Journal of Cotton Research

RESEARCH Open Access

Association mapping and domestication analysis to dissect genetic improvement process of upland cotton yield-related traits in China

GUO Chunping[1†], PAN Zhenyuan[1†], YOU Chunyuan[2], ZHOU Xiaofeng[3], HUANG Cong[4], SHEN Chao[4], ZHAO Ruihai[1], YANG Qingyong[1,5], ZHU Longfu[1,4], SHAHZAD Raheel[6], MENG Fande[7*], LIN Zhongxu[1,4*] and NIE Xinhui[1*]

Abstract

Background: Cotton fiber yield is a complex trait, which can be influenced by multiple agronomic traits. Unravelling the genetic basis of cotton fiber yield-related traits contributes to genetic improvement of cotton.

Results: In this study, 503 upland cotton varieties covering the four breeding stages (BS1–BS4, 1911–2011) in China were used for association mapping and domestication analysis. One hundred and forty SSR markers significantly associated with ten fiber yield-related traits were identified, among which, 29 markers showed an increasing trend contribution to cotton yield-related traits from BS1 to BS4, and 26 markers showed decreased trend effect. Four favorable alleles of 9 major loci ($R^2 \geq 3$) were strongly selected during the breeding stages, and the candidate genes of the four strongly selected alleles were predicated according to the gene function annotation and tissue expression data.

Conclusions: The study not only uncovers the genetic basis of 10 cotton yield-related traits but also provides genetic evidence for cotton improvement during the cotton breeding process in China.

Keywords: Upland cotton, Genome wide association study, Yield-related traits, Favorable alleles

Background

Cotton is one of the most important industrial crop in the word, which has been cultivated for over 7 000 years, providing the important raw material for the textile industry (Fang et al. 2017a; Maik et al. 2015). Among the four cultivated cotton species, *Gossypium hirsutum* (upland cotton) is the most widespread species due to the high adaptability and yield, which takes up approximately 95% of cotton production in the word (Chen et al. 2007). Hence, the genetic improvement of upland cotton to increase cotton fiber yield is one of the most important goals of cotton breeding.

Fiber yield is a complex trait, which can be influenced by multiple traits, including seed cotton weight (SCW), lint weight (LW), lint percentage (LP), effective boll number (EBN), plant height (PH), first fruit spur height (FFSH), fruit spur branch number (FSBN), first fruit branch position (FFBP), flowering period (FP), and whole growth period (WGP) (Li et al. 2018b; Sun et al. 2018). Unravelling the genetic basis of cotton fiber yield-related traits contributes to cotton fiber production increase. Based on linkage analysis, many QTLs for cotton yield traits have been identified (An et al. 2010; Deng et al. 2019; Gore et al. 2014; Liu et al. 2012). However, the

* Correspondence: dwskjj@163.com; linzhongxu@mail.hzau.edu.cn; xjnxh2004130@126.com
†Guo CP and Pan ZY contributed equally to this work.
[7]Agricultural Science Research Institute of the 5th Division of Xinjiang Production and Construction Corps, Shuanghe 833408, Xinjiang, China
[1]Key Laboratory of Oasis Ecology Agricultural of Xinjiang Production and Construction Corps, Agricultural College, Shihezi University, Shihezi 832003, Xinjiang, China
Full list of author information is available at the end of the article

construction of linkage population used for fine mapping, such as recombinant inbred lines (RIL), and near-isogenic line (NIL), is a time-consuming process (Ma et al. 2020; Li et al. 2018a; Zhang et al. 2020). In recent years, genome-wide association studies (GWAS), which is based on linkage disequilibrium (LD), has been widely used in plants to identify various traits related QTLs by using the natural populations (Huang et al. 2016; Mengistu et al. 2016; Yang et al. 2014). In cotton, GWAS has been conducted for the genetic dissection of yield-related traits. Two hundred and fifty-one significant loci were detected to be associated with lint yield yields by using 651 simple sequence repeats (SSRs) and 323 accessions of *Gossypium hirsutum* L. (Jia et al. 2014). Thirteen cotton yield-related QTLs and 44 cotton fiber quality related QTLs were detected, respectively, by using 198 SSR markers and 302 elite upland cotton accessions (Ademe et al. 2017). Forty-three marker loci were detected to be associated with cotton yield traits by using 201 pleomorphic markers and 403 accessions (Dong et al. 2018). The identification of cotton yield-related QTLs lays the foundation for gene cloning and marker assisted selective (MAS) breeding.

The natural population resources possess widely genetic variation, which are suitable to study species domestication, such as maize (Hufford et al. 2012; Xue et al. 2016), rice (Zheng et al. 2019), soybean (Zhang et al. 2019; Zhou et al. 2015), tomato (Soltis et al. 2019), watermelon (Zhao et al. 2019), and cotton (Du et al. 2018; Nie et al. 2020). During crop breeding process, the favorable alleles were constantly enriched, so the varieties collected from different breeding stages could help us uncover the artificially selected loci. In this study, 503 upland cotton varieties covering main breeding stages in China were used to construct a natural population. Ten cotton yield-related traits were investigated in multiple environments. An association analysis was performed based on best linear unbiased predictions (BLUP) data from multiple environments, and 176 polymorphic SSRs, after which, the identified loci were used for domestication analysis and favorable alleles selection. Three major loci for LP and one loci for FS and WGP were strongly selected during these breeding stages. This study aimed to explore the genetic architecture of 10 cotton yield-related traits, and to uncover the genetic improvement during the cotton breeding process in China.

Materials and methods

Materials

In this study, 503 upland cotton inbred cultivars cultivated in, or introduced to China were collected to construct the natural mapping population (Table S1), which represent extensive genetic variation resources related to fiber yield-related traits and covered four breeding stages in China (Huang 2007). The stages were divided according to the breeding objectives: breeding stage1 (BS1, 1911–1950), primary improvement stage; breeding stage2 (BS2, 1951–1980), high yield upland cotton varieties; breeding stage3 (BS3, 1981–1999), high yield, high quality and disease resistance; breeding stage4 (BS4, 2000–2011), hybrid and insect-resistant cotton (Table S2).

Field experiments and phenotype data collection

Ten cotton yield-related traits of the 503 cotton inbred cultivars were measured under multiple environments. Five phenotypic data, including SCW, LW, LP, EBN, and PH were collected from 4 locations in China (Shihezi (SHZ), Xinjiang; Kuerle (KEL), Xinjiang; Yuanyang (YY), Henan; Huanggang (HG), Hubei) in 2012 and 2013. The phenotypic data of FFSH, and FSBN were collected from these locations in China (SHZ, KEL and YY), too, in 2012 and 2013. The phenotypic data of FFBP, FP, and WGP were collected from 2 locations (SHZ and YY) in 2012 and 2013. The experiment followed a randomized complete block design with single row plot and two replications.

Phenotypic data analysis

The variance, correlation, and repeatability analysis of the phenotypic data in multiple environments were conducted by R programming language.

Best linear unbiased predictions (BLUP) were used to estimate phenotypic traits across multiple environments based on a linear model by R software (http://www.r-project.org).

The basic statistical analysis of phenotypic data, including minimum (Min), maximum (Max), mean, standard deviation (SD), and coefficient of variation (CV), were conducted by IBM SPSS Statistics ver. 21.

Association analysis

One hundred and seventy-nine polymorphic SSR markers covering the whole genome were taken from the previous study, where the linkage disequilibrium and population structure analysis were also done (Nie et al. 2016). TASSEL V2.1 software was used to detect the association relationship between BLUP phenotypic data and genotypic data in three models, including GLM (P + G) + Q, GLM (P + G) + PCA and MLM (G + P + Q + K). The *P* values of markers associated with QTLs were regulated by the method of multiple testing correction by controlling the false discovery rate (Benjamini and Hochberg 1995).

Effect and evolution analysis of the associated markers in four breeding stages, and the selection of favorable alleles for major loci

The effect of the associated markers were evaluated in the four breeding stages based on the ten traits. Nine

major loci ($R^2 \geq 3$) were used for accumulation of favorable alleles analysis. The effects of different alleles were evaluated by using the phenotypic data, after that, the favorable alleles were used for frequency analysis in the four breeding stages, as well as the favorable alleles carrier materials selection.

Candidate gene annotation and prediction

The four strongly selected alleles associated to cotton yield-related traits were mapped to the TM-1 genome (*Gossypium hirsutum*, Huazhong Agricultural University (HAU) assembly) (Wang et al. 2019). The candidate regions for marker–trait loci were set around the LD decay distance as 400 kb. The genes in the candidate regions were used for key candidate genes selection by using annotation.

Gene expression patterns

The tissue expression levels of the candidate genes were obtained from Huazhong Agricultural University (unpublished).

Results

Performances of cotton yield-related traits of the 503 upland cotton germplasm resources in multiple environments

BLUP was used to determine the phenotypic data of cotton yield-related traits in multiple environments (Table S3), and the BLUP data of ten yield-related traits was used for association analysis. The average phenotypic coefficient of variation (CV) for 10 yield traits ranged from 2.85% (WGP) to 12.31% (FFSH) (Table 1). The highest CV was observed in FFSH (12.09%), the lowest in WGP (2.28%). The highest heritability was in LP (0.93), and the lowest in FSBN (0.46), ranging from 0.75 to 0.87 in the other seven traits (Table 1).

The phenotypic trends of yield-related traits in different environments were shown in Fig. 1. Among them, LP (Fig. 1c) showed most stable in the 8 environments; SCW (Fig. 1a), LW (Fig. 1b), EBN (Fig. 1d), PH (Fig. 1e), FFSH (Fig. 1f), FSBN (Fig. 1g), FFBP (Fig. 1h), and WGP (Fig. 1j) were relatively stable in the same site in different years; SCW (Fig. 1a) and LW (Fig. 1b) in HG showed relatively lower than other sites, while EBN (Fig. 1d) and PH (Fig. 1e) showed relatively higher in HG; FFSH (Fig. 1f) in HN showed relatively lower than other sites, while FSBN (Fig. 1g) showed relatively higher in HN.

Table 1 Statistics of cotton yield-related traits in the 503 upland cotton germplasm resources

	Min	Max	Mean	SD	CV/%	H2
SCW/g	3.73	5.87	4.87	0.3	6.18	0.81
LW/g	1.19	2.4	1.88	0.19	9.89	0.87
LP/%	27.07	47.38	38.31	3.04	7.92	0.93
EBN	13.33	21.74	17.44	1.37	7.87	0.6
PH/cm	72.44	116.17	92.81	5.54	5.97	0.76
FFSH/cm	8.67	21.34	13.75	1.69	12.31	0.75
FSBN	8.24	10.23	9.31	0.27	2.9	0.46
FFBP	4.43	6.92	5.85	0.41	6.97	0.75
FP/d	56.26	68.17	62.68	2.56	4.09	0.84
WGB/d	109.09	126.28	117.99	3.36	2.85	0.79

The correlations between two environments were obtained for the ten cotton yield-related traits (Fig. 2). Among the 10 yield-related traits, the average correlations between environments were ranked as LP (0.62) > FP (0.57) > WGP (0.51) > FFBP (0.45) > LW (0.43) > FFSH (0.38) > SCW (0.32) > PH (0.31) > FSBN (0.16) > EBN (0.15). It was with far more interest to further analyze one trait between environments. For example, the correlations of FFSH ranged from 0.16 (FFSH_12KEL and FUHML_12YY) to 0.75 (FFSH_12SHZ and FUHML_13 SHZ). In addition, the average correlation for FFSH in the same site in different years was 0.57, while the average correlation for FFSH in the same year from different sites was 0.34 (Fig. 2).

Molecular genetic diversity and population structure

A total of 179 polymorphic markers obtained from the previous study (Nie et al. 2016) were used for genetic analysis, which contained 426 allele loci. The average genetic similarity coefficient variation among the 503 cultivars was 0.552 (Nie et al. 2016). The linkage disequilibrium (LD) of this population was analyzed using 179 SSR markers. Based on r^2 estimates, only 2.09% ($r^2 \geq 0.05$) and 1.30% ($r^2 \geq 0.1$) of the markers showed significant LD, which is suitable for association mapping (Nie et al. 2016).

Population structure was determined by three methods. e.g., PCA (Principal component analysis) plots, Nei's genetic distance, and STRUCTURE software, in the previous study, and the population were divided into 7 subgroups (Nie et al. 2016).

Association mapping of cotton yield-related traits

The association analysis was carried out based on the genotypic data, PCA matrix, genetic relationship matrix and BLUP data of ten cotton yield-related traits (Bradbury et al. 2007; Jakobsson and Rosenberg 2007). Through association analysis, the effects of GLM (P + G) + Q, GLM (P + G) + PCA and MLM (G + P + Q + K) models in association analysis were compared (Fig. S1). For LW and EBN, the above three models had similar effect in controlling population structure; For PH, FFBP, FP and WGP, GLM-Q and MLM-Q-K models were better than GLM-PCA. For SCW, FFSH and FSBN, GLM-PCA model was better than GLM-Q and MLM-Q-K.

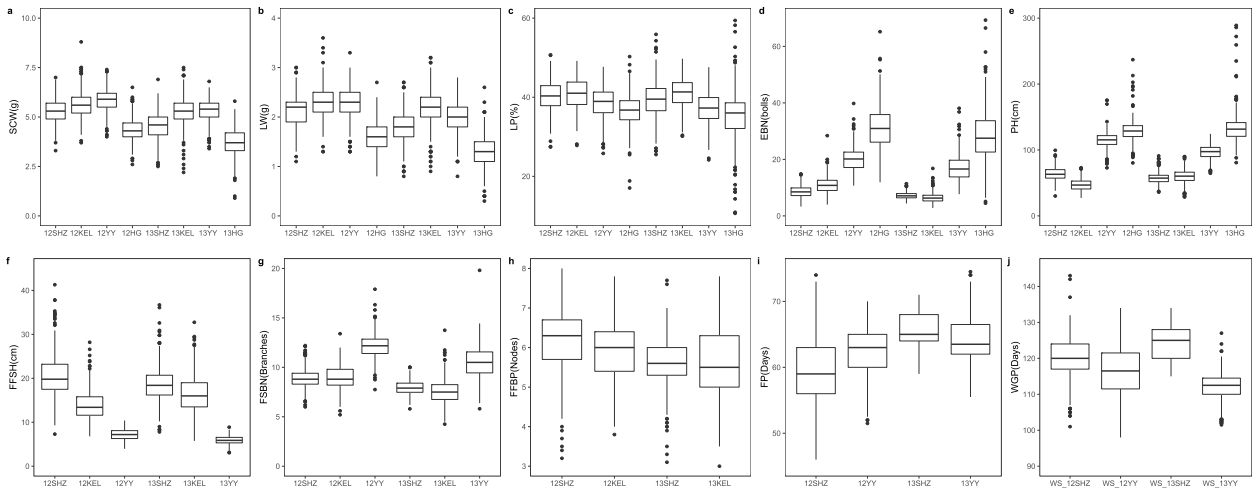

Fig. 1 The boxplots of the changing trends of ten yield-related traits in multiple environments. 12SHZ (Shihezi in 2012), 12KEL (Kuerle in 2012), 12YY (Yuanyang in 2012), 12HG (Huanggang in 2012), 13SHZ (Shihezi in 2013), 13KEL (Kuerle in 2013), 13YY (Yuanyang in 2013), 13HG (Huanggang in 2013)

For LP, GLM-Q model was better than GLM-PCA and MLM-Q-K model. According to the results, the effect of the MLM-Q-K model in the association analysis is more stable than other two models, and the effect of the 10 yield-related traits in controlling population structure is smaller. Therefore, the MLM-Q-K model was suitable for controlling the population structure of ten yield-related traits.

After filtering the minimum alleles (≤5%), 179 polymorphic SSR markers (426 alleles) were used for association analysis based on MLM-Q-K model. At the $P \leq 0.05$ and $P \leq 0.01$ levels, 140 (78.21%) and 45 (25.14%) markers were associated with yield-related traits, respectively (Table S4). An average of 5.4 associated markers were detected on each chromosome, ranging from 1 to 11, with the maximum of 11 markers on chromosomes D05 (Fig. 3). At $P \leq 0.05$ level, the phenotypic variation interpretation rates (PVE) of the associated markers ranged from 0.48% (NAU4884a and NAU3468a) to 3.89% (NAU3377db) with an average of 1.30%, while at $P \leq 0.01$ level, the PVE ranged from 1.03% (MON-CGR5866a) to 3.89% (NAU3377db), with an average of 2.14% (Table S4). In addition, one marker was generally associated with multiple traits (Fig. 3). For example, NAU2095 (on Chromosome A01) was associated with FFSH, PH, FSBN, WGP, LP, and EBN; MON_

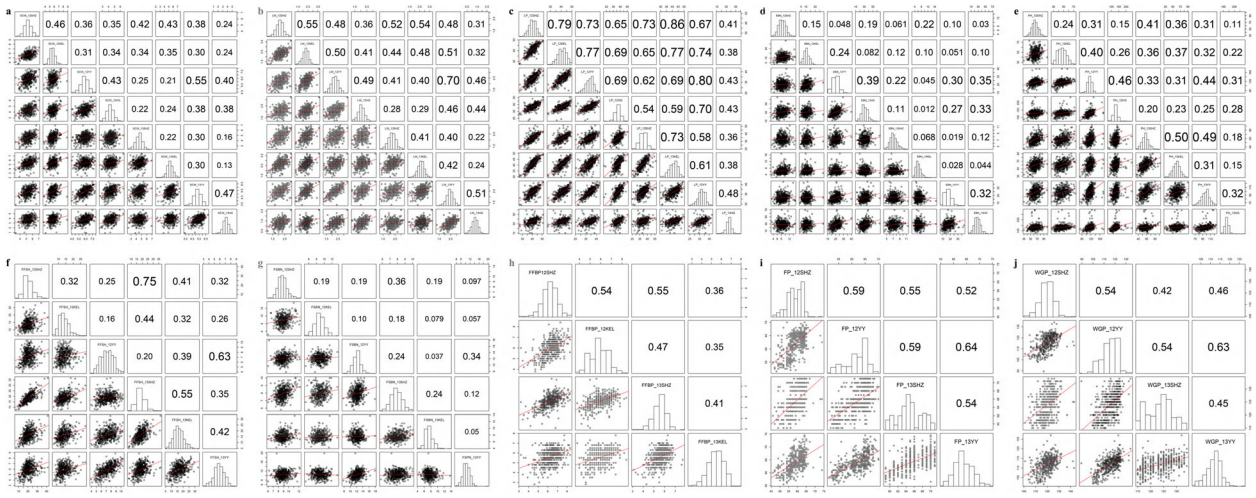

Fig. 2 The correlations of ten yield-related traits in multiple environments. **a** Correlations of SCW in eight environments. **b** Correlations of LW in eight environments. **c** Correlations of LP in eight environments. **d** Correlations of EBN in eight environments. **e** Correlations of PH in eight environments. **f** Correlations of FFSH in six environments. **g** Correlations of FSBN in six environments. **h** Correlations of FSBP in four environments. **i** Correlations of FP in four environments. **j** Correlations of WGP in four environments

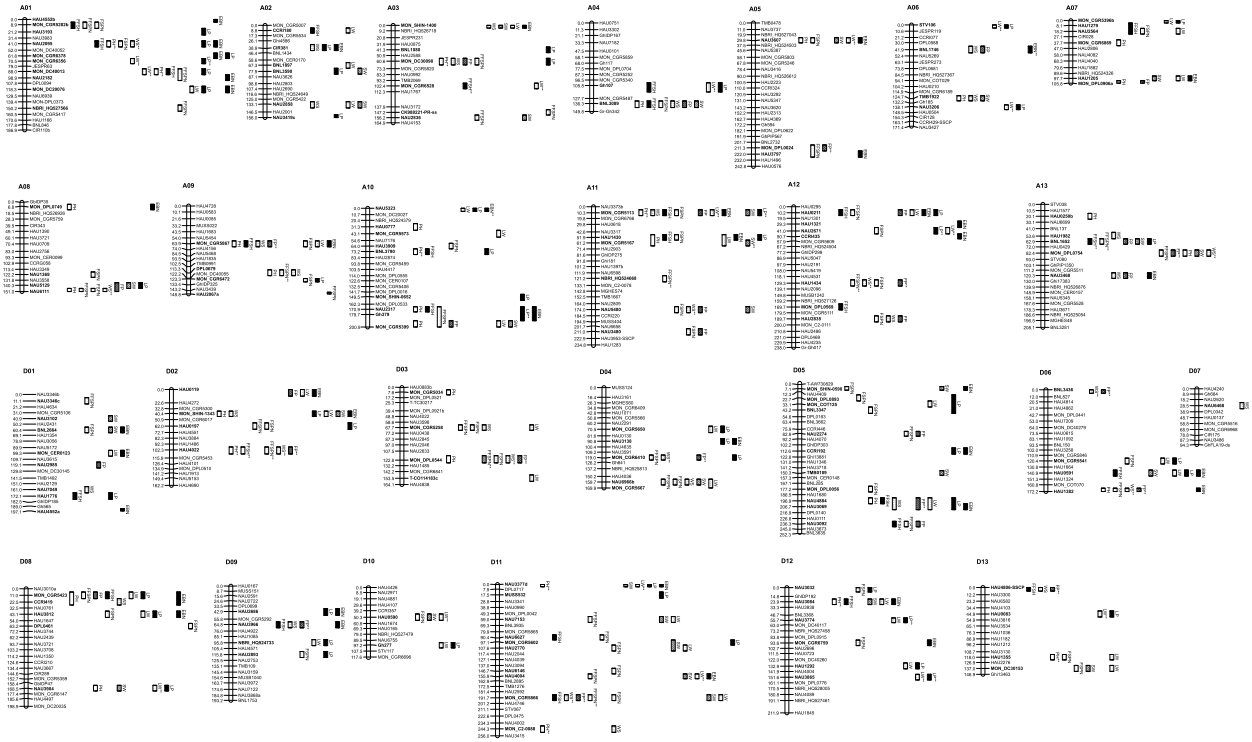

Fig. 3 The distribution of the SSR markers associated with the yield-related traits on 26 chromosomes of upland cotton

DC40013 (on Chromosome A01) was associated with LW, PH, FFSH, FFBP, and LP; NAU2564 (on Chromosome A07) was associated with FFSH, FSBN, and LW; MON_CGR5113 (on Chromosome A11) was associated with PH, WGP, FFSH, FSBN, FP, LW, EBN, SCW, and LP; HAU0211 (on Chromosome A12) was associated with FFSH, FFBP, FP, LW, EBN, and LP; HAU4022 (on Chromosome D02) was associated with PH, FFSH, FFBP, WGP, and FP at an extremely significant level ($P < 0.01$); HAU1355 (on Chromosome D13) was simultaneously associated with PH, FSBN, and LW; NAU3084 (on Chromosome D12) was associated with PH, FFSH, SCW, LW, and EBN (Fig. 3, Table S4).

LP was associated with most loci among the 10 yield-related traits, with a number of 84 ($P < 0.05$) and 11 ($P < 0.01$). PVE ranged from 0.66% (MON-CGR6410a and BNL1652a) to 3.55% (NAU2671a), with an average of 1.32% ($P < 0.05$) and 2.49 ($P < 0.01$) (Table S4).

FFSH was associated with least loci among the 10 yield-related traits, which ranged from 5 ($P < 0.01$) to 39 ($P < 0.05$). The PVE ranged from 0.65% (NAU4884b) to 3.50% (HAU4022b) (Table S4).

The number of loci associated with LW was detected in the range of 20 ($P < 0.01$) to 81 ($P < 0.05$), and the PVE ranged from 0.48% (NAU4884a) to 2.54% (MON-CGR5113c), with an average of 1.12% (Table S4).

There were 50 ($P < 0.05$), and 21 ($P < 0.01$) loci associated with FP. The PVE ranged from 1.01% (NAU4884b) to 3.61% (HAU1434b), with an average of 1.83% (Table S4).

There were 48 ($P < 0.05$) and 1 ($P < 0.01$) loci associated with SCW. The PVE ranged from 0.73% (NAU3904a) to 2.89% (MON-CGR5167b) (Table S4).

There were 55 ($P < 0.05$) and 2 ($P < 0.01$) loci associated with EBN. The PVE ranged from 0.6% (MON-SHIN-1343a) to 2.71% (NAU5323a) (Table S4).

There were 49 ($P < 0.05$) and 13 ($P < 0.01$) loci associated with PH. The PVE ranged from 0.67% (NAU3607a) to 3.89% (NAU3377db) (Table S4).

There were 56 (only at $P < 0.05$) loci associated with FSBN. The PVE ranged from 0.67% (HAU1430b) to 2.49% (MON-SHIN-0598a) (Table S4).

There were 59 ($P < 0.05$) and 7 ($P < 0.01$) loci associated with FFBP. The PVE ranged from 0.58% (MON-CGR5423d) to 2.29% (DPL0679a) (Table S4).

There were 46 ($P < 0.05$) and 14 ($P < 0.01$) loci associated with WGP. The PVE ranged from 0.48% (NAU3468a) to 3.39% (MON_DPL0754a) (Table S4).

Effect value analysis of associated markers in four breeding stages

According to the history of cotton breeding stages in China, the 503 upland cotton germplasm resources were divided into 4 groups (Table S2), from breeding stage1 (BS1) to breeding stage 4 (BS4), which represented abroad variation in each experiment site. The ten cotton

yield-related traits were compared among the four breeding stages. According to the results, the BS4 almost represented the highest level of 9 traits, except WGP, which showed a decreased trend during the breeding process; LW, LP, PH, and FFSH showed an increasing trend from BS1 to BS4, while the others showed little changes or random fluctuations during the four breeding stages (Fig. 4).

To further uncover the effect of the associated markers on the ten cotton yield-related traits, the phenotypic effect value of each locus in each breeding stage was assessed (Fig. 5). For 48 associated markers of SCW, the average phenotypic effect value of the allele loci in each stage (BS1-BS4) was − 0.11, − 0.003, 0.01, − 0.003, respectively (Table S5). The phenotypic effect range of each locus in each stage was − 4.96 to 0.31, − 0.44 to 0.38, − 0.37 to 0.24, and − 0.21 to 0.27, respectively (Table S5). As shown in Fig. 5a, the phenotypic effect showed an upward trend during the four breeding stages in the marker loci of HAU0590a (− 0.09 to 0.02), NAU6966ba (− 0.03 to 0.04), and NAU3607a (− 4.96 to 0.27), while NAU2858a (− 0.08 to 0.05) showed a downward trend (Fig. 5a, Table S5).

For the 81 associated markers of LW, the average phenotypic effect value in each stage (BS1-BS4) was − 0.10, − 0.01, − 0.02, − 0.01, respectively. The phenotypic effect range of each locus in each stage was − 1.76 to 0.26, − 0.51 to 0.12, − 1.89 to 0.10, and − 0.32 to 0.14, respectively (Table S5). As shown in Fig. 5b, the phenotypic effects showed an upward trend in the following loci: NBRI_HQ524733d (− 0.13 to 0.01), NAU3607a (− 1.76 to 0.14), NBRI_HQ524733c (− 0.14 to 0.01), MON_CGR5399c (0.01 to 0.07) and NAU3084c (− 0.22 to-0.01). HAU2770a (0.01 to 0.06), NBRI_HQ524733a (− 0.001 to 0.04) and NAU3774a (0.003 to 0.08) showed a downward trend (Fig. 5b, Table S5).

For the 84 alleles of LP, the average phenotypic effect value in each stage was − 0.6854, − 0.23, − 0.04, − 0.04, respectively. The phenotypic effect range of each locus in each stage was − 1.57 to 3.00, − 8.80 to 2.93, − 3.85 to 4.94, and − 2.70 to 4.35, respectively (Table S5). As shown in Fig. 5c, the phenotypic effects of the four breeding stages showed an upward trend in the following loci: HAU0197b (− 1.58 to 0.28), HAU0083c (− 5.15 to − 0.05), HAU3812a (− 0.40 to 0.38), NAU2095a (− 2.01 to 0.44), and MON_CGR6356b (− 3.04 to − 1.03). MON_CGR6356a (0.0003 to 0.44), MON_CGR6472a (0.025 to 1.45), NAU3774a (0.04 to 2.29), and HAU0197a (− 0.001 to 0.61) showed a decreasing trend (Fig. 5c, Table S5).

For the 55 associated markers of EBN, the average phenotypic effect value in each stage was − 0.28, 0.05, − 0.02 and − 0.02, with the variation range of − 2.99 to 1.74, − 1.37 to 2.29, − 1.12 to 1.39 and − 2.33 to 1.54, respectively (Table S5). As shown in Fig. 5d, the phenotypic effects showed an upward trend in MON_CGR5867a (− 0.91 to 0.34), MON_CGR6378c (− 2.99 to 0.81), HAU4552a (− 0.31 to 0.09), HAU0119c (− 1.19 to 0.45), and HAU1382a (− 0.92 to − 0.08). CCRI192a (− 1.06 to 0.17) and HAU3966b (− 0.22 to − 0.03) showed a decreasing trend (Fig. 5d, Table S5).

Among the associated markers of PH, the phenotypic effect of DPL0475a showed an upward trend, while MON_CGR5167b and BNL3790c showed a downward trend (Fig. 5e). Among the associated markers of FFSH, the phenotypic effect of HAU0211c showed an upward trend, while MON_CGR5423e showed a downward trend (Fig. 5f). Among the associated markers of FSBN, the phenotypic effect of BNL3347a and MON_CGR5867c showed an upward trend, while HAU2835b, HAU1355a, MON_CGR5866b, HAU0197a, HAU1434b and DPL0461c showed downward trend (Fig. 5g).

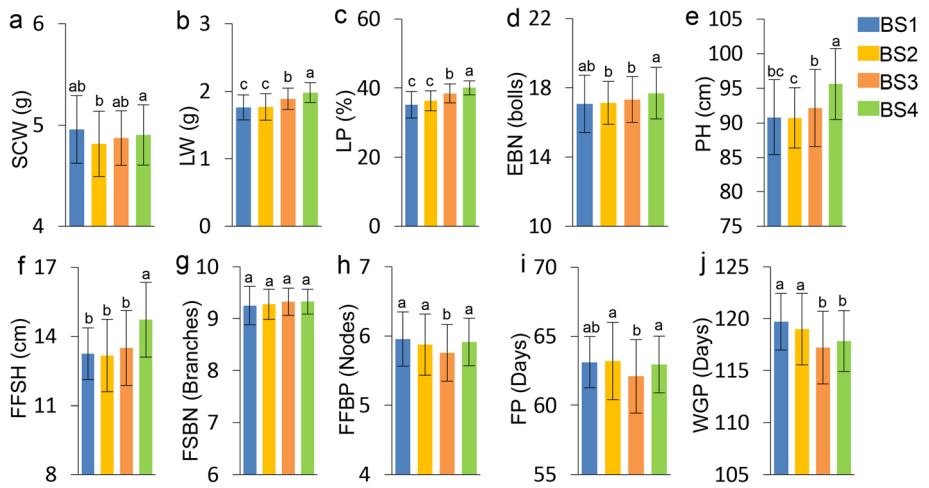

Fig. 4 The statistic analysis and comparison of the ten cotton yield-related traits in the four breeding stages

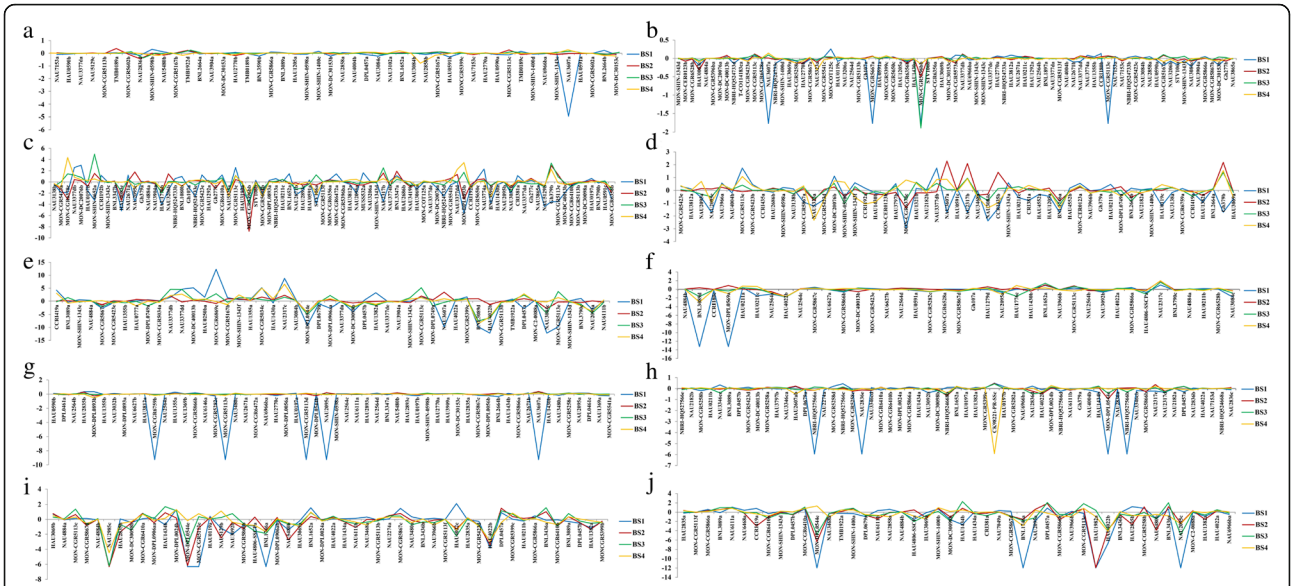

Fig. 5 The change trends of phenotypic effect of marker loci associated with ten yield-related traits in the four breeding stages. **a** SCW (seed cotton weight), **b** LW (lint weight), **c** LP (lint percentage), **d** EBN (effective boll number), **e** PH (plant height), **f** FFSH (first fruit spur height), **g** FSBN (fruit spur branch number), **h** FFBP (first fruit branch position), **i** FP (flowering period), **j** WGP (whole growth period). The four different colored lines indicate four different breeding stages

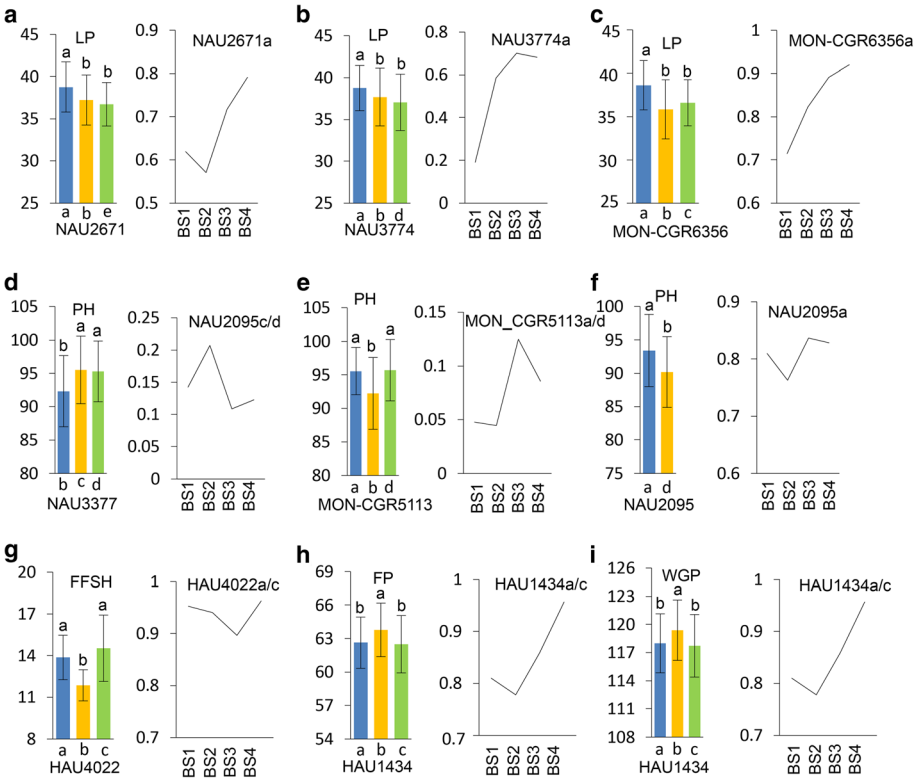

Fig. 6 The effects of different alleles of the major loci ($R^2 \geq 3.0$) and the favorable allele frequency trends in the four breeding stages. The phenotypic statistics of different alleles of NAU2671, NAU3774, MON-CGR6356, NAU3377, MON-CGR5113, NAU2095, HAU4022, and HAU1434 locus was shown in a-i (left chart). And the favorable allele frequency trends in the four breeding stages are shown in a-i (right chart)

Among the associated markers of FFBP, the phenotypic effect of MON_DPL0544e showed an upward trend, while MON_CGR5866a and CK98221_PR_SSc showed a downward trend (Fig. 5h). Among the associated markers of FP, the phenotypic effect of MON_DPL0544e, MON_CGR5113e, HAU4022b, MON_CGR5399c, and DPL0475b increased from BS1 to BS4, while NAU3468a and NAU3966b decreased (Fig. 5i). Among the associated markers of WGP, the phenotypic effect of MON_DPL0544e increased from BS1 to BS4, while phenotypic effect of MON_CGR5866a, NAU2858a and NAU3966b decreased.

Accumulation of favorable alleles for major loci in four cotton breeding stages and typical carrier materials

Nine major loci ($R^2 \geq 3$) were used for accumulation of favorable alleles analysis, including three loci for LP (NAU2671, NAU3774, and MON-CGR6356), three loci for PH (NAU3377, MON-CGR5113, and NAU2095), one locus for FFSH (HAU4022), one locus for FP (HAU1434), and one locus for WGP (HAU1434). NAU2671 has three alleles, i.e., NAU2671a, NAU2671b, and NAU2671e, among which, NAU2671a was the favorable allele (Fig. 6a), showing the highest FP. The frequency of NAU2671a showed an increasing trend during the four breeding stages (0.62 to 0.79) (Fig. 6a). NAU3774 has three alleles, too, i.e., NAU3774a, NAU3774b, and NAU3774d, among which, NAU3774a was the favorable allele. The frequency of NAU3774a showed an increasing trend during the four breeding stages (0.19 to 0.70) (Fig. 6b). MON-CGR6356 also had three alleles, i.e., MON-CGR6356a, MON-CGR6356b, and MON-CGR6356c, among which, MON-CGR6356a was the favorable allele. The frequency of MON-CGR6356a showed an increasing trend during the four breeding stages (0.71 to 0.92) (Fig. 6c). NAU3377c/d are favorable alleles of PH, which could increase the plant height. The frequency of NAU3377c/d showed a certain fluctuation, and the frequencies of BS3 and BS4 were slightly decreased (Fig. 6d). MON-CGR5113a/d were favorable alleles of PH. The frequency of MON-CGR5113a/d showed an increasing trend from BS1 to BS3, but a little decrease in BS4 (Fig. 6e). NAU2095a is a favorable allele of PH, and the frequency of which showed an irregular fluctuation among the four breeding stages, while the frequencies of BS3 and BS4 were slightly increased. HAU4022a/c were favorable alleles of FFSH, the frequency of which was relatively stable among the four breeding stages. HAU1434a/c were favorable alleles of both FP and WGP, which showed an increasing trend from BS1 to BS4 (0.80 to 0.95) (Fig. 6i).

Additionally, the typical carrier materials possessing the favorable allele were shown in Table S6. For LP, the umber carrier materials with three favorable alleles (NAU2671a, NAU3774a, and MON-CGR6356a) was 136, such as ZY1, ZY27, ZY28, and ZY44. For PH, there were 19 materials carried three favorable alleles (NAU3377c/d, MON-CGR5113a/d, and NAU2095a), such as ZY13, ZY55 and ZY71. For FFSH, there were 473 materials carrying the favorable alleles HAU4022a/c. For FP and WGP, there were 429 materials carrying the favorable alleles HAU1434a/c. However, there were only three materials carrying 8 favorable alleles, including ZY92, ZY398, and ZY87 (Table S6). ZY92 was derived from the BS3, while ZY398 and ZY87 were derived from BS4.

Candidate gene annotation and prediction

Four artificially selected major loci during the breeding process, including three loci of LP (NAU2671, NAU3774, and MON-CGR6356), one locus of FP and WGP (HAU1434), were used for candidate gene annotation and prediction. Candidate genes were listed within the candidate regions of loci, which were set around the LD decay distance as 400 kb. There were 34, 42, 20 and 22 candidate genes for 4 loci (NAU2671, NAU3774, MON-CGR6356, and HAU1434), respectively (Table S7). One key candidate gene for each locus was predicted according to the functional annotation. For NAU2671 of LP, Ghir_A12G024420 encoded a homeobox protein BEL1, which was expressed in all tested tissues (Fig. 7a); in *Arabidopsis*, BEL1-LIKE HOMEODOMAIN6 was involved in secondary cell wall formation of interfascicular fiber (Liu et al. 2014). For NAU3774 of LP, Ghir_D12G026070 encoded a Cytochrome B5 protein, which was highly expressed in the developing fibers (Fig. 7b); in *Arabidopsis*, Cytochrome b5 worked as an obligate electron shuttle for syringyl lignin biosynthesis, which also played an important role in the cotton fiber development (Gou et al. 2019). For MON-CGR6356 of LP, Ghir_A01G016840 encoded auxin-induced 5NG4 protein, which was expressed specially in the developing fibers (Fig. 7c); auxin-induced protein 5NG4 was associated with culm cellulose content in bread wheat (Kaur et al. 2017). For HAU1434 of FP and WGP, Ghir_A12G013050 encoded a SBP domain protein, which was highly expressed specially in the floral organ (Fig. 7d); SBP domain proteins were widely reported to be involved in flowering initiation and flower development (Cardon et al. 1997; Hou et al. 2017; Yamasaki et al. 2006).

Discussion

Extensive genetic variation and multiple environments contribute to QTL mapping

Genome-wide association study using natural population is a powerful strategy to effectively fine map QTL due to

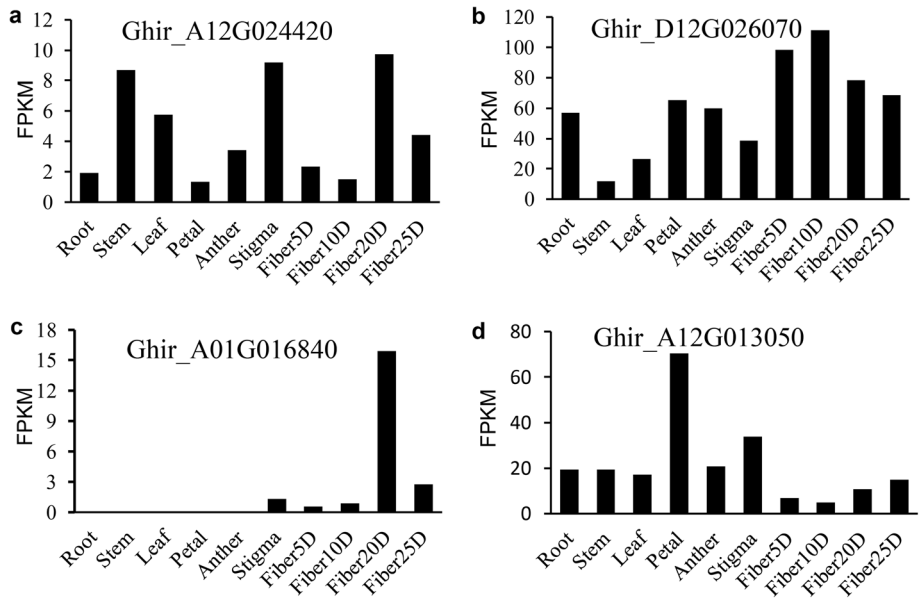

Fig. 7 Expression profiles of candidate genes in various tissues. **a**. Tissue expression of Ghir_A12G024420. **b** tissue expression of Ghir_D12G026070. **c** tissue expression of Ghir_A01G016840. **d** Tissue expression of Ghir_A12G013050

a great number of historical recombination events that lead to the rapid decay of linkage disequilibrium (LD) (Flint-Garcia et al. 2003; Li et al. 2013). The population size and diversity of germplasm resources could affect the LD and the resolution of QTL. In our study, 503 germplasms derived from the United States, the former Soviet Union, and China were divided into 7 subgroups by using the SSR markers, which represent extensive genetic variations and are suitable for GWAS. In addition, quantitative traits are generally susceptible to environments, so the QTLs stable in multiple environments were more reliable, which can be used for gene cloning and marker assistance selection (MAS) (Raihan et al. 2016; Wang et al. 2015). In this study, the data of 5 traits were collected from 8 environments (4 typical cotton growing areas in China in 2 years), and BLUP phenotypic data were used to provide comprehensive multi-environmental accurate phenotypic values for association analysis. According to the results, the extensive genetic variation and multiple environments of the study make the QTLs more reliable.

In order to verify the reliability of the association mapping results, the associated markers were compared with the previous studies. As a result, some of the markers were consistent with previous studies. For example, an LP- and SCW-related loci, BNL3590 (Chr A02), was reported in two studies, where it was associated with SCW (An et al. 2010; Mei et al. 2013). MON_CGR5399 (Chr A10) was reported to be associated with LW in previous studies, while, in the study, it was associated with multiple traits, including LW, FP, PH, FFBP and SCW (Wang et al. 2015). NAU3377 (Chr D11) was reported to be associated with LP in previous studies, while in the study (Wang et al. 2007), NAU3377 (Chr D11) was not only associated with LP, but also with EBN, LW, PH, SCW. NAU3774 (Chr D12) was reported to be associated with LW, while in the study, NAU3774 was associated with both LW and LP (Wang et al. 2015). Additionally, there were more new loci identified in this study and could be used for further genetic analysis of cotton yield-related traits.

The artificial selection during the breeding process contributes to the increased cotton production

As a natural fiber resource, Gossypium was cultivated about 7 000 years ago (Fang et al. 2017a). During the process of cultivation, its fiber productivity has been increased duo to natural genome polyploidization and artificial selection (Jiang et al. 1998; Yuan et al. 2015). Artificial selection could influence the allele frequency through the selection of preferred traits in the traditional breeding process (Luikart et al. 2003). Mutation and recombination are two important factors in determining the efficiency of selection (Nachman and Payseur 2012; Noor and Bennett 2009). The comparison of the genome sequence of *Gossypium hirsutum* CRI-12 family (including CRI-12, its parental cultivars, and progeny cultivars) revealed that 1 029 haplotype blocks might be recombined under artificial selection (Lu et al. 2019). Strong artificial selection during domestication has resulted in reduced genetic diversity but stronger linkage disequilibrium and higher extents of selective sweeps (Ma et al.

2019). In our study, LW, LP, PH, and FFSH showed significant increasing trends during the cotton breeding process, while WGP showed a significant decreased trend, which was just consistent with the evolution of cotton varieties in Xinjiang, China, and resulted in higher yield but shorter growing period cotton varieties. The effect and evolution analysis of markers associated with yield traits in the four breeding stages have been assessed. Twenty-nine markers showed an increasing trend contribution to cotton yield-related traits from BS1 to BS4, which could be caused by increasing favorable allele frequency during artificial selection breeding process. However, 26 associated markers showed decreased trend effect from BS1 to BS4, which showed that there was a great potential to gain cotton yield by increasing the favorable alleles of those loci. Additionally, the favorable allele frequency of major loci were detected, and three favorable alleles (NAU2671a, NAU3774a, and MON-CGR6356a) associated with LP were strongly selected during the cotton breeding process, which contributed to the increasing LP of cotton; one favorable allele associated with WGP was strongly selected, which contributed the decreasing WGP of cotton. According to the results, cotton breeding process is a genetic improvement to increase frequency and number of favorable alleles. However, there were still some favorable alleles that were not affected or negative selected during the evolution, which could be as a potential genetic resource for cotton genetic improvement in the future.

The loci identified lay the foundation for gene cloning and molecular breeding

With the developing of genome sequencing technology, the reference genome sequences of *Gossypium hirsutum* have been well developed (Wang et al. 2019), which provide lots of useful information for candidate gene identification from the results of GWAS (Huang et al. 2018; Nie et al. 2020). According to the reference genome, the SSR markers significantly associated with cotton yield-related traits could be anchored on the physical maps, which provided candidate regions based on LD decay distance (Huang et al. 2018). The annotated genes in these regions could be used for further verification by functional annotation or using reverse genetics methods (Fang et al. 2017b; Li et al. 2017; Xu et al. 2020). In our study, at the level $P \leq 0.01$, 45 markers were identified to be associated with cotton yield-related traits, among which, 8 major loci were used for favorable allele identification. Additionally, the materials which carried the favorable alleles were identified, which could be used as donors to improve the cotton yield, such as ZY92, ZY398, and ZY87, which carried 8 favorable alleles. Four strongly selected alleles were used for candidate predication according to gene functional annotation and tissue expression. As a result, three candidate genes were identified for LP, which were associated with fiber development according to the function of homologous in *Arabidopsis* and wheat (Gou et al. 2019; Kaur et al. 2017; Liu et al. 2014). One candidate gene was identified for FP and WGP, which was specifically expressed in flower organs, and the homologous in other plants were involved in flowering initiation and flower development (Cardon et al. 1997; Yamasaki et al. 2006).

Conclusion

In this study, 140 markers significantly associated with ten fiber yield-related traits were identified by using 503 upland cotton varieties covering the four breeding stages in China, among which, 29 loci showed an increasing trend contribution to cotton yield-related traits from BS1 to BS4, and 26 loci showed a decreased trend effect. Four favorable alleles of 9 major loci ($R^2 \geq 3$) were strongly selected during the breeding stages, and the candidate genes of the four strongly selected alleles were predicated according to the gene function annotation and tissue expression data. The study not only uncovers the genetic basis of 10 cotton yield-related traits but also provides genetic evidence for cotton improvement during the cotton breeding process in China.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-021-00087-3.

Additional file 1.
Additional file 2.
Additional file 3.
Additional file 4.
Additional file 5.
Additional file 6.
Additional file 7.
Additional file 8.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable comments and helpful suggestions which help to improve the manuscript.

Authors' contributions

Nie XH, Lin ZX and Pan ZY designed the experiments. Meng FD provided the cotton germplasm resources. Guo CP, Pan ZY, Nie XH, You CY, Huang C, Shen C, Zhou XF and Zhao RH performed the experiments. Nie XH, Guo CP and Pan ZY wrote the main manuscript text and prepared all Figures. Guo CP, Pan ZY, Huang C, Shen C and Yang QY performed data analysis. Nie XH, Pan ZY, Lin ZX, Zhu LF, and Shahzad R revised and polished the manuscript. All authors contributed in the interpretation of results and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (31760402), Young and Middle-aged Science and Technology Leading Talents of Xinjiang Production and Construction Corps (2019CB027).

Availability of data and materials
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All Authors have provided ethical approval and consent to participate as well as consent for publication.

Competing interests
The authors have declared that no competing interests exist.

Author details
[1]Key Laboratory of Oasis Ecology Agricultural of Xinjiang Production and Construction Corps, Agricultural College, Shihezi University, Shihezi 832003, Xinjiang, China. [2]Cotton Research Institute, Shihezi Academy of Agriculture Science, Shihezi 832011, Xinjiang, China. [3]Cotton Institute, Xinjiang Academy of Agriculture and Reclamation Science, Shihezi 832000, Xinjiang, China. [4]National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences & Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China. [5]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, Hubei, China. [6]Department of Biotechnology, Faculty of Science and Technology, Universitas Muhammadiyah Bandung, Bandung, West Java 40614, Indonesia. [7]Agricultural Science Research Institute of the 5th Division of Xinjiang Production and Construction Corps, Shuanghe 833408, Xinjiang, China.

Received: 31 January 2021 Accepted: 22 March 2021
Published online: 04 May 2021

References
Ademe MS, He SP, Pan ZE, et al. Association mapping analysis of fiber yield and quality traits in upland cotton (*Gossypium hirsutum* L.). Mol Gen Genomics. 2017;292(6):1267–80. https://doi.org/10.1007/s00438-017-1346-9.

An CF, Jenkins JN, Wu JX, et al. Use of fiber and fuzz mutants to detect QTL for yield components, seed, and fiber traits of upland cotton. Euphytica. 2010; 172:21–34. https://doi.org/10.1007/s10681-009-0009-2.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc: Series B (Methodological). 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bradbury PJ, Zhang Z, Kroon DE, et al. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308.

Cardon GH, Höhmann S, Nettesheim K, et al. Functional analysis of the *Arabidopsis thaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. Plant J. 1997;12(2):367–77. https://doi.org/10.1046/j.1365-313x.1997.12020367.x.

Chen ZJ, Scheffler BE, Dennis E, et al. Toward sequencing cotton (*Gossypium*) genomes. Plant Physiol. 2007;145(4):1303–10. https://doi.org/10.1104/pp.107.107672.

Deng XY, Gong JW, Liu AY, et al. QTL mapping for fiber quality and yield-related traits across multiple generations in segregating population of CCRI 70. J Cotton Res. 2019;2(1):13. https://doi.org/10.1186/s42397-019-0029-y.

Dong CG, Wang J, Chen QJ, et al. Detection of favorable alleles for yield and yield components by association mapping in upland cotton. Genes Genomics. 2018;40(7):725–34. https://doi.org/10.1007/s13258-018-0678-0.

Du XM, Huang G, He SP, et al. Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. Nat Genet. 2018;50(6):796–802. https://doi.org/10.1038/s41588-018-0116-x.

Fang L, Gong H, Hu Y, et al. Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. Genome Biol. 2017a;18(1): 33. https://doi.org/10.1186/s13059-017-1167-5.

Fang L, Wang Q, Hu Y, et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nat Genet. 2017b;49(7):1089–98. https://doi.org/10.1038/ng.3887.

Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003;54(1):357–74. https://doi.org/10.1146/annurev.arplant.54.031902.134907.

Gore MA, Fang DD, Poland J, et al. Linkage map construction and quantitative trait locus analysis of agronomic and fiber quality traits in cotton. Plant Genome. 2014;7(1):1–10. https://doi.org/10.3835/plantgenome2013.07.0023.

Gou MY, Yang XM, Zhao YJ, et al. Cytochrome b5 is an obligate electron shuttle protein for syringyl lignin biosynthesis in Arabidopsis. Plant Cell. 2019;31(6): 1344–66. https://doi.org/10.1105/tpc.18.00778.

Hou H, Yan X, Sha T, et al. The SBP-box gene *VpSBP11* from Chinese wild *vitis* is involved in floral transition and affects leaf development. Int J Mol Sci. 2017; 18(7):1493. https://doi.org/10.3390/ijms18071493.

Huang C, Shen C, Wen TW, et al. SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population. Mol Gen Genomics. 2018;293(4):793–805. https://doi.org/10.1007/s00438-018-1419-4.

Huang XH, Yang SH, Gong JY, et al. Genomic architecture of heterosis for yield traits in rice. Nature. 2016;537(7622):629–33. https://doi.org/10.1038/nature19760.

Huang ZK. Chinese cotton varieties and their genealogies. Beijing: China Agriculture Press; 2007.

Hufford MB, Xu X, van Heerwaarden J, et al. Comparative population genomics of maize domestication and improvement. Nat Genet. 2012;44(7):808–11. https://doi.org/10.1038/ng.2309.

Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801–6. https://doi.org/10.1093/bioinformatics/btm233.

Jia YH, Sun XW, Sun JL, et al. Association mapping for epistasis and environmental interaction of yield traits in 323 cotton cultivars under 9 different environments. PLoS One. 2014;9:e95882. https://doi.org/10.1371/journal.pone.0095882.

Jiang C, Wright RJ, El-Zik KM, et al. Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci U S A. 1998;95(8):4419–24. https://doi.org/10.1073/pnas.95.8.4419.

Kaur S, Zhang X, Mohan A, et al. Genome-wide association study reveals novel genes associated with culm cellulose content in bread wheat (*Triticum aestivum* L.). Front Plant Sci. 2017;8:1913. https://doi.org/10.3389/fpls.2017.01913.

Li C, Zhao TL, Yu HR, et al. Genetic basis of heterosis for yield and yield components explored by QTL mapping across four genetic populations in upland cotton. BMC Genomics. 2018a;19(1):910. https://doi.org/10.1186/s12864-018-5289-2.

Li FJ, Wen WE, He ZH, et al. Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. Theor Appl Genet. 2018b;131(9):1903–24. https://doi.org/10.1007/s00122-018-3122-6.

Li H, Peng ZY, Yang XH, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013; 45(1):43–50. https://doi.org/10.1038/ng.2484.

Li TG, Ma XF, Li NY, et al. Genome-wide association study discovered candidate genes of Verticillium wilt resistance in upland cotton (*Gossypium hirsutum* L.). Plant Biotechnol J. 2017;15:1520–32. https://doi.org/10.1111/pbi.12734.

Liu RZ, Wang BH, Guo WZ, et al. Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. Mol Breed. 2012;29(2):297–311. https://doi.org/10.1007/s11032-011-9547-0.

Liu YY, You SJ, Taylor-Teeples M, et al. BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of *REVOLUTA*. Plant Cell. 2014;26(12):4843–61. https://doi.org/10.1105/tpc.114.128322.

Lu XK, Fu XQ, Wang DL, et al. Resequencing of cv CRI-12 family reveals haplotype block inheritance and recombination of agronomically important genes in artificial selection. Plant Biotechnol J. 2019;17(5):945–55. https://doi.org/10.1111/pbi.13030.

Luikart G, England PR, Tallmon DA, et al. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4(12): 981–94. https://doi.org/10.1038/nrg1226.

Ma LL, Su Y, Nie HS, et al. QTL and genetic analysis controlling fiber quality traits using paternal backcross population in upland cotton. J Cotton Res. 2020;3: 22. https://doi.org/10.1186/s42397-020-00060-6.

Ma XF, Wang ZY, Li W, et al. Resequencing core accessions of a pedigree identifies derivation of genomic segments and key agronomic trait loci

during cotton improvement. Plant Biotechnol J. 2019;17(4):762–75. https://doi.org/10.1111/pbi.13013.
Maik W, Abid MA, Cheema HM, et al. From qutn to Bt cotton: development, adoption and prospects. A review. Tsitol Genet. 2015;49(6):73–85.
Mei HX, Zhu XF, Zhang TZ. Favorable QTL alleles for yield and its components identified by association mapping in Chinese upland cotton cultivars. PLoS One. 2013;8(12):e82193. https://doi.org/10.1371/journal.pone.0082193.
Mengistu DK, Kidane YG, Catellani M, et al. High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. Plant Biotechnol J. 2016;14(9):1800–12. https://doi.org/10.1111/pbi.12538.
Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos Trans R Soc Lond Ser B Biol Sci. 2012;367(1587):409–21. https://doi.org/10.1098/rstb.2011.0249.
Nie XH, Huang C, You CY, et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbreed cultivars in China. BMC Genomics. 2016;17(1):352. https://doi.org/10.1186/s12864-016-2662-x.
Nie XH, Wen TW, Shao PX, et al. High-density genetic variation maps reveal the correlation between asymmetric interspecific introgressions and improvement of agronomic traits in upland and Pima cotton varieties developed in Xinjiang, China. Plant J. 2020;103(2):677–89. https://doi.org/10.1111/tpj.14760.
Noor MA, Bennett SM. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb). 2009;103:439–44. https://doi.org/10.1038/hdy.2009.151.
Raihan MS, Liu J, Huang J, et al. Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 × SK maize population. Theor Appl Genet. 2016;129(8):1465–77. https://doi.org/10.1007/s00122-016-2717-z.
Soltis NE, Atwell S, Shi G, et al. Interactions of tomato and botrytis cinerea genetic diversity: parsing the contributions of host differentiation, domestication, and pathogen variation. Plant Cell. 2019;31(2):502–19. https://doi.org/10.1105/tpc.18.00857.
Sun ZW, Wang XF, Liu ZW, et al. A genome-wide association study uncovers novel genomic regions and candidate genes of yield-related traits in upland cotton.Theor Appl Genet. 2018;131(11):2413–25. https://doi.org/10.1007/s00122-018-3162-y.
Wang BH, Guo WZ, Zhu XF, et al. QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. J Genet Genomics. 2007;34(1):35–45. https://doi.org/10.1016/S1673-8527(07)60005-8.
Wang HT, Huang C, Guo HL, et al. QTL mapping for fiber and yield traits in upland cotton under multiple environments. PLoS One. 2015;10(6):e0130742. https://doi.org/10.1371/journal.pone.0130742.
Wang MJ, Tu LL, Yuan DJ, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet. 2019;51(2):224–9. https://doi.org/10.1038/s41588-018-0282-x.
Xu SD, Pan ZY, Yin FF, et al. Identification of candidate genes controlling fiber quality traits in upland cotton through integration of meta-QTL, significant SNP and transcriptomic data. J Cotton Res. 2020;3(1):34. https://doi.org/10.1186/s42397-020-00075-z.
Xue S, Bradbury PJ, Casstevens TM, Holland JB. Genetic architecture of domestication-related traits in maize. Genetics. 2016;204(1):99–113. https://doi.org/10.1534/genetics.116.191106.
Yamasaki K, Kigawa T, Inoue M, et al. An *Arabidopsis* SBP-domain fragment with a disrupted C-terminal zinc-binding site retains its tertiary structure. FEBS Lett. 2006;580(8):2109–16. https://doi.org/10.1016/j.febslet.2006.03.014.
Yang N, Lu YL, Yang XH, et al. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet. 2014;10(9):e1004573. https://doi.org/10.1371/journal.pgen.1004573.
Yuan DJ, Tang ZH, Wang MJ, et al. The genome sequence of Sea-Island cotton (*Gossypium barbadense*) provides insights into the allopolyploidization and development of superior spinnable fibres. Sci Rep. 2015;5(1):17662. https://doi.org/10.1038/srep17662.
Zhang D, Zhang HY, Hu ZB, et al. Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. PLoS Genet. 2019;15(7):e1008267. https://doi.org/10.1371/journal.pgen.1008267.
Zhang Z, Li JW, Jamshed M, et al. Genome-wide quantitative trait loci reveal the genetic basis of cotton fibre quality and yield-related traits in a *Gossypium hirsutum* recombinant inbred line population. Plant Biotechnol J. 2020;18(1):239–53. https://doi.org/10.1111/pbi.13191.
Zhao GW, Lian Q, Zhang ZH, et al. A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat Genet. 2019;51(11):1607–15. https://doi.org/10.1038/s41588-019-0522-8.
Zheng J, Wu H, Zhu HB, et al. Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. New Phytol. 2019;223(2):705–21. https://doi.org/10.1111/nph.15807.
Zhou ZK, Jiang Y, Wang Z, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015;34(4):441–14. https://doi.org/10.1038/nbt.3096.