RESEARCH    Open Access

# QTL mapping of agronomic and economic traits for four F$_2$ populations of upland cotton

LI Hongge[†], PAN Zhaoe[†], HE Shoupu, JIA Yinhua, GENG Xiaoli, CHEN Baojun, WANG Liru, PANG Baoyin and DU Xiongming[*]

## Abstract

**Background:** Upland cotton (*Gossypium hirsutum*) accounts for more than 90% of the annual world cotton output because of its high yield potential. However, yield and fiber quality traits often show negative correlations. We constructed four F$_2$ populations of upland cotton, using two normal lines (4133B and SGK9708) with high yield potential but moderate fiber quality and two introgression lines (Suyuan04–3 and J02–247) with superior fiber quality, and used them to investigate the genetic basis underlying complex traits such as yield and fiber quality in upland cotton. We also phenotyped eight agronomic and economic traits and mapped quantitative trait loci (QTLs).

**Results:** Extensive phenotype variations and transgressive segregation were found across the segregation populations. We constructed four genetic maps of 585.97 centiMorgan (cM), 752.45 cM, 752.45 cM, and 1 163.66 cM, one for each of the four F$_2$ populations. Fifty QTLs were identified across the four populations (7 for plant height, 27 for fiber quality and 16 for yield). The same QTLs were identified in different populations, including qBW4 and qBW2, which were linked to a common simple sequence repeat (SSR) marker, NAU1255. A QTL cluster containing eight QTLs for six different traits was characterized on linkage group 9 of the 4133B × Suyuan04–3 population.

**Conclusions:** These findings will provide insights into the genetic basis of simultaneous improvement of yield and fiber quality in upland cotton breeding.

**Keywords:** Upland cotton, Yield traits, Fiber quality traits, QTL mapping, Simultaneous improvement

## Background

Cotton represents the main source of natural textile fibers in the world and is the most prevalent raw material used in the textile industry (Wang et al. 2018). High yield and fine fiber-quality are prerequisites to meet the ever-increasing demand. Upland cotton (*Gossypium hirsutum*) accounts for more than 90% of the global cotton production because of its high yield potential and broad adaptability, but it has moderate fiber quality, whereas *G. barbadense* produces exceptionally fine-quality fibers, but with lower fiber yield (Cai et al. 2014; Hu et al. 2019).

Most agronomic and economic traits, such as yield and fiber quality, are quantitative traits that are controlled by multiple loci/genes. Moreover, environmental influence is substantial in the control and expression of these traits. Significant negative correlations between fiber quality traits and yield traits have been reported (Wang et al. 2015; Liu et al. 2018; Zhang et al. 2020). Dissecting the genetic basis of yield and fiber quality is essential for simultaneous improvement of yield and fiber quality.

Molecular genetic methods, especially molecular markers, have been applied widely in cotton in last decade. Recently, the development of the molecular markers was accelaerated with the release of assembled genome sequences of *G. hirsutum* (Li et al. 2015; Zhang et al. 2015; Wang et al. 2018;

* Correspondence: dujeffrey8848@hotmail.com
[†]LI Hongge and PAN Zhaoe contributed equally to this work.
State Key Laboratory of Cotton Biology / Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China

Yang et al. 2019) and *G. barbadense* (Liu et al. 2015; Yuan et al. 2015). Numerous genetic linkage maps, including the intraspecific map of *G. hirsutum* and the interspecific map between *G.hirsutum* and *G.barbadense*, have been constructed using restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). Thousands of quantitative trait loci (QTLs) for yield and fiber quality in cotton are documented in CottonQTLdb (Release 2.3, Said et al. 2013; Said et al. 2015). However, to date, there are few studies about the simultaneous dissection of the genetic basis underlying complex traits and their genetic correlations in multiple upland cotton populations by QTL mapping.

In the present study, we used four $F_2$ populations derived from the hybridization between two *G. hirsutum* normal lines (4133B and SGK9708) and two introgression lines (Suyuan04–3 and J02–247). Four corresponding genetic linkage maps were constructed using SSR markers. QTL mapping was implemented with the integration of the genotypic and phenotypic data of eight agronomic and economic traits, including yield and fiber quality. Our findings will not only contribute to dissecting the genetic basis underlying yield and fiber quality and their genetic correlations but also provide insights into the simultaneous improvement of yield and fiber quality in upland cotton breeding.

## Materials and methods
### Plant materials and field experiments
Two *G. hirsutum* normal lines (4133B and SGK9708) with high yield potential but moderate fiber quality, and two introgression lines (Suyuan04–3 and J02–247) with superior fiber quality were used as the parents to generate four $F_2$ populations in this study. SGK9708 was derived from CCRI41, which is a widely planted cultivar with wide adaptability; 4133B was derived from the hybridization of SGK9708 and the offspring of Gan4104 and CZA (70)33 and has high combining ability; Suyuan04–3 was derived from the distant hybridization of [83–811 × (86–1 × G.*armourianum*)]; and J02–247 was derived from the cross of Suyin45 × Sukang310, and has large cotton bolls as well as superior fiber length and strength. The cotton materials were provided by the National Mid-term Gene Bank for Cotton of China.

In 2014, the seeds of the four parents were sown in Anyang, Henan Province and four cross combinations, 4133B × Suyuan04–3 (hereafter referred to as 4Su), 4133B × J02–247 (4 J), SGK9708 × J02–247 (SgJ), and SGK9708 × 4133B (Sg4) were constructed. Winter breeding of cotton was carried out in Hainan Province. In 2015, the four $F_2$ populations, 4Su, 4 J, SgJ and Sg4, which consisting of 271, 248, 276, and 304 individuals, respectively, were planted in Anyang. Each row was 7 m in length and 0.8 m apart, with 20~22 plants planted in each row. Field management practices were implemented according to local farming practices.

### Trait measurements and statistical analysis
In mid-September, all of the plants in the four $F_2$ populations were investigated for plant height (PH). During the harvesting season, all of the seed cotton was collected and boll weight (BW) and lint percentage (LP) were calculated after the seed cotton had been weighed and ginned. Fiber quality traits, namely fiber length (FL), fiber strength (FS), fiber length uniformity (FU), micronaire (MIC), and fiber elongation (FE), were tested using an HVI 1000 (Uster Technologies, Switzerland) in the Cotton Quality Supervision, Inspection and Testing Center, Ministry of Agriculture, Anyang, China.

The descriptive statistics, namely the maximum, minimum, and mean values, standard deviation, and coefficient of variation (CV), for the eight traits across the four populations, were processed using Microsoft Excel 2013. A correlation matrix was calculated and visualized using the corrplot package in R (Wei and Simko, 2016).

### SSR markers analysis
Young leaves were collected from each plant and stored at −80 °C. Genomic DNA of individuals from the $F_2$ populations and their parents was extracted from young leaf tissues using a modified cetyltrimethylammonium bromide (CTAB) method (Paterson et al. 1993).

Polymorphism detection for the four pairs of parents was performed using 5 713 SSR primers. The primers that amplified stable polymorphic products were selected for genotyping the $F_2$ populations. The SSR primer sequences were downloaded from CottonGen (https://www.cottongen.org; Yu et al. 2014). We used a local Basic Local Alignment Search Tool (BLAST) program (Altschul et al. 1990) to map the SSRs to a physical map. The SSR sequences were queried against the *G. hirsutum* genome sequences (Wang et al. 2018) and the top BLAST hit was selected for further analysis. The separating and silver staining of polymerase chain reaction (PCR) amplified products were performed as detailed by Feng et al. (2015).

### Genetic linkage map construction
The genetic linkage map was constructed using Join-Map 4.0 with the regression mapping method and logarithm of odds (LOD) threshold of 5.0. The Kosambi function was used to convert the recombination frequencies to map distances.

### QTL mapping and analysis
WinQTL Cartographer 2.5 was applied to identify QTLs with the composite interval mapping (CIM) method. The parameters were set as 1.0 cM for the mapping step,

5 for control markers, and 1 000 for permutation tests. QTLs were considered significant if the corresponding LOD score was > 2.5. The additive effect, dominant effect, and $R^2$ (percent of phenotypic variance explained by a QTL) were estimated. QTLs detected for the eight traits were named as q-trait-linkage group number (McCouch et al. 1988). A graphic representation of the linkage groups and QTLs were created using Map-Chart 2.2 (Voorrips 2002).

The action mode of a QTL was represented as dominance degree (|D/A|), i.e., an absolute value of dominant effect (D) divided by additive effect (A) (Stuber et al. 1987). The value was considered as additive if the dominance degree was < 0.2, partial dominance for 0.2~0.8, dominance for 0.81~1.2, overdominance for > 1.2.

The QTLs identified in this study were compared with those in CottonQTLdb (Said et al. 2015) to determine whether the identified QTLs were novel or had been detected previously. QTLs identified in the present study that shared the same or overlapping confidence intervals with QTLs in the CottonQTLdb based on common marker position were considered as QTLs that had been identified in previous studies.

## Results

### Phenotypic variation of the four $F_2$ populations

The phenotypes of eight agronomic and economic traits across four $F_2$ populations were evaluated. Extensive phenotype variations and transgressive segregation were observed (Table 1 and Fig. 1). Transgressive segregation means that the phenotypic values of some individuals were better than those of the superior parent or worse than those of the inferior parent (Reyes 2019). The CV values revealed differences in variability among the eight traits (Table 1). The CV value for LP was low (5.96–7.98%), whereas both CV values of PH and BW were high and similar (PH: 16.9–21.95%; BW: 15.66–19.7%). Among FL, FS, FU, FE, and MIC, the CV value was lowest for FU (1.59–2.61%) and highest for MIC (13.87–22%). Frequency distribution analysis showed normal distribution for seven of the traits, MIC was the exception (Fig. 1), suggesting that these traits were quantitative traits controlled by multiple genes and suitable for QTL mapping.

### Correlation analysis

Correlation analysis between 32 sets of phenotypic data from the eight traits across the four $F_2$ populations revealed significant correlations for different traits within and between populations (Fig. 2). BW and LP were significantly negatively correlated (– 0.87 < r < – 0.62) in three populations (4Su, 4 J, Sg4), whereas BW had significant positive correlations with FS, FU, FE, and MIC (0.13 < r < 0.67) in the 4 J and SgJ populations (Fig. 2).

Negative correlations between BW and FL, FS, FU, and FE (– 0.89 < r < – 0.82) and positive correlations between BW and MIC (r = 0.82) were observed in the 4Su population (Fig. 2). Conversely, LP was positively correlated with FL, FS, FU, and FE (0.95 < r < 0.99) and negatively correlated with MIC (r = – 0.90) (Fig. 2). For the fiber quality traits, positive correlations were found between FL and FS, FU, FE; FS and FU, FE; FU and FE in all four populations (0.19 < r < 0.998) (Fig. 2).

PH and BW in the 4 J population were positively correlated with PH, LP, FL, FS, FU, and FE in the 4Su population (0.25 < r < 0.88) and negatively correlated with BW and MIC in the 4Su population (– 0.776 < r < – 0.772). Conversely, negative correlations (– 0.83 < r < – 0.15) and positive correlations (0.68 < r < 0.77) between LP, FL in 4 J population and corresponding traits in 4Su population were observed (Fig. 2). PH, LP, and FE in the Sg4 population were positively correlated with LP, FL, FS, FU, and FE in 4Su population (0.11 < r < 0.17) and negatively correlated with BW (– 0.14 < r < – 0.13); LP, FE, and MIC in the Sg4 population were positively correlated with PH and BW and negatively correlated with LP and FL (– 0.24 < r < – 0.15) in 4 J population (0.14 < r < 0.27) (Fig. 2).

Overall, within populations, most of the correlations were negative between the two yield traits (BW and LP), whereas, most of the correlations were positive among the fiber quality traits, as well as between BW and the fiber quality traits (Fig. 2). Significant correlations were found between multiple traits among the 4Su, 4 J, and Sg4 populations (Fig. 2), suggesting the influence of the common parent 4133B on the traits.

### Genetic map construction

Five thousand seven hundred thirteen SSR primers were used to detect polymorphisms in the four pairs of parents. Seven hundred thirty-nine polymorphism SSR primers with clearly amplified bands were retained, including 203 polymorphism primers between 4133B and Suyuan04–3 (Additional file 5: Table S1a), 208 between 4133B and J02–407(Additional file 5: Table S1b), 158 between SGK9708 and J02–407 (Additional file 5: Table S1c), and 170 between SGK9708 and 4133B (Additional file 5: Table S1d). The polymorphism rates of the primers for the four comparisons were 3.55, 3.64, 2.77, and 2.98%, respectively.

Joinmap 4.0 was employed to construct a genetic linkage map. For the 4Su population, 71 markers were assigned to 10 linkage groups (LGs) with a total map length of 585.97 cM (Table 2, Additional file 1: Fig. S1, Additional file 6: Table S2a). The average length of the LGs was 58.6 cM, and the average distance of markers was 8.25 cM. The longest LG, LG9, contained the most markers (27), and half of the LGs contained only three markers.

**Table 1** Phenotypic variation of 8 traits in four F$_2$ Populations

| Populaton | Traits | P$_1$ | P$_2$ | F$_2$ populations | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Minimum | Maximum | Mean | SD | CV/% |
| 4Su | PH/cm | 100.04 | 86.56 | 39.00 | 128.00 | 82.64 | 18.14 | 21.95 |
| | BW/g | 6.59 | 6.54 | 2.50 | 8.79 | 5.37 | 0.97 | 18.07 |
| | LP/% | 42.85 | 41.62 | 33.33 | 51.38 | 41.25 | 2.55 | 6.19 |
| | FL/mm | 28.80 | 29.40 | 26.01 | 33.33 | 29.77 | 1.47 | 4.94 |
| | FS/cN/tex | 29.30 | 34.57 | 23.42 | 38.71 | 30.09 | 2.84 | 9.43 |
| | FU/% | 85.87 | 86.47 | 79.70 | 89.40 | 85.34 | 1.67 | 1.95 |
| | FE/% | 6.85 | 6.80 | 6.20 | 7.30 | 6.85 | 0.20 | 2.91 |
| | MIC | 5.27 | 5.00 | 2.21 | 5.49 | 4.14 | 0.71 | 17.08 |
| 4 J | PH/cm | 100.04 | 89.74 | 45.00 | 129.00 | 86.45 | 16.38 | 18.95 |
| | BW/g | 6.59 | 7.06 | 2.87 | 7.80 | 5.28 | 0.89 | 16.87 |
| | LP/% | 42.85 | 41.37 | 32.37 | 47.84 | 39.85 | 2.38 | 5.96 |
| | FL/mm | 28.80 | 30.67 | 27.24 | 34.47 | 31.00 | 1.59 | 5.14 |
| | FS/cN/tex | 29.30 | 33.00 | 24.21 | 40.08 | 31.86 | 2.66 | 8.34 |
| | FU/% | 85.87 | 85.43 | 79.20 | 90.60 | 86.00 | 1.58 | 1.84 |
| | FE/% | 6.85 | 6.95 | 6.60 | 7.70 | 7.17 | 0.20 | 2.78 |
| | MIC | 5.27 | 5.27 | 2.24 | 5.64 | 4.36 | 0.61 | 13.96 |
| SgJ | PH/cm | 82.72 | 89.74 | 45.00 | 124.00 | 80.78 | 13.65 | 16.90 |
| | BW/g | 6.20 | 7.06 | 2.41 | 7.63 | 4.77 | 0.94 | 19.70 |
| | LP/% | 43.63 | 41.37 | 20.32 | 45.33 | 37.97 | 2.74 | 7.22 |
| | FL/mm | 29.33 | 30.67 | 26.60 | 34.67 | 30.72 | 1.32 | 4.29 |
| | FS/cN/tex | 28.20 | 33.00 | 24.11 | 39.79 | 31.08 | 2.70 | 8.68 |
| | FU/% | 85.08 | 85.43 | 78.60 | 89.80 | 85.27 | 2.22 | 2.61 |
| | FE/% | 6.70 | 6.95 | 6.10 | 7.50 | 6.95 | 0.23 | 3.37 |
| | MIC | 5.15 | 5.27 | 2.02 | 5.90 | 4.08 | 0.90 | 22.00 |
| Sg4 | PH/cm | 82.72 | 100.04 | 44.00 | 130.00 | 89.33 | 16.78 | 18.79 |
| | BW/g | 6.20 | 6.59 | 3.11 | 7.30 | 5.26 | 0.82 | 15.66 |
| | LP/% | 43.63 | 42.85 | 35.13 | 51.57 | 43.53 | 3.47 | 7.98 |
| | FL/mm | 29.33 | 28.80 | 25.51 | 32.59 | 29.66 | 1.06 | 3.55 |
| | FS/cN/tex | 28.20 | 29.30 | 23.40 | 32.90 | 27.46 | 1.62 | 5.90 |
| | FU/% | 85.08 | 85.87 | 81.50 | 90.50 | 86.09 | 1.37 | 1.59 |
| | FE/% | 6.70 | 6.85 | 6.20 | 7.50 | 6.86 | 0.20 | 2.95 |
| | MIC | 5.15 | 5.27 | 2.81 | 5.97 | 4.70 | 0.65 | 13.87 |

Note: a: The P$_1$ of 4Su, 4 J, SgJ and Sg4 population is 4133B, 4133B, SGK9708 and SGK9708, respectively; b: The P$_2$ of 4Su, 4 J, SgJ and Sg4 population is Suyuan04–3, J02–247, J02–247 and 4133B, respectively; c: SD,Standard Deviation; d: CV,Coefficient of variation

For the 4 J population, 61 markers were assigned to 10 linkage groups with a total map length of 752.45 cM (Table 2, Additional file 2: Fig. S2, Additional file 6: Table S2b). The average length of the LGs was 75.2 cM, and the average distance of markers was 12.34 cM. LG7 contained the most markers (21) and LG3 contained the least markers (3).

For the SgJ population, 83 markers, approximately half of the 158 polymorphism markers, were assigned to 15 linkage groups with a total map length of 855.04 cM (Table 2, Additional file 3: Fig. S3,

Additional file 6: Table S2c). The average length of the linkage groups was 57 cM. The longest average distance of markers was 21.46 cM on LG13 and the shortest was 1.06 cM on LG14.

For the Sg4 population, 52 markers, approximately one-third of the 170 polymorphism markers were assigned to nine linkage groups with a total map length of 1 163.66 cM (Table 2, Additional file 4: Fig. S4, Additional file 6: Table S2d). The average length of linkage groups was 129.3 cM, and the average distance of markers was 22.38 cM.
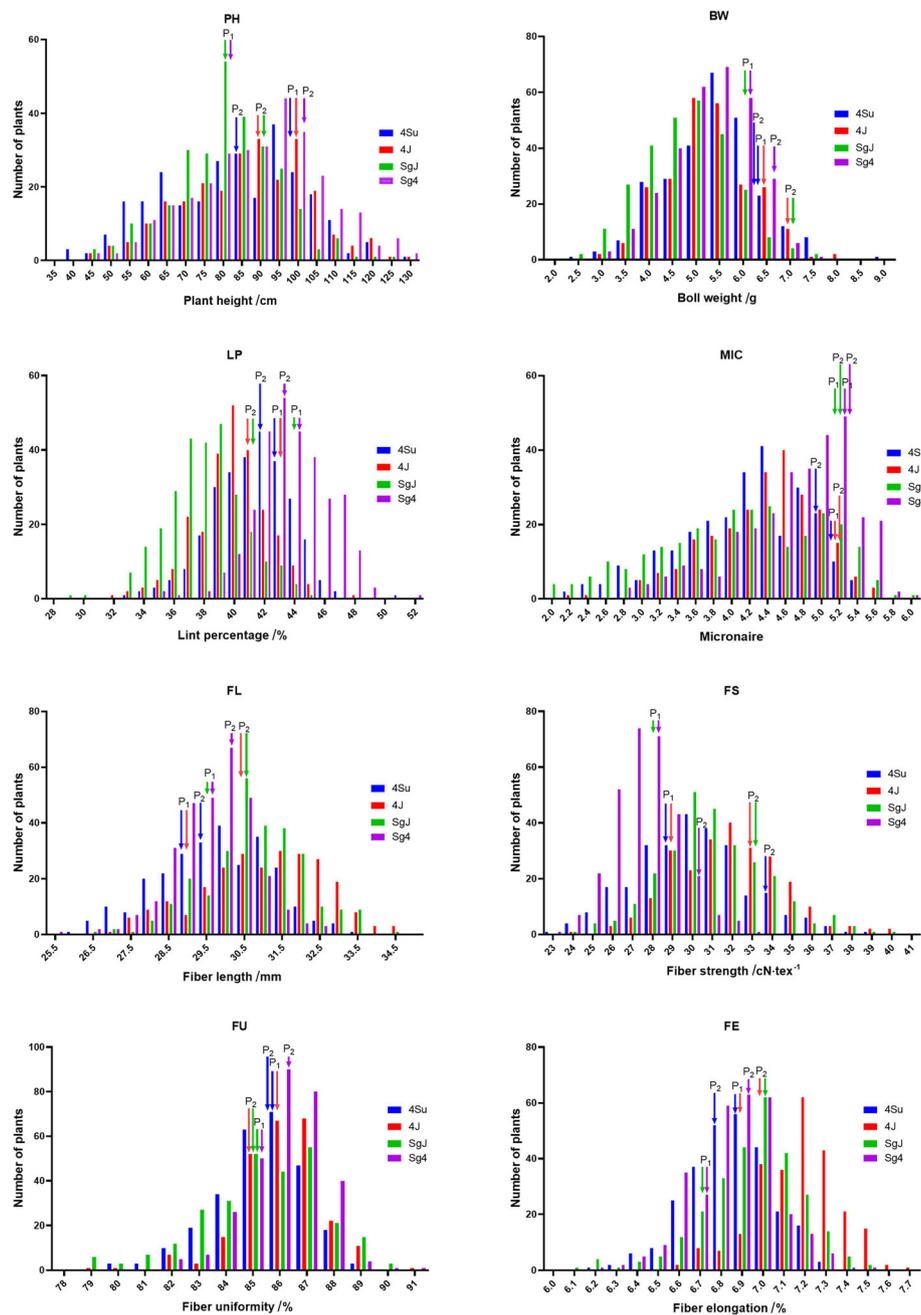
**Fig. 1** Histogram of the frequency distribution for 8 traits across 4 populations. Plant height (PH), boll weight (BW), lint percentage (LP), fiber length (FL), fiber strength (FS), fiber length uniformity (FU), micronaire (MIC), fiber elongation (FE)

## QTL mapping

The WinQTL Cartographer 2.5 with the CIM algorithm was used to identify the QTLs for eight traits in the four $F_2$ populations. As a result, a total of 50 QTLs with $R^2$ of 0.1–59.24% were identified, 27 were for fiber quality traits and 16 were for yield traits. A total of 23, 4, 8, and 15 QTLs were detected in the 4Su, 4 J, SgJ, and Sg4 populations, respectively (Table 3, Fig. 3). LG9 in the 4Su

population harbored the highest number of QTLs (13), following by LG6 (6) and LG1 (5) in the Sg4 population.

Seven QTLs for PH were identified, but six of them in the 4Su population had only minor effects (0.11% < $R^2$ < 4.02%; Table 3, Fig. 3). The additive effects of QTLs qPH2−1 and qPH2−2, which with the higher $R^2$ (2.66% and 4.02%), were positive, indicating that the favorable alleles were from Suyuan04−3. The action modes of
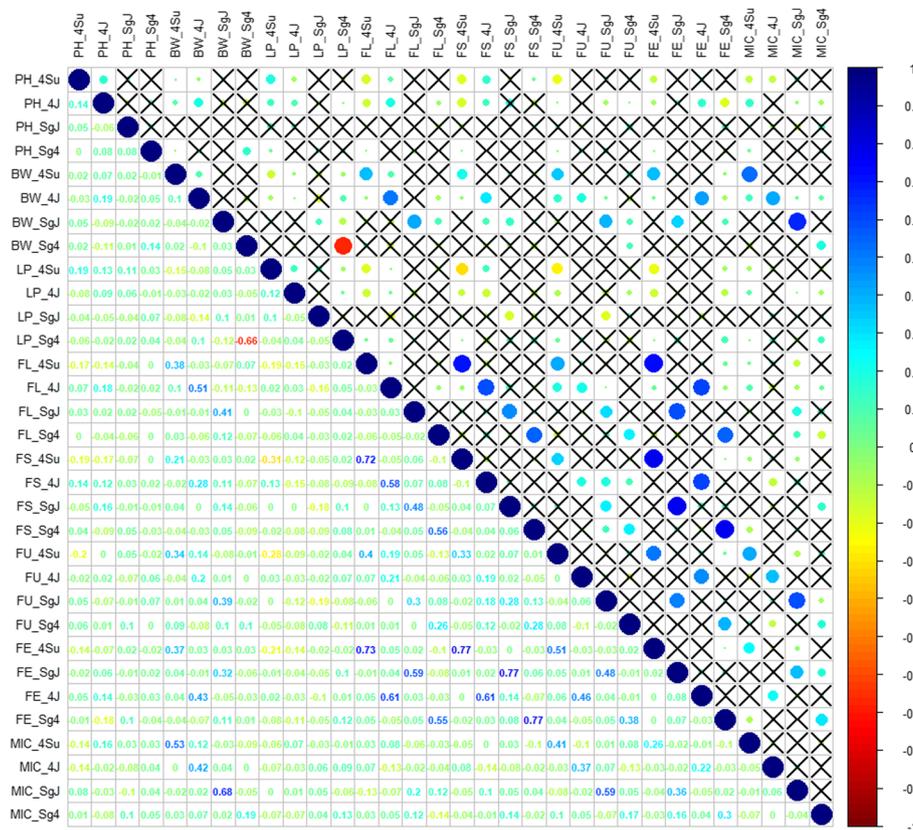
**Fig. 2** Correlation analysis among 8 traits. The magnitude of the correlation is indicated by different colors of numbers at the left diagonal and circles at the right diagonal. The circles marked by cross indicate no significant correction was observed between traits ($P > 0.05$)

qPH2−1 and qPH2−2 were over-dominance according to the dominance degree values.

Eight QTLs for BW were identified with $R^2$ of 1.17%~9.31% in the 4 J (1), SgJ (1), and Sg4 (6) populations (Table 3, Fig. 3). It is noteworthy the LGs that harbored one of QTLs in the 4 J (qBW4) and SgJ (qBW2) populations were anchored to chromosome A05, and the common SSR marker, NAU1255, was detected close to the QTL interval implying that NAU1255 was closely linked to BW. Furthermore, the directions of the additive and dominance effects of these QTLs were the same.

Eight QTLs for LP were identified with $R^2$ of 1.68% ~ 18.11% in the 4Su (2), 4 J (2), and SgJ (4) populations (Table 3, Fig. 3). The additive effects of two major QTLs, qLP7$_{4J}$ and qLP2 with $R^2 > 10$% were negative, indicating that the favorable alleles were from J02−247. The action modes of qLP7$_{4J}$ and qLP2 were over-dominance and dominance, respectively.

Eleven QTLs for FL were identified in the 4Su (6), 4 J (1), and SgJ (4) populations (Table 3, Fig. 3). Multiple QTLs were in the same LG of a population, for example, qFL9−1, qFL9−2, and qFL9−3 with $R^2$ of 0.35%~7.70% were in LG9 of the 4Su population.

Interestingly, both LG7 in the 4Su population and LG6 in the SgJ population were anchored to chromosome A13 (Table 2). The common SSR markers, BNL2449 and NAU1211, were detected near QTLs qFL7$_{4Su}$ and qFL6, hinting that BNL2449 and NAU1211 may be closely linked to FL. In addition, the additive effect of QTL qFL2−2 was positive, suggesting that the favorable alleles come from the male parents, Suyuan04−3 and J02−247, that is endowed with superior fiber quality.

Five QTLs for FS were identified: four with $R^2$ of 2.95%~7.15% in the 4Su population and one major QTL with $R^2$ of 15.10% in the Sg4 population (Table 3, Fig. 3). The additive effects of these four QTLs in the 4Su population were positive, whereas the additive effect of the one major QTL in the Sg4 population was negative, implying that the parent, 4133B may not have conferred the favorable allele.

Only two QTLs for FU were identified with $R^2$ of 0.10%~1.21% in the same LG of the 4Su population (Table 3, Fig. 3).

Four QTLs for FE were identified with $R^2$ of 0.16%~5.62% in the 4Su, SgJ and Sg4 populations (Table 3, Fig. 3). The additive effect of one QTL, qFE8, was

**Table 2** Summary of 4 genetic linkage maps

| Linkage Maps | LG | Anchored Chromosome | LG length /cM | Marker Number of LG | Average Marker Interval /cM |
|---|---|---|---|---|---|
| 4Su | LG1 | A02 | 23.10 | 3 | 7.70 |
| | LG2 | A03 | 25.10 | 4 | 6.27 |
| | LG3 | A05 | 110.78 | 7 | 15.83 |
| | LG4 | A07 | 112.29 | 14 | 8.02 |
| | LG5 | A08 | 49.11 | 4 | 12.28 |
| | LG6 | A12 | 47.60 | 3 | 15.87 |
| | LG7 | A13 | 14.92 | 3 | 4.97 |
| | LG8 | D04 | 50.40 | 3 | 16.80 |
| | LG9 | D09/D11 | 148.20 | 27 | 5.49 |
| | LG10 | D10 | 4.47 | 3 | 1.49 |
| 4 J | LG1 | A01a | 57.46 | 7 | 8.21 |
| | LG2 | A01b | 114.11 | 5 | 22.82 |
| | LG3 | A03 | 3.50 | 3 | 1.17 |
| | LG4 | A05 | 96.99 | 21 | 4.62 |
| | LG5 | A12a | 115.23 | 4 | 28.81 |
| | LG6 | A12b | 83.28 | 4 | 20.82 |
| | LG7 | A12c | 137.15 | 4 | 34.29 |
| | LG8 | D01 | 64.95 | 5 | 12.99 |
| | LG9 | D07 | 49.76 | 4 | 12.44 |
| | LG10 | D09 | 30.02 | 4 | 7.51 |
| SgJ | LG1 | A02 | 68.75 | 4 | 17.19 |
| | LG2 | A05 | 87.86 | 13 | 6.76 |
| | LG3 | A09 | 56.60 | 4 | 14.15 |
| | LG4 | A11 | 56.39 | 4 | 14.10 |
| | LG5 | A12 | 66.04 | 4 | 16.51 |
| | LG6 | A13 | 63.24 | 6 | 10.54 |
| | LG7 | D01 | 28.31 | 7 | 4.04 |
| | LG8 | D02 | 44.59 | 6 | 7.43 |
| | LG9 | D05a | 90.77 | 7 | 12.97 |
| | LG10 | D05b | 25.66 | 5 | 5.13 |
| | LG11 | D07 | 63.76 | 3 | 21.25 |
| | LG12 | D09 | 61.02 | 10 | 6.10 |
| | LG13 | D10a | 64.38 | 3 | 21.46 |
| | LG14 | D10b | 3.17 | 3 | 1.06 |
| | LG15 | D12 | 74.50 | 4 | 18.63 |
| Sg4 | LG1 | A01a | 260.79 | 9 | 28.98 |
| | LG2 | A01b | 86.60 | 6 | 14.43 |
| | LG3 | A05 | 82.07 | 3 | 27.36 |
| | LG4 | A07 | 72.66 | 5 | 14.53 |
| | LG5 | A13 | 96.95 | 4 | 24.24 |
| | LG6 | D02/A05 | 257.58 | 13 | 19.81 |
| | LG7 | D08 | 80.68 | 3 | 26.89 |
| | LG8 | D09a | 167.76 | 6 | 27.96 |
| | LG9 | D09b | 58.58 | 3 | 19.53 |

**Table 3** Summary of QTLs identified in four populations for 8 traits

| Traits | QTLs | Population | LG | Position /cM | LOD | $R^2$ /% | Add. | Dom. | Dom.De |
|--------|------|-----------|-----|-------------|------|---------|-------|--------|--------|
| PH | qPH1 | 4Su | LG1 | 0.01 | 3.93 | 1.55 | −5.05 | −4.11 | 0.81 |
| | qPH2–1 | 4Su | LG2 | 7.61 | 4.37 | 2.66 | 0.34 | 26.39 | 78.28 |
| | qPH2–2 | 4Su | LG2 | 13.61 | 4.06 | 4.02 | 0.66 | −26.83 | 40.74 |
| | qPH4 | 4Su | LG4 | 95.31 | 4.18 | 2.55 | −7.01 | −4.82 | 0.69 |
| | qPH9–1 | 4Su | LG9 | 105.81 | 2.67 | 0.58 | −4.12 | −6.09 | 1.48 |
| | qPH9–2 | 4Su | LG9 | 118.61 | 2.52 | 1.98 | −6.12 | −7.57 | 1.24 |
| | qPH6 | Sg4 | LG6 | 172.41 | 3.10 | 0.11 | 2.30 | 8.27 | 3.59 |
| BW | qBW4 | 4 J | LG4 | 13.31 | 4.08 | 5.66 | −0.34 | −0.07 | 0.19 |
| | qBW2 | SgJ | LG2 | 33.21 | 4.46 | 1.98 | −0.30 | −0.22 | 0.74 |
| | qBW1–1 | Sg4 | LG1 | 87.01 | 3.98 | 1.17 | −0.14 | −0.75 | 5.25 |
| | qBW1–2 | Sg4 | LG1 | 227.51 | 9.01 | 9.31 | 0.54 | 0.35 | 0.64 |
| | qBW6–1 | Sg4 | LG6 | 17.51 | 2.66 | 4.26 | 0.24 | 0.02 | 0.08 |
| | qBW6–2 | Sg4 | LG6 | 94.51 | 3.32 | 2.74 | 0.30 | 0.46 | 1.54 |
| | qBW6–3 | Sg4 | LG6 | 131.81 | 3.49 | 2.21 | 0.25 | 0.35 | 1.41 |
| | qBW9 | Sg4 | LG9 | 7.71 | 2.87 | 2.44 | −0.22 | − 0.08 | 0.39 |
| LP | qLP9 | 4Su | LG9 | 95.31 | 3.40 | 1.68 | −0.79 | −0.85 | 1.07 |
| | qLP10 | 4Su | LG10 | 2.71 | 5.21 | 1.75 | −0.93 | −0.74 | 0.80 |
| | qLP5$_{4J}$ | 4 J | LG5 | 31.01 | 3.15 | 7.49 | 0.52 | −0.85 | 1.63 |
| | qLP7$_{4J}$ | 4 J | LG7 | 22.01 | 5.03 | 18.11 | −1.01 | 1.27 | 1.25 |
| | qLP2 | SgJ | LG2 | 1.01 | 11.31 | 11.74 | −9.67 | 9.47 | 0.98 |
| | qLP5$_{SgJ}$ | SgJ | LG5 | 26.41 | 3.15 | 5.82 | −0.95 | − 0.01 | 0.01 |
| | qLP6 | Sg4 | LG6 | 181.41 | 5.07 | 5.02 | 0.81 | 0.97 | 1.19 |
| | qLP7$_{Sg4}$ | Sg4 | LG7 | 63.51 | 2.83 | 7.63 | 0.53 | −1.47 | 2.77 |
| FL | qFL2–1 | 4Su | LG2 | 4.61 | 2.90 | 7.70 | 0.08 | −1.57 | 19.30 |
| | qFL2–2 | 4Su | LG2 | 20.11 | 2.94 | 4.52 | −0.79 | 0.03 | 0.04 |
| | qFL7$_{4Su}$ | 4Su | LG7 | 6.01 | 2.82 | 6.01 | 0.41 | −0.26 | 0.65 |
| | qFL9–1 | 4Su | LG9 | 96.31 | 7.90 | 3.25 | 0.72 | 0.86 | 1.20 |
| | qFL9–2 | 4Su | LG9 | 102.81 | 8.05 | 5.41 | 0.77 | 0.59 | 0.77 |
| | qFL9–3 | 4Su | LG9 | 120.61 | 4.48 | 1.62 | 0.54 | 1.08 | 2.01 |
| | qFL5 | 4 J | LG5 | 64.31 | 3.32 | 4.22 | 0.11 | −2.18 | 19.44 |
| | qFL6 | SgJ | LG6 | 7.01 | 2.60 | 2.70 | 0.42 | 0.21 | 0.51 |
| | qFL7$_{SgJ}$ | SgJ | LG7 | 6.01 | 2.88 | 6.76 | 0.36 | −0.27 | 0.73 |
| | qFL15–1 | SgJ | LG15 | 26.61 | 2.61 | 0.35 | 0.15 | 0.49 | 3.19 |
| | qFL15–2 | SgJ | LG15 | 68.61 | 2.74 | 3.43 | 0.81 | 0.12 | 0.15 |
| FS | qFS1 | 4Su | LG1 | 3.01 | 2.51 | 2.95 | 0.75 | 0.14 | 0.19 |
| | qFS7 | 4Su | LG7 | 10.61 | 2.80 | 3.13 | 0.77 | 0.12 | 0.15 |
| | qFS9–1 | 4Su | LG9 | 96.31 | 8.72 | 7.15 | 1.52 | 1.16 | 0.76 |
| | qFS9–2 | 4Su | LG9 | 101.81 | 7.75 | 6.09 | 1.35 | 0.73 | 0.54 |
| | qFS2 | Sg4 | LG2 | 15.91 | 3.71 | 15.10 | −0.72 | 0.73 | 1.02 |
| FU | qFU9–1 | 4Su | LG9 | 41.71 | 2.92 | 1.24 | 0.40 | 0.38 | 0.95 |
| | qFU9–2 | 4Su | LG9 | 72.21 | 2.52 | 0.10 | −1.11 | 2.15 | 1.93 |
| FE | qFE9 | 4Su | LG9 | 96.31 | 4.08 | 0.16 | 0.06 | 0.12 | 2.07 |
| | qFE8 | SgJ | LG8 | 24.21 | 3.73 | 5.62 | −0.07 | 0.00 | 0.04 |
| | qFE1 | Sg4 | LG1 | 245.01 | 3.76 | 1.46 | 0.05 | 0.06 | 1.33 |

**Table 3** Summary of QTLs identified in four populations for 8 traits (Continued)

| Traits | QTLs | Population | LG | Position /cM | LOD | $R^2$ /% | Add. | Dom. | Dom.De |
|--------|------|-----------|-----|------------|------|--------|-------|-------|--------|
|        | qFE6 | Sg4 | LG6 | 152.81 | 4.29 | 1.05 | 0.04 | 0.13 | 3.57 |
| MIC    | qMIC9–1 | 4Su | LG9 | 41.71 | 2.91 | 1.63 | 0.00 | 0.32 | 92.03 |
|        | qMIC9–2 | 4Su | LG9 | 100.81 | 2.51 | 6.29 | −0.20 | 0.13 | 0.66 |
|        | qMIC1–1 | Sg4 | LG1 | 52.31 | 3.58 | 0.47 | −0.02 | −0.74 | 30.64 |
|        | qMIC1–2 | Sg4 | LG1 | 88.01 | 5.54 | 0.15 | −0.02 | −0.95 | 47.68 |
|        | qMIC2 | Sg4 | LG2 | 68.31 | 4.10 | 59.24 | 0.10 | −0.92 | 9.41 |

Note: *Add.* Additive effect; *Dom* Dominant effect; *Dom. De* Dominance degree

negative and the action mode was additive, whereas the additive effects of the other three QTLs were positive and the action modes were over-dominance.

Five QTLs for MIC were detected in three LGs in the 4Su and Sg4 populations (Table 3, Fig. 3). A major QTL, qMIC2 with $R^2$ of up to 59.24%, was in LG2 of the Sg4 population, the other four QTLs were minor with $R^2$ 0.15%~6.29%. The dominance degree values of all QTLs, except qMIC9–2, were up to 9.41~92.03, suggesting the action modes were over-dominance.

A hotspot region was detected in LG9 of the 4Su population (Fig. 3a). Three QTLs (qFL9–1, qFS9–1, qFE9) were identified only at the position of 96.31 cM and the expansion of this region from 95.31 cM to 105.81 cM revealed the presence of eight QTLs for six traits: PH (105.81 cM), LP (95.31 cM), FL (96.31 cM, 102.81 cM), FS (96.31 cM, 101.81 cM), FE (96.31 cM) and MIC (100.81 cM). Therefore, this hotspot region may be an important genome region that affects agronomic and economic traits in cotton. Two other QTLs, qFU9–1 and qMIC9–1, were identified in the same LG9 at 41.71 cM.

### QTLs comparison and analysis
We compared all of the identified QTLs with the QTLs in CottonQTLdb database. The results showed that one-fifth of our QTLs (10/50) overlapped with previously reported QTLs, illustrating the reliability of our QTL mapping and indicating the other 40 QTLs were novel QTLs. The 10 common QTLs were reported to be associated with FL (4), FS (2), PH (1), BW (1), LP (1), and FE (1) traits. QTLs for FL were the most identified QTLs in both the present study (11) and CottonQTLdb database (494), which may have increased the probability of a hit.

QTLs for different traits that shared the same or overlapping confidence intervals were considered to be in QTL clusters. In the present study, a total of nine QTL clusters were identified in the 4Su (5), 4 J (1), and Sg4 (3) populations. The QTL cluster harboring the most QTLs was the hotspot region described above, with eight QTLs for six traits. Another QTL cluster in the same LG (LG9 in the 4Su population) contained QTLs for FU and MIC (Fig. 3a).

As we know, BW and LP represented yield traits, FL, FS, FU, FE and MIC represented fiber quality traits. With this prerequisite, the analysis of paired trait QTLs was employed. There were 19 paired trait QTLs within six paired traits (BW and FL, or FE; LP and FL, FS, FU, or FE) that had significant medium or high positive correlations ($|r| > 0.3$) in the $F_2$ populations. Six of the 19 paired trait QTLs had the same direction of addictive effects (Additional file 7: Table S3).

## Discussion
To dissect the genetic basis underlying yield and fiber quality as well as their genetic correlations, two upland cotton normal lines (4133B and SGK9708) and two introgression lines (Suyuan04–3 and J02–247) were selected as parents, and four $F_2$ populations were constructed. Among these populations, the female parents of 4Su, 4 J and SgJ were potential high yield lines, and the male parents were superior fiber quality lines. Thus, extensive phenotypic variation was observed in the cross combinations whose parents had a distant kinship. All of the targeted traits exhibited normal distribution patterns across the four $F_2$ populations (Table 1, Fig. 1), suggesting that these traits were quantitative traits controlled by multiple genes.

Furthermore, all of the traits exhibited transgressive segregation and many individuals with transgressive phenotypes were found (Table 1, Fig. 1). For example, all of the median values of FL and FS in the 4Su, 4 J, and SgJ populations were higher than or nearly 30, fiber with two quality values over 30 (FL ≥ 30 mm and FS ≥ 30 cN·tex$^{-1}$) is generally considered as fine quality one. In plant breeding, transgressive segregation provides an adaptive advantage for traits (Reyes 2019). To a certain extent, high yield and fine–quality fibers are the outcome of the adaptation of cotton. Therefore, it is not surprising that many instances of transgressive segregation were observed for these traits in the $F_2$ populations. Furthermore, some of these transgressive lines can be used to breed for high-quality fiber. However, these characteristics imply that the favorable alleles of the fiber quality traits were generally from the introgression lines parents.
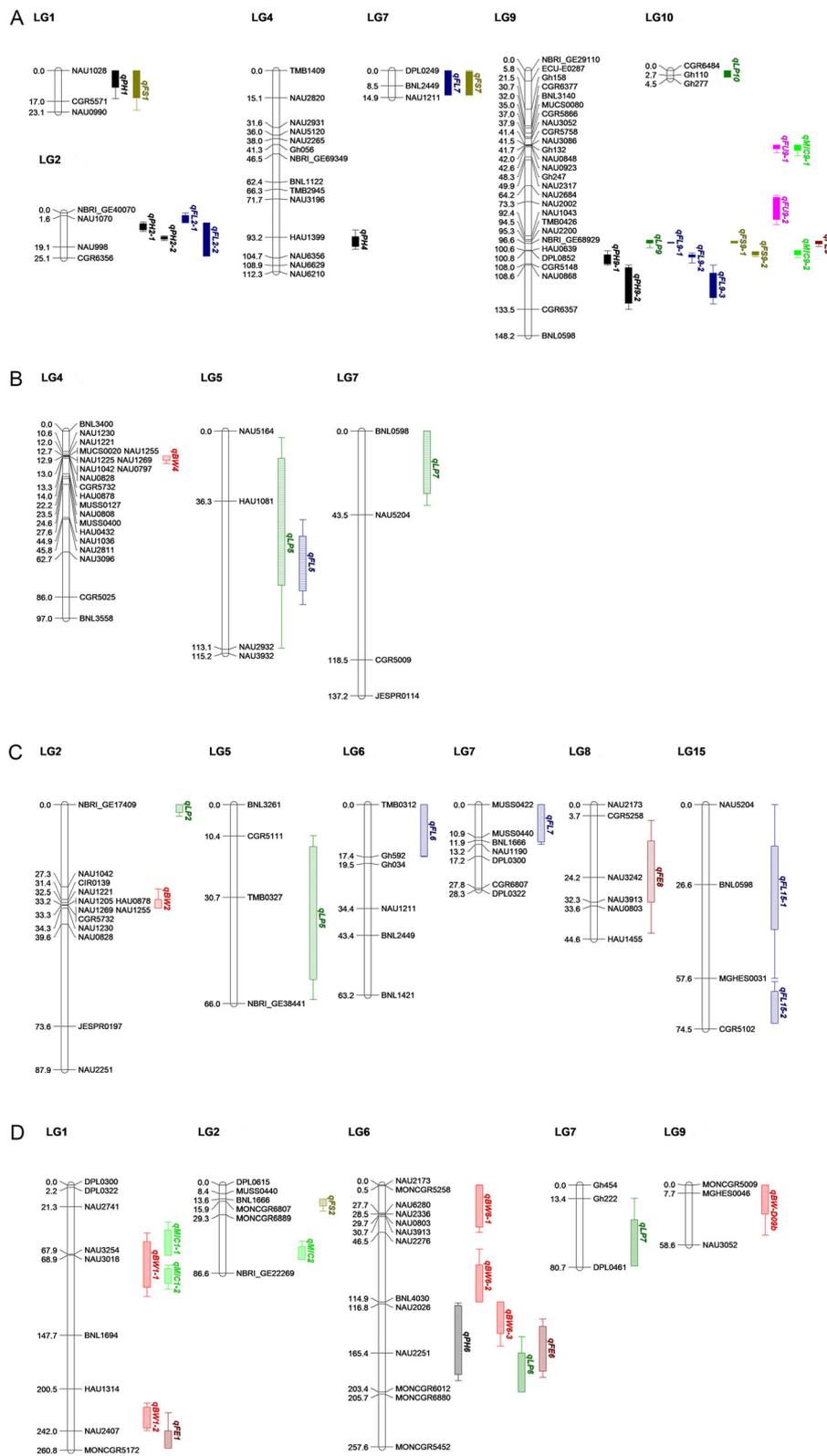
**Fig. 3** QTLs identified for 8 traits across 4 populations. **a** 4Su population; **b** 4 J population; **c** SgJ population; **d** Sg4 population

Quantitative traits are known to be influenced by the environment. Therefore, to identify stable QTLs, mapping populations are usually planted in multiple environments (Tang et al. 2015; Diouf et al. 2018; Zhang et al. 2020). However, we detected many stable QTLs including qBW4 and qBW2, using the four $F_2$ populations. Although qBW4 was detected in the 4 J population and qBW2 was detected in the SgJ population, they had a common marker and the corresponding LGs were on the same chromosome (A05). Therefore, these two QTLs could be considered as one. This study has demonstrated an alternative method of detecting stable QTLs through multiple populations.

The detection of QTLs cluster is consistent with the results of previous studies, that found that QTLs for fiber quality were clustered on the same chromosome, D09, where the majority of markers in LG9 were mapped, harboring important loci for fiber quality traits (He et al. 2007; Qiao et al. 2019). These results illustrated that QTLs in clusters may be closely linked or have pleiotropic effects (Vikram et al. 2015; Zhao et al. 2016; You et al. 2019; Yuan et al. 2018), which explains the significant phenotypic correlations or linkage drag found between related traits (Zhang et al. 2020). The simultaneous improvement of paired trait QTLs that have the same QTL additive effect direction and significant medium or high positive correlations should be made (Zhang et al. 2020). In the present study, we identified six paired trait QTLs with significant positive correlation and additive mode of gene action (Additional file 7: Table S3) that may be valuable for further simultaneous improvement of yield and fiber quality traits.

We found that qLP9 for LP and qFL9−1 and qFL9−2 for FL were in the same QTL cluster in LG9 of the 4Su population. Furthermore, a high positive correlation between LP and FL and the same direction of the addictive effect were observed. We plan to use the detected QTL clusters in a future breeding program by fine−mapping the QTL clusters using large scale segregating populations and gene-editing technology to break the negative correlation and further improve yield and fiber quality.

## Conclusions

In this study, four $F_2$ populations were derived from hybridization between two G. hirsutum normal lines and two introgression lines. Four corresponding genetic linkage maps were constructed and QTL mapping was implemented following the integration of phenotypic data of eight agronomic and economic traits. Fifty QTLs across the four populations were detected; 27 were for fiber quality traits and 16 were for yield traits. QTLs in the same cluster, such as qLP9 for LP and qFL9−1 for FL, were prioritized for further research. Our results will help to dissect the genetic basis underlying yield and fiber quality, and lay a promising foundation for simultaneously improving yield and fiber quality in upland cotton breeding programs.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-020-00076-y.

**Additional file 1: Figure S1**. The genetic map of 4Su population.

**Additional file 2: Figure S2**. The genetic map of 4 J population.

**Additional file 3: Figure S3**. The genetic map of SgJ population.

**Additional file 4: Figure S4**. The genetic map of Sg4 population.

**Additional file 5: Table S1**. The polymorphism primers information between four pairs of parents.

**Additional file 6: Table S2**. The linkage group of four populations.

**Additional file 7: Table S3**. Integrated analysis of paired trait QTLs and direction of additive.

### References

Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Cai C, Ye W, Zhang T, et al. Association analysis of fiber quality traits and exploration of elite alleles in upland cotton cultivars/accessions (Gossypium hirsutum L.). J Integr Plant Biol. 2014;56(1):51–62. https://doi.org/10.1111/jipb.12124.

Diouf L, Magwanga RO, Gong W, et al. QTL mapping of fiber quality and yield-related traits in an intra-specific upland cotton using genotype by sequencing (GBS). Int J Mol Sci. 2018;19(2):441. https://doi.org/10.3390/ijms19020441.

Feng H, Guo L, Wang G, et al. The negative correlation between fiber color and quality traits revealed by QTL analysis. PLoS One. 2015;10(6):e0129490. https://doi.org/10.1371/journal.pone.0129490.

He DH, Lin ZX, Zhang XL, et al. QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of Gossypium hirsutum x Gossypium barbadense. Euphytica. 2007;153(1–2):181–97. https://doi.org/10.1007/s10681-006-9254-9.

Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet. 2019;51(4):739–48. https://doi.org/10.1038/s41588-019-0371-5.

Li F, Fan G, Lu C, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30. https://doi.org/10.1038/nbt.3208.

Liu R, Gong J, Xiao X, et al. GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. Front Plant Sci. 2018;9:1067. https://doi.org/10.3389/fpls.2018.01067.

Liu X, Zhao B, Zheng H, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep. 2015;5:14139. https://doi.org/10.1038/srep14139.

McCouch SR, Kochert G, Yu ZH, et al. Molecular mapping of rice chromosomes. Theor Appl Genet. 1988;76(6):815–29. https://doi.org/10.1007/bf00273666.

Paterson AH, Brubaker CL, Wendel JF. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol Biol Rep. 1993;11(2):122–7. https://doi.org/10.1007/bf02670470.

Qiao W, Yan G, Shi J, et al. QTL mapping of fiber quality in low-generation upland cotton populations and functional annotation of candidate genes. Cotton Sci. 2019;31(4):282–96. https://doi.org/10.11963/1002-7807.qwqhq.20190530.

Reyes BGL. Genomic and epigenomic bases of transgressive segregation - new breeding it paradigm for novel plant phenotypes. Plant Sci. 2019;288:110213. https://doi.org/10.1016/j.plantsci.2019.110213.

Said JI, Knapka JA, Song M, et al. Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* x *G. barbadense* populations. Mol Genet Genomics. 2015;290(4):1615–25. https://doi.org/10.1007/s00438-015-1021-y.

Stuber CW, Edwards MD, Wendel JF. Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Sci. 1987;27(4):639–48. https://doi.org/10.2135/cropsci1987.0011183X002700040006x.

Tang S, Teng Z, Zhai T, et al. Construction of genetic map and QTL analysis of fiber quality traits for upland cotton (*Gossypium hirsutum* L.). Euphytica. 2015;201(2):195–213. https://doi.org/10.1007/s10681-014-1189-y.

Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered. 2002;93(1):77–8. https://doi.org/10.1093/jhered/93.1.77 .

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 M, Tu L, Yuan D, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat Genet. 2018; 51:224–9. https://doi.org/10.1038/s41588-018-0282-x.

Yang Z, Ge X, Yang Z, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. Nat Commun. 2019;10(1):2989. https://doi.org/10.1038/s41467-019-10820-x.

You WJ, Feng YR, Shen YH, et al. Silencing of *NbCMT3s* has pleiotropic effects on development by interfering with autophagy-related genes in *Nicotiana benthamiana*. Plant Cell Physiol. 2019;60(5):1120–35. https://doi.org/10.1093/pcp/pcz034.

Yu J, Jung S, Cheng CH, et al. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 2014;42(D1):D1229–36. https://doi.org/10.1093/nar/gkt1064.

Yuan D, Tang Z, Wang M, 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;5:17662. https://doi.org/10.1038/srep17662.

Yuan X, Huang P, Wang R, et al. A zinc finger transcriptional repressor confers pleiotropic effects on rice growth and drought tolerance by down-regulating stress-responsive genes. Plant Cell Physiol. 2018;59(10):2129–42. https://doi.org/10.1093/pcp/pcy133.

Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–7. https://doi.org/10.1038/nbt.3207.

Zhang Z, Li J, 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 X, Daygon VD, McNally KL, et al. Identification of stable QTLs causing chalk in rice grains in nine environments. Theor Appl Genet. 2016;129(1):141–53. https://doi.org/10.1007/s00122-015-2616-8.