *Gh_A07G218800* and *GhHOX3* assessed whether these genes exhibit heterosis. Consequently, the heterotic alleles of these genes need to be examined. Because these two genes are not closely linked on cotton chromosomes, the allelic combination of the loci may lead to diverse cotton fiber qualities. All of the candidate genes for fiber quality and yield should be investigated more thoroughly to clarify their biological function.

Common QTLs of GCA and phenotype

The identification of significant loci for the GCA with DNA markers may improve the efficiency of hybrid predictions and provide targets for MAS during cotton hybrid breeding. In the present study, 66 stable QTLs were identified for eight traits. Moreover, 29 of these 66 QTLs (43.94%) were concurrently detected for the female parent phenotype and the GCA for BW, FL, FS, SCI, MIC, and LP. The genomic loci commonly detected for the female parent phenotype and the GCA may be explained by the high degree of correlation between the female parent phenotype and the GCA values for BW, FL, FS, SCI, MIC, and LP ($0.76 < r < 0.92$, $P < 0.05$) (Figure S8). However, 37 QTLs for the GCA were not detected for the female parent phenotype. These results are similar to those reported previously. For example, one study reported that, among 58 heterotic loci, only seven were also detected by a QTL analysis involving the data of chromosome segment introgression line population in cotton (Guo *et al.*, 2013). Another study detected 17 and 12 QTLs for yield and yield components, respectively, based on the mid-parent heterosis data for XZ and XZV hybrids (Shang *et al.*, 2015). These results indicate that the phenotype and GCA are likely controlled by two different genetic and molecular mechanisms.

Elite parents selected in the present study

We selected 20 elite accessions with top 5% GCA values in the eight analyzed traits, and we subsequently evaluated the distribution of the favorable haplotypes that we identified in these accessions. The results obtained showed that

the mean proportion of the favorable haplotypes (FH) and the hybrid haplotypes (HH) was 74.82%, ranging from 50.00 to 92.86% (Table S17). This result implied that pyramiding superior haplotypes of GCA would have a positive effect on GCA performance. Additionally, we analyzed whether these 20 accessions have been utilized in the cotton breeding program. We found that six cultivars (including Lu343, Zhong1421, Zhong1441, Zhongzi2574, CIR81, and CIR82) have been developed using SGK9708 as a parent, with PD6186 having been used as a parent to breed Han8959. Except for these two accessions, we found no evidence for the other 18 accessions having been used in cotton breeding. Consequently, these 20 accessions can be utilized in future hybrid cotton breeding.

In conclusion, the present study comprises one large-scale approach for applying high-throughput sequencing to investigate the molecular genetic basis of combining ability in cotton. The identified SNPs of combining ability may increase the efficiency of the selection of appropriate parents and superior F_1 hybrids, with possible implications for future hybrid breeding.

EXPERIMENTAL PROCEDURES

Plant materials

In the present study, 282 female parents were crossed with four male parents in accordance with the North Carolina II mating scheme to generate 1128 hybrids. All of the accessions came from the main cotton-growing regions of China [the Yangtze River Region (YtRR, 54), the Yellow River Region (YRR, 157), the Northwestern Inland Region (NIR, 16), the Southern China Region (SCR, 2), and the Northern Specific Early Maturation Region (NSEMR, 9)], as well as historically introduced varieties and germplasm resources lines from the USA (25) and other countries (OTH, 23). All of the accessions were preserved in the Gene Bank of Institute of Cotton Research of Chinese Academy of Agriculture Sciences, with detailed information being provided in Table S2.

Field experiments

All of the female parents and the F_1 hybrids were evaluated in 2012 and 2013 in the YRR and the YtRR in China. The YRR included Anyang (36°08′N, 114°48′E) and Xinxiang (35°18′N, 113°54′E), and the YtRR included Changde (29°00′N, 111°39′E) and Jingzhou (30°32′N, 112°55′E). Field experiments were arranged in a randomized complete block design with three replicates. All materials were planted in single-row plots (width 0.8 m, length 8 m). We made every effort to control the experimental error for this large-scale field experiment. First, our experiment was carried out in the field using the same fertilization as far as possible. Second, we planted two control varieties and guarding rows in each replication. Third, field management, including fertilizer application, irrigation, weed management, and insect pest control, both throughout the growing season and during harvest, was kept the same as much as possible.

Data collection and statistical analysis

Randomly selected 30 naturally opened bolls of the hybrids and parents were harvested manually. The fiber quality traits, including fiber length (FL, mm), fiber strength (FS, cN/tex), fiber elongation (FE, %), fiber uniformity (FU, %), spinning consistency index (SCI, %) and micronaire (MIC), were measured with the HVI9000 system (Uster Technologies AG, Charlotte, NC, USA) at the Supervision and Testing Center of Cotton Quality, Ministry of Agriculture, Anyang, Henan province, China. Yield component traits including boll weight (BW, g) and lint percentage (LP, %) were recorded. Descriptive statistics for eight fiber quality- and yield-related traits of the female parent and F_1 hybrids are presented in Table S3.

Analysis of variance (ANOVA) was performed with a GLM procedure in SAS, version 9.21 (SAS Institute, Cary, NC, USA). The significant genotypic variance of each trait was further partitioned to GCA, SCA, and experimental error (Hallauer *et al.,* 1981; Kearsey and Pooni, 1996). The effects of male parents, female parents, male parents × female parents, and environment were calculated using variance analysis with reference to a statistics book (Mo *et al.,*1982).

We calculated the additive genetic variance of male parents (σ^2_m), female parents (σ^2_f), non-additive genetic variance of male parents × female parents (σ^2_{mf}), genetic variance of F_1 (σ^2_G), environmental variance (σ^2_w), phenotypic variance of F_1 (σ^2_P), narrow-sense heritability (h^2), and broad-sense heritability (H^2). These parameters were calculated using: $\sigma^2_m = (MS_{males} - MS_{males \times females})/rf$; $\sigma^2_f = (MS_{female} - MS_{males \times females})/rm$; $\sigma^2_{mf} = (MS_{males \times females} - MS_{males \times females \times environments})/rn$; $\sigma^2_w = MS_{error}$; and $\sigma^2_G = \sigma^2_m + \sigma^2_f + \sigma^2_{mf}$; $h^2 = (\sigma^2_m + \sigma^2_f)/\sigma^2_P$ and $H^2 = \sigma^2_G/\sigma^2_P$, respectively.

The GCA was calculated using: $g_{i(f)} = \bar{y}_{i(f)} - \bar{y}$, where $g_{i(f)}$ is the GCA of the *i*th female parent, $\bar{y}_{i(f)}$ is the phenotypic value for the hybrid derived from the *i*th female parent, and $\bar{y}$is the mean phenotypic value for all hybrids. The SCA was calculated using: $S_{ij} = y_{ij} - \bar{y} - g_{i(f)} - g_{j(m)}$, where y_{ij} is the phenotypic value of the F_1 hybrid between the *i*th and *j*th parents, $g_{i(f)}$ is the GCA of the *i*th female parent, and $g_{j(m)}$ is the GCA of the *j*th male parent. Descriptive statistics for the GCA values for eight analyzed agronomic traits are presented in Table S3. The correlation between the female parent trait and the GCA was assessed with the 'correlation' function of PRISM, version 7.00 (GraphPad Software Inc., San Diego, CA, USA).

SNP genotyping

Genomic DNA was extracted from the fresh leaves of the 286 parental lines according to an established CTAB method (Paterson *et al.*, 1993). The purified DNA was digested with FastDigest *TaqI* (Fermentas; Thermo Scientific, Waltham, MA, USA) at 65°C for 10 min. Bar-coded adapters were ligated to the digested DNA fragments with T4 DNA ligase (Enzymatics, Beverly, MA, USA), during 1 h of incubation at 22°C. Samples were then heated at 65°C for 20 min, after which the 24 samples were pooled. The DNA fragments (400–600 bp) were purified from a 2% agarose gel with the QIA quick Gel Extraction kit (Qiagen, Valencia, CA, USA). The adapter-ligated DNA fragments were amplified via a PCR with Phusion High-fidelity DNA polymerase (Finnzymes; Thermo Scientific). The amplified fragments were separated by agarose gel electrophoresis, and the DNA fragments (400–600 bp) were purified using a QIA quick PCR Purification kit (Qiagen, Hilden, Germany). Finally, the purified libraries were quantified with a 2100 Bioanalyzer Instrument (Agilent Technologies Inc., Santa Clara, CA, USA). The libraries were sequenced using the Hiseq 2000 system (Illumina, San Diego, CA, USA). The raw reads were aligned to the *G. hirsutum* L. TM-1 reference genome (https://cottonfgd.org/about/download.html) with the 'mem -t 8' parameter of BWA (Yang *et al.*, 2019). GATK (McKenna *et al.,* 2010) and SAMTOOLS packages (Li

et al., 2009) were used for SNP calling, after which the SNPs with a high missing-data rate (> 20%) and a low minor allele frequency (< 5%) were eliminated. The generated sequencing data have been deposited into the NCBI database (accession number: PRJNA353524). The genotypes of F_1 hybrids can be deduced by the genotypes of the parents because the heterozygous SNPs in one of the two parents are scored as missing. Finally, 36 331, 15 294, 42 213, and 33 460 SNPs were deduced for F_1 populations A, C, D, and E, respectively.

Phylogenetic and population structure analyses

We performed a phylogenetic analysis of all parental lines according to a neighbor-joining statistical method involving the *P* distance of TREEBEST, version 1.9.2 (http://treesoft.sourceforge.net/treebest.shtml). The phylogenetic tree was visually edited with FIGTREE (http://tree.bio.ed.ac.uk). The population structure of parental genotypes was analyzed with STRUCTURE, version 2.3.4 (Falush *et al.,* 2003). Specifically, the number of assumed genetic clusters (*K*) ranged from 2 to 10, with 10 000 iterations for each run. Principal component analysis of the SNPs was conducted using EIGENSOFT, version 6.0.1 (Price *et al.,* 2006), and the first three principal components were used for the analysis of the genetic structure of the 286 parental lines (Figure S1). The *F*st values were calculated with VCFTOOLS, version 0.1.14 (http://vcftools.sourceforge.net) (100-kb windows sliding 20 kb with the following parameter: --window-pi 100000 --window-pi-step 20000) (Danecek *et al.*, 2011). The familiar relatedness among the parental lines was assessed by calculating a kinship matrix using the VanRaden method in TASSEL, version 5.2.14 (Bradbury *et al.,* 2007), based on the 'scaled identity by state' (VanRaden, 2008; Endelman and Jannink, 2012).

GWAS

We performed single-locus GWAS with 306 814 filtered SNP (a missing-data rate < 20% and a minor allele frequency > 5%) in EMMAX (Kang *et al.*, 2010). The *P* value threshold for significant associations was 1.63×10^{-6} ($0.5/n$); therefore, those SNPs with $-\log_{10}$ (*P*) greater than 5.79 were considered as the significant SNPs for female parent phenotype and GCA (Wang *et al.*, 2012; Yang *et al.*, 2013). The $-\log_{10}$ (*P*) thresholds for SCA of F_1 populations A, C, D, and E were 4.86, 4.49, 4.93, and 4.83, respectively. Manhattan plots and quantile-quantile plots were constructed with R script. Multi-locus GWAS were implemented using MRMLM, version 1.3, to verify the SNPs identified by single-locus GWAS. The MRMLM package, including six multi-locus GWAS methods (mrMLM, ISIS EM-BLASSO, FASTmrEMMA, pLARmEB, FASTmrMLM, and pKWmEB), is available via: http://cran.r-project.org/web/packages/mrMLM/index.html (Wang *et al.*, 2016; Tamba *et al.*, 2017; Zhang *et al.*, 2017; Ren *et al.*, 2018; Tamba and Zhang, 2018; Wen *et al.*, 2018). Default values were used for all parameters. The significant association thresholds were set to LOD = 3.0.

To define the QTL range, we split the female parent genomes into haplotype blocks using HAPLOVIEW (Barrett *et al.,* 2005) and the recombinant confidence interval method (Gabriel *et al.*, 2002). The QTL regions were determined based on the range of the corresponding haplotype blocks (Zhang *et al.*, 2015). All genomic positions provided in the present study were based on the *G. hirsutum* L. TM-1 reference genome (Yang *et al.*, 2019).

Favorable haplotype identification

We selected significant SNPs for FS_GCA, SCI_GCA, and BW_GCA to investigate the allelic variation, respectively, and those SNPs with the same allelic variation frequency were divided into one haplotype. In the present study, the favorable haplotypes were defined as the haplotypes that were shown to be beneficial for trait improvement of cotton. According to the results of the GWAS, corresponding phenotypic data of haplotypes were used to compare the genetic effect between haplotypes and haplotypes with larger trait values (except for micronaire), defined as favorable haplotypes.

ACKNOWLEDGEMENTS

We thank the Gene Bank of Institute of Cotton Research of Chinese Academy of Agricultural Sciences for providing the germplasms. This research was supported by grants from the National Key Research and Development Program of China (2016YFD0100203, 2016YFD0101401) and the National Natural Science Foundation of China (Grant No. 31571716).

AUTHOR CONTRIBUTIONS

XD conceived and designed the experiments; YJ, JS, and MSI collected materials; QW, HQ, JL, HL, JY, ZM, DX, JY, JZ, ZL, ZC, X-LZ, XZ, GZ, LL, HZ, LW, and BP contributed to phenotyping; SH and ZP performed RAD resequencing data production. GS performed GWAS and population structure analysis; XG, YQ and ZS worked on data analysis. XG wrote the paper. All authors reviewed and approved the final manuscript submitted for publication.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The RAD-seq data used in this study were submitted to the NCBI under accession number PRJNA353524. Other relevant data can be found within the manuscript and its supporting materials.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. The variance explained by the first 10 principal components.

Figure S2. Distribution of pairwise relative kinship values for 286 accessions.

Figure S3. Comparison of the GCA values of female parents cultivated in different cotton-grown regions. The mean GCA value was compared using one-way ANOVA followed by a Tukey's multiple comparisons test. Different letters indicate a significant difference among groups ($P < 0.05$).

Figure S4. Screening of genomic divergence region in two represented groups. The top panel indicates the comparisons between Group I versus the remaining accessions, the bottom panel indicates the comparisons between Group III-2 versus the remaining accessions.

Figure S5. Identification of the candidate gene for FS_GCA on chromosome A08. (a) Manhattan plots displaying the GWAS result of FS_GCA. (b) LD heat map surrounding the SNPs estimated on chromosome A08. (c) Performance of FS_GCA for three haplotypes of the significant SNP in female parents. The mean GCA

value was compared using one-way ANOVA followed by a Tukey's multiple comparisons test. Different letters indicate significant a difference among haplotypes ($P < 0.05$). (d) The transcriptomic pattern of the candidate gene located in the LD block. R, S, and L represent root, stem, and leaf, respectively.

Figure S6. Identification of the candidate gene for FL_GCA on chromosome D13. (a) Manhattan plots displaying the GWAS result of FL_GCA. (b) LD heat map surrounding the SNP estimated on chromosome D13. (c) Performance of FL_GCA for two genotypes of the significant SNP (**$P < 0.01$, two-tailed *t*-test). (d) Transcriptomic pattern of the candidate gene located in the LD block. R, S, and L represent root, stem, and leaf, respectively.

Figure S7. Summary of GWAS results for the GCA value of the eight analyzed traits. (a–d) Manhattan plots and quantile-quantile plots for FS_GCA. (e–h) Manhattan plots and quantile-quantile plots for SCI_GCA. (i–l) Manhattan plots and quantile-quantile plots for BW_GCA. (m–p) Manhattan plots and quantile-quantile plots for FL_GCA. (q–t) Manhattan plots and quantile-quantile plots for MIC_GCA. (u–x) Manhattan plots and quantile-quantile plots for LP_GCA. (y–ab) Manhattan plots and quantile-quantile plots for FE_GCA. (ac–af) Manhattan plots and quantile-quantile plots for FU_GCA.

Figure S8. Correlation (*r*) between female parent trait and the GCA for eight analyzed traits.

Table S1. Summary of the number of SNPs, PIC, and gene diversity.

Table S2. The list of 286 cotton accessions used in the present study, including the cotton-growing region and phylogenetic groups.

Table S3. Statistical analyses of the phenotype of the female parent, F_1 hybrid, and the GCA value of the female parent.

Table S4. List of accessions for which the GCA value is in the top 5% and bottom 5%.

Table S5. List of the 19 F_1 hybrids that showed positive SCA values (except for micronaire) for fiber yield traits, fiber quality traits, or all traits.

Table S6. The population genetic differentiation statistics (*Fst*) between different groups.

Table S7. Summary of significant SNPs and common SNPs associated with eight fiber yield and quality-related traits identified by the single-locus GWAS method.

Table S8. List of the significant SNPs associated with the eight agronomic traits detected for the female parents in four environments by the single-locus GWAS method.

Table S9. Information for the 233 QTLs detected for the female parent phenotype.

Table S10. List of the significant SNPs associated with the eight agronomic traits detected for the female parents in four environments by the multi-locus GWAS method.

Table S11. Summary of the significant SNPs associated with the GCA values of the eight fiber yield and quality-related traits identified by the single-locus GWAS method.

Table S12. List of the significant SNPs associated with the GCA values of the female parents identified by the single-locus GWAS method.

Table S13. List of the significant SNPs associated with the GCA values of the female parents identified by the multi-locus GWAS method.

Table S14. List of the significant SNPs associated with the SCA values of the F_1 hybrids identified by the single-locus GWAS method.

Table S15. List of the significant SNPs associated with the SCA values of the F_1 hybrids identified by the multi-locus GWAS method.

Table S16. Pleiotropic QTLs identified in the present study.

Table S17. Haplotype proportions in the 20 elite accessions.

REFERENCES

Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Bruce, A.B. (1910) The Mendelian theory of heredity and the augmentation of vigor. *Science*, **32**, 627–628.

Crow, J.F. (1948) Alternative hypotheses of hybrid vigor. *Genetics*, **33**, 477–487.

Danecek, P., Auton, A., Abecasis, G. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Du, X.M., Huang, G., He, S.P. *et al.* (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796–802.

East, E.M. (1936) Heterosis. *Genetics*, **21**, 375–397.

Endelman, J.B. and Jannink, J.L. (2012) Shrinkage estimation of the realized relationship matrix. *G3: Genes Genomics Genet.* **2**, 1405–1413.

Falush, D., Stephens, M. and Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**(4), 1567–1587.

Fang, L., Wang, Q., Hu, Y. *et al.* (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **49**, 1089–1098.

Gabriel, S.B., Schaffner, S.F., Nguyen, H. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.

Galanopoulou-Sendouca, S. and Roupakias, D. (1999) Performance of cotton F_1 hybrids and its relation to the mean yield of advanced bulk generations. *Eur. J. Agron.* **11**, 53–62.

Giraud, H., Bauland, C., Falque, M. *et al.* (2017) Reciprocal Genetics: Identifying QTL for general and specific combining abilities in hybrids between multiparental populations from two maize (*Zea mays* L.) heterotic groups. *Genetics*, **207**, 1167–1180.

Guo, X., Guo, Y., Ma, J., Wang, F., Sun, M., Gui, L., Zhou, J., Song, X., Sun, X. and Zhang, T. (2013) Mapping heterotic loci for yield and agronomic traits using chromosome segment introgression lines in cotton. *J. Int. Plant Biol.* **55**, 759–774.

Hallauer, A.R., Marcelo, J.C. and Miranda Filho, J.B. (1981) *Quantitative genetics in maize breeding*. Ames, IA: Iowa State University Press.

Huang, C., Nie, X.H., Shen, C., You, C.Y., Li, W., Zhao, W.X., Zhang, X.L. and Lin, Z.X. (2017) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. *Plant Biotechnol. J.* **15**, 1374–1386.

Huang, X., Wei, X., Sang, T. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967.

Huang, X., Zhao, Y., Wei, X. *et al.* (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39.

Islam, M.S., Thyssen, G.N., Jenkins, J.N., Zeng, L.H., Delhom, C.D., McCarty, J.C., Deng, D.D., Hinchliffe, D.J., Jones, D.C. and Fang, D.D. (2016) A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genom.*, **17**, 903.

Jinks, J.L. and Jones, R.M. (1958) Estimation of the components of heterosis. *Genetics*, **43**, 223–234.

Jones, D.F. (1917) Dominance of linked factors as a means of accounting for heterosis. *Genetics*, **2**, 466–479.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C. and Eskin, E. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354.

Kearsey, M.J. and Pooni, H.S. (1996) *The genetical analysis of quantitative traits*. London, UK: Chapman and Hall.

Kliebenstein, D.J., Lambrix, V.M., Reichelt, M., Gershenzon, J. and Mitchell-Olds, T. (2001) Gene duplication in the diversification of secondary metabolism: tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in Arabidopsis. *Plant Cell*, **13**, 681–693.

Kump, K.L., Bradbury, P.J., Wisser, R.J. *et al.* (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* **43**, 163–168.

Lacape, J.M., Llewellyn, D., Jacobs, J. *et al.* (2010) Meta-analysis of cotton fiber quality QTLs across diverse environments in a *Gossypium hirsutum* × *G. barbadense* RIL population. *BMC Plant Biol.* **10**, 132.

Larièpe, A., Moreau, L., Laborde, J. *et al.* (2017) General and specific combining abilities in a maize (*Zea mays* L.) test-cross hybrid panel: relative importance of population structure and genetic divergence between parents. *Theor. Appl. Genet.* **130**, 403–417.

Li, C.Q., Xu, X.J., Dong, N., Ai, N.J. and Wang, Q.L. (2016) Association mapping identifies markers related to major early-maturating traits in upland cotton (*Gossypium hirsutum* L.). *Plant Breed.* **135**, 483–491.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Liang, Q., Shang, L., Wang, Y. and Hua, J. (2015) Partial dominance, overdominance and epistasis as the genetic basis of heterosis in upland cotton (*Gossypium hirsutum* L.). *PLoS One*, **10**, e0143548.

Liu, R., Wang, B., Guo, W., Qin, Y., Wang, L., Zhang, Y. and Zhang, T. (2012) Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. *Mol. Breed.* **29**, 297–311.

Ma, Z., He, S., Wang, X. *et al.* (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **50**, 803–813.

McKenna, A., Hanna, M., Banks, E. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

Meijón, M., Satbhai, S.B., Tsuchimatsu, T. and Busch, W. (2013) Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. *Nat. Genet.* **46**, 77.

Meredith, W.R. and Bridge, R.R. (1972) Heterosis and gene action in cotton, *Gossypium hirsutum* L.1. *Crop Sci.* **12**, 304–310.

Mo, H.D. (1982) The analysis of combining ability in p×q mating pattern. *Journal of Yangzhou University* **3**(3), 8–14.

Nakatsubo, T., Mizutani, M., Suzuki, S., Hattori, T. and Umezawa, T. (2008) Characterization of Arabidopsis thaliana pinoresinol reductase, a new type of enzyme involved in lignan biosynthesis. *J. Biol. Chem.* **283**, 15550–15557.

Paterson, A.H., Brubaker, C.L. and Wendel, J.F. (1993) A rapid method for extraction of cotton (*Gossypium spp.*) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* **11**, 122–127.

Powers, L. (1944) An expansion of Jones's theory for the explanation of heterosis. *Am. Nat.* **78**, 275–280.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., and Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**(8), 904–909.

Qin, Y., Liu, R., Mei, H., Zhang, T. and Guo, W. (2009) QTL mapping for yield traits in Upland cotton (*Gossypium hirsutum* L.). *Acta Agron. Sin.* **35**, 1812–1821.

Reif, J.C., Gumpert, F.M., Fischer, S. and Melchinger, A.E. (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. *Genetics*, **176**, 1931–1934.

Ren, W.L., Wen, Y.J., Dunwell, J.M. and Zhang, Y.M. (2018) pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity*, **120**, 208–218.

Richey, F.D. (1942) Mock-dominance and hybrid vigor. *Science*, **96**, 280–281.

Shan, C.M., Shangguan, X.X., Zhao, B. *et al.* (2014) Control of cotton fibre elongation by a homeodomain transcription factor *GhHOX3*. *Nat. Commun.* **5**, 5519.

Shang, L., Liang, Q., Wang, Y., Zhao, Y., Wang, K. and Hua, J. (2016a) Epistasis together with partial dominance, over-dominance and QTL by environment interactions contribute to yield heterosis in upland cotton. *Theor. Appl. Genet.* **129**, 1429–1446.

Shang, L., Ma, L., Wang, Y. *et al.* (2016b) Main effect QTL with dominance determines heterosis for dynamic plant height in upland cotton. *G3: Genes Genomics Genet.* **6**, 3373–3379.

Shang, L., Wang, Y., Cai, S., Ma, L., Liu, F., Chen, Z., Su, Y., Wang, K. and Hua, J. (2016c) Genetic analysis of Upland cotton dynamic heterosis for boll number per plant at multiple developmental stages. *Sci. Rep.* **6**, 35515.

Shang, L., Wang, Y., Cai, S., Wang, X., Li, Y., Abduweli, A. and Hua, J. (2015) Partial dominance, overdominance, epistasis and QTL by environment interactions contribute to heterosis in two upland cotton hybrids. *G3: Genes Genomics Genet.* **6**, 499–507.

Shen, C., Jin, X., Zhu, D. and Lin, Z.X. (2017) Uncovering SNP and indel variations of tetraploid cottons by SLAF-seq. *BMC Genom.* **18**, 247.

Shull, G.H. (1908) The composition of a field of maize. *J. Hered.* **4**, 296–301.

Sprague, G.F. and Tatum, L.A. (1942) General *vs.* specific combining ability in single crosses of corn. *Agron. J.* **34**, 923–932.

Su, J.J., Fan, S.L., Li, L.B. *et al.* (2016a) Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Front. Plant Sci.* **7**, 1576.

Su, J.J., Pang, C.Y., Wei, H.L. *et al.* (2016b) Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genom.*, **17**, 687.

Sun, Z., Wang, X., Liu, Z. *et al.* (2017) Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L. *Plant Biotechnol. J.* **15**, 982–996.

Tamba, C.L., Ni, Y.L. and Zhang, Y.M. (2017) Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* **13**, e1005357.

Tamba, C.L. and Zhang, Y.-M. (2018) A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*, 341784.

VanRaden, P.M. (2008) Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423.

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y. and Zheng, Y. (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant Sci.* **196**, 125–131.

Wang, M.J., Tu, L.L., Lin, M. *et al.* (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587.

Wang, S.B., Feng, J.Y., Ren, W.L., Huang, B., Zhou, L., Wen, Y.J., Zhang, J., Dunwell, J.M., Xu, S. and Zhang, Y.M. (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6**, 19444.

Wang, X., Yan, Y., Li, Y., Chu, X., Wu, C. and Guo, X. (2014) *GhWRKY40*, a multiple stress-responsive cotton WRKY gene, plays an important role in the wounding response and enhances susceptibility to *Ralstonia solanacearum* infection in transgenic *Nicotiana benthamiana*. *PLoS One*, **9**, e93577.

Wang, X., Yu, Y., Sang, J., Wu, Q., Zhang, X. and Lin, Z. (2013) Intraspecific linkage map construction and QTL mapping of yield and fiber quality of *Gossypium barbadense*. *Aust. J. Crop Sci.* **7**, 1252–1261.

Wen, J., Zhao, X.W., Wu, G.R. *et al.* (2015) Genetic dissection of heterosis using epistatic association mapping in a partial NCII mating design. *Sci. Rep.* **5**, 18376.

Wen, Y.J., Zhang, H., Ni, Y.L., Huang, B., Zhang, J., Feng, J.Y., Wang, S.B., Dunwell, J.M., Zhang, Y.M. and Wu, R. (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinformatics*, **19**, 700–712.

Werner, C.R., Qian, L., Voss-Fels, K.P., Abbadi, A., Leckband, G., Frisch, M. and Snowdon, R.J. (2018) Genome-wide regression models considering general and specific combining ability predict hybrid performance in oilseed rape with similar accuracy regardless of trait architecture. *Theor. Appl. Genet.* **131**, 299–317.

Wu, Y.T., Yin, J.M., Guo, W.Z., Zhu, X.F. and Zhang, T.Z. (2004) Heterosis performance of yield and fibre quality in F_1 and F_2 hybrids in upland cotton. *Plant Breed.* **3**, 285–289.

Yang, Z., Ge, X., Yang, Z. *et al.* (2019) Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat. Commun.* **10**, 2989.

Yang, Z., Li, Z. and Bickel, D.R. (2013) Empirical Bayes estimation of posterior probabilities of enrichment: a comparative study of five estimators of the local false discovery rate. *BMC Bioinformatics*, **14**, 87.

Zhang, J., Feng, J.Y., Ni, Y.L., Wen, Y.J., Niu, Y., Tamba, C.L., Yue, C., Song, Q. and Zhang, Y.M. (2017) pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity*, **118**, 517–524.

Zhang, T.Z., Hu, Y., Jiang, W.K. *et al.* (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. *acc.* TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537.

Zhang, T.Z., Qian, N., Zhu, X.F., Chen, H., Wang, S., Mei, H.X. and Zhang, Y.M. (2013) Variations and transmission of QTL alleles for yield and fiber qualities in upland cotton cultivars developed in China. *PLoS One*, **8**, e57220.

Zhao, Q., Zeng, Y., Yin, Y. *et al.* (2015) Pinoresinol reductase 1 impacts lignin distribution during secondary cell wall biosynthesis in *Arabidopsis*. *Phytochemistry*, **112**, 170–178.

Zhao, X.W., Li, B., Zhang, K., Hu, K.N., Yi, B., Wen, J., Ma, C.Z., Shen, J.X., Fu, T.D. and Tu, J.X. (2016) Breeding signature of combining ability improvement revealed by a genomic variation map from recurrent selection population in *Brassica napus*. *Sci. Rep.* **6**, 29553.

Zhou, H., Xia, D., Zeng, J., Jiang, G.H. and He, Y.Q. (2017) Dissecting combining ability effect in a rice NCII-III population provides insights into heterosis in indica-japonica cross. *Rice*, **10**, 39.