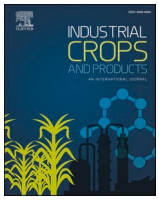
Industrial Crops & Products 215 (2024) 118663

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop



Unraveling the genetic basis of superior traits in *Gossypium barbadense*: From phenotype to genotype

Yongsheng Cai [a], Yanying Qu [a], Long Yang [a], Jun Liu [b], Peng Huo [c], Yajie Duan [a], Dongcai Guo [a], Qiang Zhou [a], Ying Li [a], Quanjia Chen [a,*], Kai Zheng [a,*]

[a] *Engineering Research Centre of Cotton of Ministry of Education, College of Agriculture, Xinjiang Agricultural University, Urumqi 830052, China*
[b] *Xinjiang Jinfengyuan Seeds Co., Ltd., Aksu 843100, China*
[c] *Department of Agriculture and Rural Affairs of Xinjiang, Urumqi 830000, China*

A R T I C L E I N F O

Keywords:
G. barbadense
Genetic map
QTLs
RNA-seq
Candidate gene
Genetic correlation

A B S T R A C T

G. barbadense is renowned for its high-quality fiber and is highly regarded as a natural material in the textile industry. However, research on cotton has mainly focused on *G. hirsutum*, and progress in sea island cotton has been relatively slow. There was limited understanding of the genetic loci and transcriptional regulatory mechanisms of important traits, and the genetic basis for the development of superior traits remains unclear. In this study, a recombinant inbred line (RIL) population was constructed using two sea island cotton parents from different cotton regions. The phenotypic characteristics of the two parents and the RIL population, as well as the phenotypic correlations, heritability, and genetic models of the RIL population, were comprehensively analyzed. The RIL population was genotyped using whole-genome resequencing technology, and a high-density intraspecific linkage map was constructed, consisting of 5295 bin markers and a total genetic map distance of 2721.79 centimorgans (cM). Phenotypic data collected from five different environments were used to detect 169 quantitative trait loci (QTLs) using the composite interval mapping method. Among these QTLs, 30 were related to agronomic traits, 61 were related to yield traits, and 78 were related to fiber quality traits. Additionally, 17 QTL clusters were detected, and the additive effects of these clusters explained the correlation between different phenotypic traits. Candidate genes for stable QTLs related to agronomic and yield traits were identified through variant annotation and functional prediction. Two bin markers associated with lint percentage (LP) were validated as potential breeding markers. Furthermore, using RNA-seq data of cotton fiber from the RIL population parents, the dynamic changes in gene expression during different developmental stages of cotton fiber were revealed, and important differentially expressed genes in the secondary cell wall development and metabolism network of fibers were identified. Importantly, through the combined analysis of fiber trait QTLs and transcriptomes, eight candidate genes involved in regulating fiber quality traits were predicted, and the genetic basis of the excellent allele site *qFL_D04_1* in the breeding history of Chinese sea island cotton was elucidated. In summary, this study provides new genetic resources and potential breeding markers for cotton variety improvement, valuable theoretical information for understanding the genetic basis of important traits in sea island cotton, and effectively promotes the development of cotton biotechnology breeding.

1. Introduction

Cotton is one of the world's important economic crops, accounting for 35% of the global total fiber production as a renewable resource (Wen et al., 2023). Cotton fiber products have excellent breathability and moisture absorption, and the unique comfort makes cotton textiles highly demanded and favored in society. In global cotton production, upland cotton, with its wide adaptability and high yield, accounts for over 95% of the total cotton production, while sea island cotton, with its superior fiber quality, is regarded as the "diamond in fibers," but its production only accounts for 2%-4% (National Cotton Council, http://www.cotton.org). Sea island cotton complements the traits of

* Correspondence to: Urumqi 830052, China.
E-mail addresses: cys0620@126.com (Y. Cai), xjyyq5322@126.com (Y. Qu), 18373009232@163.com (L. Yang), ljjaks@163.com (J. Liu), hp20080224@126.com (P. Huo), 547525782@qq.com (Y. Duan), 1907821666@qq.com (D. Guo), zhqiang715817@126.com (Q. Zhou), 1716225632@qq.com (Y. Li), chqjia@126.com (Q. Chen), zhengkai555@126.com (K. Zheng).

https://doi.org/10.1016/j.indcrop.2024.118663
Received 13 January 2024; Received in revised form 15 April 2024; Accepted 29 April 2024
Available online 9 May 2024

upland cotton (Hu et al., 2019), and breeders have adopted different breeding strategies. However, it has been proven that the offspring of traditional breeding techniques for intercrossing between sea island and upland cotton exhibit extreme segregation or even lethality, making it difficult to develop high-yielding, high-quality, and multi-resistant cotton varieties (Deng et al., 2019; Zhang et al., 2014). Under the modern cotton production model, sea island cotton production faces a series of bottlenecks, such as high homogeneity of varieties, inconsistent fiber quality, and low economic benefits, leading to a gradual reduction in planting area (Yuan et al., 2018; Zhang et al., 2023; Zheng et al., 2022), which severely affects the stability, competitiveness, and sustainable development of the sea island cotton industry chain.

The important economic traits of cotton are complex quantitative traits controlled by multiple genes, influenced by minor-effect polygenes and environmental factors (Li et al., 2023). The developmental process involves complex biological regulation of morphological and molecular dynamics (Wen et al., 2023), with extremely intricate genetic mechanisms. Furthermore, the characteristics of allopolyploid bring complexity and uncertainty to cotton improvement (Wendel and Grover, 2015). The negative correlations among important economic traits of cotton undoubtedly pose greater obstacles to simultaneous improvement. Therefore, outdated breeding theories and single breeding methods are insufficient to cultivate high-yielding, high-quality, and multi-resistant cotton varieties.

Genetic marker technology is an important tool in genetic research. It has evolved rapidly since Mendel's discovery of 7 pairs of morphological markers and Watson-Crick's revelation of the nature of genetic material (Watson and Crick, 1953). In 1994, Reinisch et al. first constructed a relatively complete restriction fragment length polymorphism (RFLP) genetic map for inter-specific sea island and upland cotton (Reinisch et al., 1994), marking the beginning of molecular genetic map research in cotton. In the early studies, most of the maps constructed were inter-specific maps considering the polymorphism between parents. However, inter-specific genetic maps resulting from sea island and upland cotton hybridization are not applicable to genetic improvement of either upland or sea island cotton (Wang et al., 2019). As a result, some researchers began to construct intra-specific genetic maps (Shappley et al., 1998), but the low polymorphism and limited number of markers resulted in low genome coverage. In 2012, the draft genome of *Gossypium raimondii* was published (Wang et al., 2012), marking a milestone in cotton biological research. Subsequently, the cost of sequencing greatly decreased, leading to an explosive growth of cotton genomic sequencing data (Li et al., 2015a, 2021; Wang et al., 2019), providing researchers with new perspectives and tools and promoting the development of population genetics. Single nucleotide polymorphism (SNP) markers, as third-generation molecular markers, have been widely used in the construction of cotton genetic maps. For example, Zhang et al. integrated four different types of markers to construct a high-density genetic map of a recombinant inbred line (RIL) population (0–153×sGK9708) with 8295 markers, a total genetic distance of 5197.17 cM, and an average distance of 0.88 cM (Zhang et al., 2020). Gu et al. used 6187 bin markers to construct a high-density genetic map of a RIL population (ND13×ND601) with a total length of 4478.98 cM, which is the first high-quality and high-density SNP genetic map based on resequencing (Gu et al., 2020). However, the intra-specific map of sea island cotton lags far behind that of upland cotton. In 2018, Fan et al. constructed the first high-density intra-specific genetic map of sea island cotton using genotyping-by-sequencing (GBS) sequencing technology. The map was constructed using a RIL population derived from the cross between sea island cotton 5917 and Pima S-7, and it contained 3557 SNP markers, a total length of 3076.23 cM, and an average distance of 1.09 cM (Fan et al., 2018).

The transition from traditional "empirical breeding" to efficient "precision breeding" can effectively improve the efficiency and predictability of breeders (Hickey et al., 2019). With the rapid development of disciplines such as genetics, bioinformatics, and molecular biology, the theory and technology of plant breeding have undergone significant changes (Song et al., 2023). The deep integration of molecular breeding information and molecular biology technology has to some extent broken through the barriers and limitations of current crop breeding (Chen et al., 2022). Currently, multiple functional databases integrating phenotypes and genomics have been established for cotton, such as the CottonMD database, CottonFGD database, and Cottongene database (Yang et al., 2023; Yu et al., 2014; Zhu et al., 2017), providing excellent gene resources and theoretical guidance for molecular breeding of cotton. However, they are still in the stage of basic research, and their contribution to molecular design breeding is relatively low (Wei et al., 2019). In particular, the accumulation of data on sea island cotton is limited, and research on the genetic localization and transcriptional regulation mechanisms of important traits is not in-depth. Faced with a shrinking planting area (FAO, https://www.fao.org/), sea island cotton is being rekindled by researchers. In the past three years, studies on natural populations of sea island cotton have been reported successively, and through genome-wide association study (GWAS) analysis, regulatory loci related to stress tolerance, disease resistance, and cotton fiber development have been identified (Jin et al., 2023; Su et al., 2020; Wang et al., 2022c; Yu et al., 2021; Zhao et al., 2022). These studies have actively promoted the genetic improvement of sea island cotton.

In this study, we used a newly bred Chinese elite parent of Xinhai cotton (Xinhai 21) as the female parent and an early variety of Pima cotton bred in the United States (06E2062) to construct a selfing population (105 families). Based on data from 12 different phenotypic traits in five environments, we explored the genetic characteristics of phenotypic traits. With the help of resequencing technology, we constructed a high-density genetic map, mapped the quantitative trait loci (QTL) of eight important traits of sea island cotton, and deciphered the genetic basis of the target traits. Finally, using transcriptome data of parental cotton fibers at different stages and published public databases as validation sets, we screened candidate genes within stable QTLs. Our research results provide theoretical support for unraveling the genetic basis of excellent traits in sea island cotton and have the potential to create breakthrough cotton new varieties using cotton biotechnology breeding techniques.

2. Materials and methods

2.1. Construction of RIL population and phenotypic investigation

The experimental materials for this study were selected as the Chinese sea island cotton variety, Xinhai 21, which has the largest cumulative planting area, as the female parent, and the high-yielding and high-lint Pima cotton variety, 06E2062, from the United States, as the male parent. In 2012, F_1 seeds were produced by crossing the male and female parents in Alar City, Xinjiang. In the same year, the F_1 generation was further propagated in Sanya, Hainan. From 2013–2019, F_2 individual plants were grown in Alar City, Xinjiang, and self-pollinated and single-seed descent was practiced until the seventh generation, resulting in a segregating population of 105 recombinant inbred lines. The parents and RIL population were planted in a three-point two-replicate trial from 2019 to 2021 in Alar City, Xinjiang (Table S1). The phenotypic evaluation of the population was conducted using a completely randomized block design, with one film and three rows of artificial sowing. The row length was 2 m, and the spacing between rows was 0.5 m. Field management was carried out according to local production practices.

Phenotypic data for plant height (PH), first fruit spur branch number (FFSBN), first fruit spur height (FFSH), effective boll number (EBN), and fruit spur branch number (FSBN) were collected during the flowering stage for the RIL population and parental lines in each survey year. In addition, 50 normal cotton bolls from the middle fruit spur of each sample were selected, weighed, and ginned to determine lint percentage (LP) and single boll weight (SBW) data. For fiber quality evaluation, 20 g of seed cotton from each sample was sent to the Cotton Quality

Inspection and Testing Center of Xinjiang Academy of Agricultural Sciences. Fiber quality parameters including fiber length (FL), fiber strength (FS), micronaire value (FM), fiber uniformity (FU), and fiber elongation (FE) were measured using the Uster HVI1000 fiber testing instrument under controlled conditions of constant temperature (20 ±1℃) and relative humidity (65±1%). Meteorological data for the study site were obtained from the National Meteorological Information Center of China (http://data.cma.cn/).

2.2. Phenotypic data analysis

Descriptive statistics and correlation analysis of the phenotypic data for the parental lines and RIL population were performed using IBM SPSS Statistics 27.0 (SPSS, Chicago, IL). The lme4 package (Bates et al., 2014) in R 4.2.1 was used to calculate the best linear unbiased estimation (BLUE) values for each RIL in five environments, and the Performance Analytics package (Peterson et al., 2014) was employed to analyze the correlation of 12 traits. Variance and heritability analysis were conducted using QTL IciMapping 4.2, utilizing the following equation: $H_b^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{G\times E}^2/n_e + \sigma_E^2/n_e n_r)$, where ne and nr represent the number of environments and replicates, respectively. To present the data analysis results in a clear manner, the ggplot2 package (Wickham, 2009) in R 4.2.1 was used for graphical visualization, including *t-test* boxplots for parental phenotypic traits, a correlation heatmap for population phenotypic traits, as well as normal distribution plots and boxplots for population phenotypic traits.

2.3. Analysis of quantitative trait main genes and multiple gene mixed inheritance

To identify the main genes and multiple gene mixed inheritance models for quantitative traits in a segregating population, and to provide reference information for the genetic basis of quantitative traits and crop breeding, the SEA v2.0 software in R 4.2.1 was used (Wang et al., 2022a). The interactive program SEA-G4F3 was employed to analyze different phenotype data. In this analysis, models with the minimum or smaller Akaike's information criterion (AIC) values were selected as candidate models. Then, fitness tests, including Homogeneity (U_1^2, U_2^2, U_3^2) test, Smirnov (${}_nW^2$) test, and Kolmogorov (D_n) test, were conducted. Models that did not reach a significant level in the fitness tests were selected from the candidate models. By comparing the genetic parameters of each model, the best genetic model was determined. Genetic parameter analysis was then performed on the best model.

2.4. DNA extraction and sequencing

Field-collected parental and RIL population leaf tissue samples were collected and immediately frozen in liquid nitrogen before being stored at −80°C. Genomic DNA was extracted using an improved cetyltrimethylammonium bromide (CTAB) method. The concentration of DNA samples was determined using a NanoDrop™ One microvolume UV-Vis spectrophotometer by Thermo Fisher Scientific. The integrity and purity of DNA samples were assessed by 1% agarose gel electrophoresis. Genomic DNA was then fragmented by ultrasonication and size-selected using magnetic beads to construct the genomic resequencing libraries. The libraries were subjected to PE150 sequencing on the DNBseq platform by BGI Genomics (Rao et al., 2020). The raw data obtained were processed to obtain high-quality clean data by splitting, removing contaminants, and trimming low-quality sequences.

2.5. SNP detection

Used the BWA software (Li and Durbin, 2009) to align the clean data to the reference genome of *G. barbadense* 3–79 (Wang et al., 2019) with the parameters "mem -t 14 -k 32 -M -R". Converted the SAM file to BAM format using Samtools. Sorted, marked duplicates, filter, and indexed the BAM file using Picard-tools (https://broadinstitute.github.io/picard/). Used the GATK software (McKenna et al., 2010) to calculate alignment rate, coverage, and depth based on the filtered BWA alignment results. Selected uniquely mapped reads and perform population SNP detection using the HaplotypeCaller program from the GATK software package (QD < 2.0 || MQ < 40.0 || FS > 60.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0). Additionally, to improve variant detection accuracy, performed quality recalibration and data output using the BaseRecalibrator and PrintReads tools, followed by variant detection using the HaplotypeCaller program again to generate the final usable VCF data.

2.6. SNP filtering

To eliminate false-positive variants caused by misalignment, we only retained high-quality SNPs (parental coverage depth ≥ 15, RIL coverage depth ≥ 5). Based on parental polymorphism, SNPs were classified into 4 types of allelic segregation (hk×hk, aa×bb, nn×np, and lm×ll). To ensure the accuracy of subsequent experiments, we selected SNP markers that exhibited polymorphism between parents and were homozygous, specifically the aa×bb type. We further filtered the high-quality SNPs using the following criteria: (1) removing sites with a missing rate greater than 15% in the RIL population; (2) filtering out sites with a chi-square test P-value less than 0.001 for skewed segregation. These retained sites were used for subsequent analysis.

2.7. Construction and quality assessment of genetic map

The SNP markers that conform to the aa×bb segregation pattern in the RIL population were processed using the "sliding window" method, as described by HUANG et al. (Huang et al., 2009), with a perl script. In this method, a set of SNPs is treated as a single marker unit, and a sliding window approach is used. A window size of 15 SNPs, with a step size of 1 SNP, was employed to detect recombination breakpoints in the RIL population and generate a recombination breakpoint map. Additionally, bin markers with positions less than 500 bp or containing fewer than 4 SNPs were filtered out. The genotypes of the offspring were classified as "A," "B," or "H" based on the geneotyping results of the RIL population. The output file obtained from this process was imported into JoinMap 4.1 for genetic map construction using regression analysis and the Kosambi function (Stam, 1993). Markers that exhibited severe non-linkage were removed, and the genetic distance between adjacent markers was calculated. Finally, a linkage genetic map was generated.

Performing co-linearity analysis between the markers' physical positions and genetic positions shown in the diagram, the ALLMAPS program was used to calculate the spearman correlation coefficient for each linkage group with the physical map (Tang et al., 2015). This analysis aimed to validate the construction of the genetic map. The ratio of genetic distance to physical distance between adjacent markers in the diagram was used to determine the recombination rate. Regions with a recombination rate greater than 20 cM/Mb were considered as recombination hotspots.

2.8. QTL localization of important traits in Sea Island cotton

QTL mapping was performed using the QTL IciMapping 4.2 software with the following parameters: step size of 1 cM, PIN of 0.001, permutation times of 1000, and LOD threshold of 2.5 (Meng et al., 2015). Inclusive Composite Interval Mapping ADDitive (ICIM-ADD) method was used to map QTLs for eight traits including PH, LP, FL, and FS. Overlapping confidence intervals for the same trait in different environments were considered as the same QTL. QTLs detected in two or more environments were defined as stable QTLs. QTLs identified in this study but not overlapping with previously reported QTLs were considered as novel. The 99% confidence interval for QTLs was used as the mapping interval. Physical confidence intervals were determined based on the

markers within the genetic confidence interval, using the reference *G. barbadense* 3–79 genome. QTLs were named following the convention of "q+trait+chromosome+QTL number" established by previous studies (McCouch, 1997). If the confidence intervals of two or more QTLs for different traits overlapped or if the highest peak of the QTLs was within 10 cM, they were considered as a QTL cluster (Zhang et al., 2020). MapChart software was used to visually present the QTL mapping results (Voorrips, 2002).

2.9. Annotation and identification of candidate genes

Stable QTLs detected under different environmental conditions, which show minimal environmental influence, are more likely to be candidate genes or markers for trait improvement and prediction. We selected the 99% confidence interval of stable QTLs as the candidate region and extracted the physical interval and gene information within this QTL from the CottonFGD website (Zhu et al., 2017). ANNOVAR software package (Wang et al., 2010a) was used to annotate the genetic variations of the population SNP and Indel data based on the *G. barbadense* 3–79 genome, and the annotation results were analyzed and summarized. The candidate genes were functionally annotated and enriched using the eggNOG-mapper online tool for Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and euKaryotic Orthologous Groups (KOG) analyses (Cantalapiedra et al., 2021). The selected candidate genes and their corresponding protein sequences were compared to the *A. thaliana* using the BLASTP software for functional prediction, utilizing the Gene search program in the CottonMD website (Yang et al., 2023).

2.10. Transcriptome analysis of cotton fibers at different developmental stages

For this study, the RIL population derived from the parents (Xinhai 21 and 06E2062) was used. Embryo sacs at 0 DPA (day post anthesis) and fibers at 5 DPA, 10 DPA, 15 DPA, 20 DPA, 25 DPA, 30 DPA, and 35 DPA were collected and immediately frozen in liquid nitrogen. The samples were then stored at −80°C for subsequent analysis. Total RNA extraction from the different cotton fiber samples was performed using the RNAprep Pure Polyphenol Plant Total RNA Extraction Kit (Tiangen Biotech). The protocol provided by the manufacturer was followed to extract total RNA from the samples at different developmental stages. The extracted RNA was then used for mRNA enrichment using magnetic beads with Oligo (dT) and cDNA library construction with insert fragments of 150–200 bp. The DNBSEQ sequencing platform was used for paired-end sequencing. The *G. barbadense* 3–79 reference genome was used for sequence alignment using HISAT2 (Kim et al., 2019). The aligned reads were then processed using the featureCounts command from the Subread package (Liao et al., 2013) to calculate the expression levels. The expression values used in this study were represented as TMM (trimmed mean of M-values) values. Differentially expressed genes (DEGs) were identified using the DEGseq package (Wang et al., 2010b), with a threshold set at $|\log_2(\text{FoldChange})| > 1$ and a P-value < 0.01. The WGCNA package (Langfelder and Horvath, 2008) in R 4.2.1 was used to construct a weighted gene co-expression network. The "blockwise Modules" function with default settings was used to obtain gene expression modules. The LinkET package (https://github.com/Hy4m/linkET) was employed to perform Mantel tests and calculate the correlation coefficients between the module eigengenes (MEs) and different fiber traits. Specific expression modules were selected and subjected to GO enrichment analysis and KEGG pathway analysis using the cluster Profiler package (Yu et al., 2012) in R 4.2.1.

2.11. Validation by qRT-PCR

To validate the accuracy of the RNA-seq data, ten differentially expressed genes were randomly selected for qRT-PCR analysis. Primer design was conducted using Primer Premier 5 software (Lalitha, 2000) (Table S2). Total RNA extraction was performed as described in the previous section, and single-stranded cDNA was synthesized using the First strand cDNA Synthesis kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The cotton *GbUBQ7* gene was chosen as the reference gene. Biological triplicates were set up, and data analysis was conducted using the $2^{-\Delta\Delta Ct}$ method (Hu et al., 2019).

2.12. Analysis of genotypic and phenotypic data in natural populations

In the initial phase of this study, a wide range of sea island cotton materials from different countries and regions were collected by our research group. Additionally, publicly available genotype data materials were obtained from the NCBI SRA database (Ma et al., 2021; Wang et al., 2022c; Yu et al., 2021; Zhao et al., 2022). From these sources, a total of 287 sea island cotton resources (Table S3, with an average sequencing depth of $\geq$ 10x) were selected to construct a natural population of sea island cotton. Among the 287 resources, 63 were approved varieties from Xinjiang, 54 were collected from the northwest region of China, 32 were collected from coastal areas of China, 92 were obtained from Central Asian countries, 26 were from the United States or South America, and 20 were from Egypt. In 2021, these materials were planted in the same environmental conditions as the RIL population in Korla and Awati, Xinjiang. The planting methods, collection of phenotypic data, SNP filtering, and descriptive statistics were performed following the same procedures as the RIL population.

3. Results

3.1. Phenotypic analysis of RIL population parents

Selection of parents with significantly different phenotypes in the parental segregation population can maximize the utilization of their genetic diversity. The cultivated variety Xinhai 21, developed in China, is a zero-fruiting spur type, while the Pima variety 06E2062, originated from the United States, is a type II-fruiting spur (Fig. 1A). Both have been widely used as breeding parents in their respective countries of origin and have made significant contributions. The average PH of Xinhai 21 is 62.57 cm, while that of 06E2062 is 55.34 cm, and the difference between them is not significant (Fig. 1B). Other traits, such as EBN, FU, and FE, show no significant differences between the parents (Fig. 1E, L, M). However, significant differences are observed between the parents in terms of FSBN, SBW, FS, and FM (Fig. 1F, G, J, K). FFSBN, FFSH, LP, and FL also show extremely significant differences between the parents (Fig. 1C, D, H, I). It is worth noting that FFSH and LP are bottlenecks for the sustainable development of sea island cotton production in China and are traits that urgently need improvement. The FFSH of 06E2062 is 17.08 cm, with a LP of 40.28%, while Xinhai 21 has a FFSH of 5.79 cm and a LP of 33.71%. The FFSH of 06E2062 exceeds that of Xinhai 21 by 11.29 cm, and its LP is 6.57% higher in comparison. However, the FL of 06E2062 is 31.18 mm, which is 19.00% shorter than that of Xinhai 21 (36.98 mm). In summary, Xinhai 21 and 06E2062 parents have different origins and do not belong to the same ecological cotton region. They exhibit significant differences in agronomic, yield, and quality traits, providing a basis for generating more genetic variability in their progeny.

3.2. Phenotypic analysis of the RIL population

Phenotypic analysis of the RIL population was conducted in five different environments, resulting in the collection of data for 12 distinct phenotypic traits. The BLUE values for the means were calculated using the lme4 statistical package. A summary of descriptive statistics, including minimum and maximum values, means, variances, coefficients of variation, skewness, and kurtosis for the 12 traits, was provided (Table S4). The range of variation between minimum and

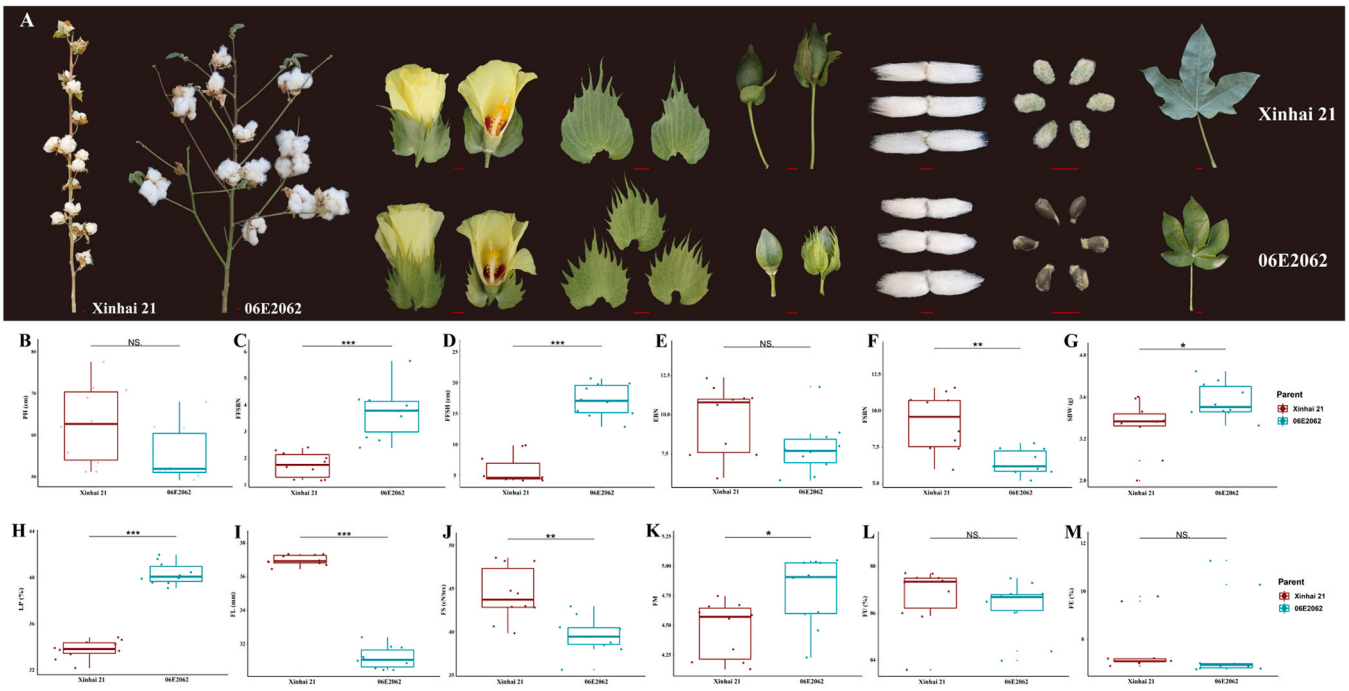


Fig. 1. Phenotypic characterization and comparative analysis of 12 traits between Xinhai 21 and 06E2062 lines. (A) The field-level phenotypic traits exhibited by *G. barbadense* Xinhai 21 and 06E2062. The scale bars in the panels are 1 cm, respectively. (B-M) A boxplot was utilized to illustrate intergroup comparisons between the lines Xinhai 21 and 06E2062 across 12 traits. *, **, *** indicates significant levels at $P < 0.05$, 0.01, 0.001, respectively, determined by Student's *t* test.

maximum values across different environments was 8.93–10.44 mm for FL, 10.06–11.78% for LP, and 8.60–15.9 cm for FFSH. The coefficients of variation for the different traits ranged 1.36%-48.72%, with FU having the lowest variation and FFSH having the highest variation. The absolute values of skewness for the different traits across environments were all less than 1, with most fiber quality traits exhibiting negative skewness and most agronomic traits and yield traits showing positive skewness, except for SBW. In terms of kurtosis analysis, the absolute values for most traits across environments were less than 1, except for 6 traits (FE, EBN, FFSH, and FFSBN), which had absolute kurtosis values greater than 1. A value of kurtosis greater than 1 indicates that the phenotypic data for that trait are more concentrated and there are a few individuals exhibiting extremely extreme phenotypes, showing significant positive or negative transgressive segregation. Furthermore, the

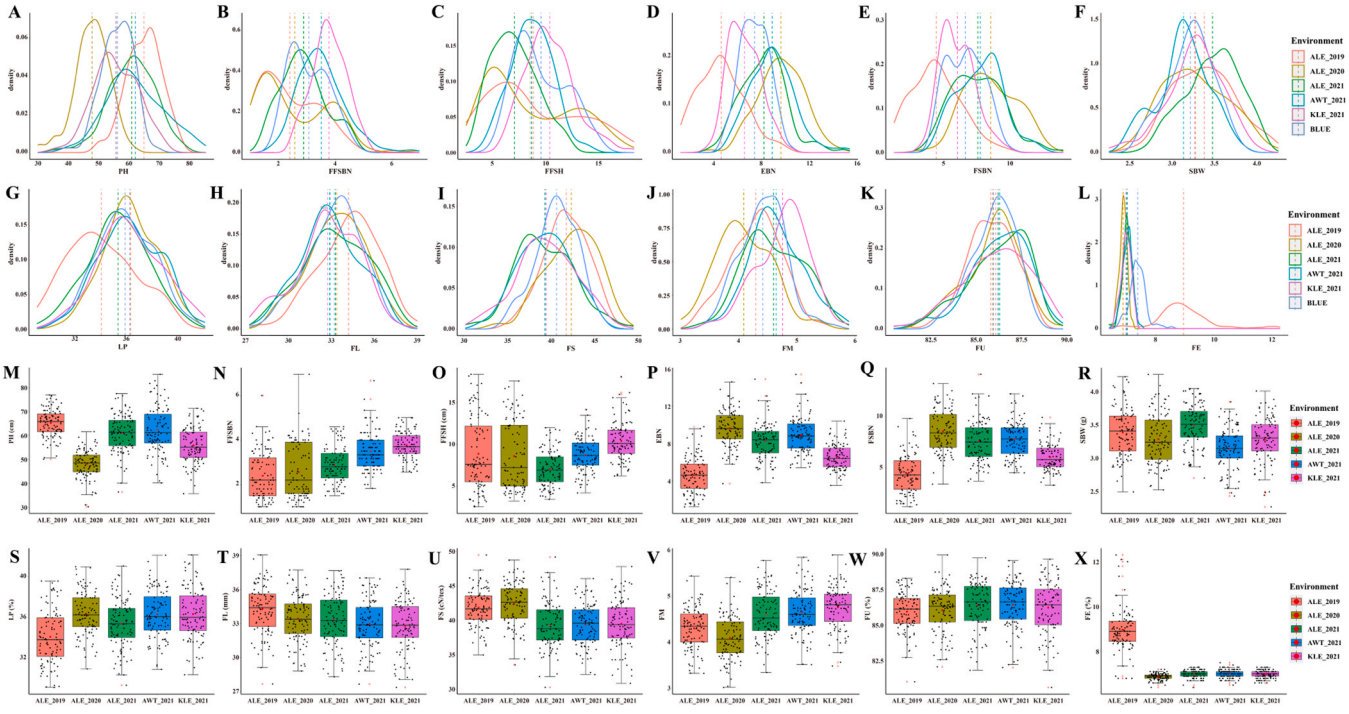


Fig. 2. Normal distribution and boxplot of 12 traits in RIL populations across different environments. (A-L) Normal distributions of different traits in the RIL population across five environments. (M-X) Boxplots of different traits in the RIL population across five environments.

Kolmogorov-Smirnov test was performed to examine whether the different traits across environments followed a normal distribution, and frequency histograms were used to clearly illustrate the normal distribution curves (Fig. 2 A-L). The results showed that FFSH and FFSBN exhibited bimodal distributions, while FE did not follow a normal distribution. The remaining traits all followed or closely approximated a normal distribution, which is consistent with the descriptive statistical results. In conclusion, the RIL population exhibits significant genetic diversity and wide phenotypic variation, and except for FE, all other traits are suitable for genetic analysis.

3.3. Correlation analysis of the RIL population phenotypes

To explore the reliability of the data, we analyzed the correlation of the same phenotypic traits in different environments. The results showed that most of the same traits exhibited strong positive correlations (correlation coefficient > 0.5 and extremely significant) in different environments. However, the correlation coefficient of FE between different environments was consistently lower than 0.5, which was significantly different from other traits (Table S5, Fig. 3A). In the correlation analysis of different traits in the same environment, there were no significant strong correlations between fiber traits and agronomic traits, as well as fiber traits and yield traits. However, there were extremely significant positive correlations between FFSBN and FFSH, EBN and FSBN, and FL and FS in different environments (Table S6). Furthermore, correlation analysis using the BLUE values revealed that there were 30 combinations with significant or extremely significant correlations, including 11 combinations with strong correlations, 10 with moderate correlations, and 9 with weak correlations (Fig. 3B). In addition, among the fiber quality traits, the FM showed extremely significant negative correlations with other fiber quality traits (except FE). These results indicate that there is weak correlation between different types of traits in the RIL population, while there is strong correlation between traits of the same type. The phenotypic traits of different types interact, constrain, and influence each other.

The phenotype of cotton is the result of the interaction between genotype and environment. Due to chemical regulation factors, the ALE-2020 environment exhibited significantly PH compared to other environments, but significantly higher EBN (Fig. 2M). In the ALE-2019 and KEL-2021 environments, there were lower boll numbers, as indicated by meteorological data (Fig. 3C, D). The high temperature during bolling period in the ALE-2019 environment and the low temperature during seedling stage in the KEL-2021 environment were unfavorable for cotton yield formation. The former resulted in a higher shedding rate, while the latter hindered the growth and development of cotton seedlings (Fig. 2M-Q). In different years at the same experimental site, we found that the average FL decreased year by year (Fig. 2T). In different experimental sites in the same year, the ALE environment was more conducive to the formation of SBW and FL, but not to the improvement of LP (Fig. 2S, T). In addition, the KEL environment had significantly

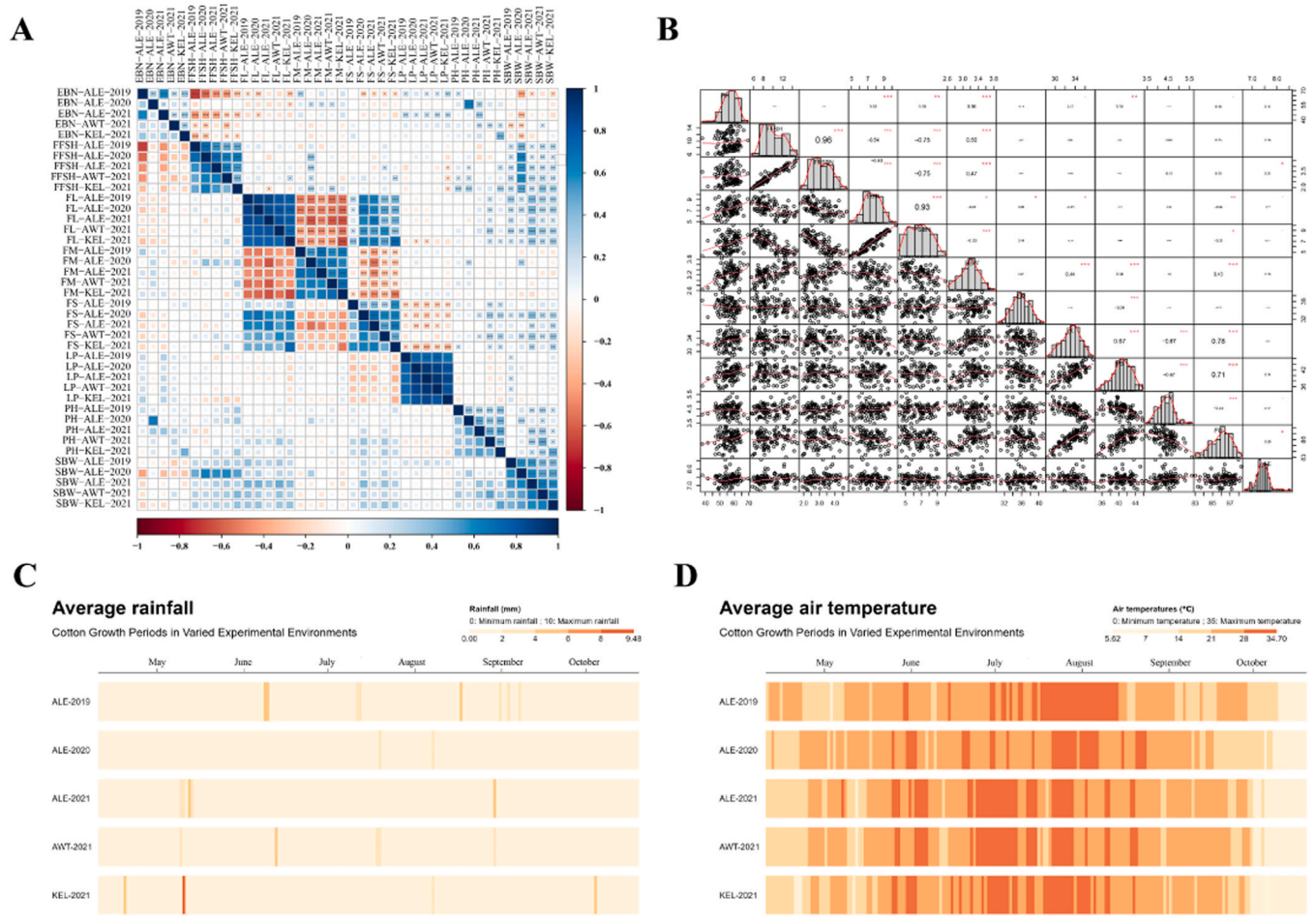


Fig. 3. Analysis of correlations and environmental factors within the RIL population during the cotton growth season. (A, B) Heatmap of the correlation between different environments across identical traits; Frequency distribution and correlation coefficients of phenotypic variations of 12 traits. Traits on the diagonal correspond to names on the horizontal and vertical coordinates. *, **, *** indicates significant levels at $P < 0.05$, 0.01, 0.001, respectively. (C) Heatmaps of average rainfall in the cotton growth season in the 5 environments in 2019–2021. (D) Heatmaps of average temperature in the cotton growth season in the 5 environments in 2019–2021.

lower PH, EBN, and FSBN compared to other environments, but there was no significant difference in fiber quality traits and LP compared to other environments (Fig. 2M-X).These results indicate that extreme temperatures significantly affect agronomic and yield traits, but have a smaller impact on fiber quality traits, suggesting that the former is more influenced by environmental effects, while the latter is more controlled by genotype and has a higher heritability.

3.4. Analysis of heritability and genetic models in RIL population

In the analysis of variance model, genotype effect, environment effect, genotype-environment interaction effect, replication-environment interaction, and error effect were all included in the genetic model. The F-test indicates that the genetic model was suitable for analyzing 12 phenotypic traits (Table S7). Further analysis of the genetic model reveals that there were significant differences among genotypes for traits such as FFSH, SBW, LP, FL, FM, and FU. The environment effect, genotype effect, and genotype-environment interaction effect were also significant. However, based on the estimation of variances, the genetic variance was much higher than the random error variance, environmental variance, and interaction variance, indicating that the phenotypic data were suitable for further genetic analysis. On the other hand, PH, FFSBN, EBN, FSBN, and FS, although reached a significant level for genotype, had genetic variance estimates lower than environmental variance or interaction variance. This suggested that these traits were more influenced by the environment and required more environmental data for genetic analysis. FE was not suitable for genetic analysis, which was confirmed by the estimation of broad-sense heritability. Except for FE, the broad-sense heritability ranged from 72.05% to 95.36% for different traits, with EBN having the lowest broad-sense heritability and LP having the highest.

The previous analysis results indicated that the FFSBN did not follow a normal distribution, the FSBN was highly correlated with the EBN, the FU had low variability, and the FE had low heritability. Therefore, we excluded these four phenotypic traits from the subsequent genetic analysis. In the analysis of the major gene + polygene mixed genetic model; we retained the three genetic models with the smallest AIC values for each trait in different environments. A total of 14 genetic models were obtained for the eight phenotypic traits in different environments, and we performed goodness-of-fit tests including U_1^2, U_2^2, U_3^2, ${}_nW^2$ and D_n (Table S8). It was found that the most suitable genetic model for the same trait differs in different environments. In order to more accurately infer the genetic models corresponding to the phenotypic traits, we not only considered the AIC values but also calculated the frequencies of different models appearing in different environments. Therefore, it was predicted that PH, LP, FS, and FM conformed to the 1 major gene + polygene additive-dominance-epistasis (MX1-AD-ADI) genetic model; the FFSH conformed to the 2 major gene additive-dominance-epistasis (2MG-ADI) genetic model; and FL, EBN, and SBW conformed to the polygene additive-dominance-epistasis (PG-ADI) genetic model. By adjusting the QTL mapping statistical model based on the genetic models provided, the influence of confounding effects can be eliminated or reduced, leading to more accurate detection signals.

3.5. Statistical analysis and SNP filtering of resequencing data

Resequencing of RIL population and parents was performed using the DNBSEQ sequencing platform from BGI Genomics. Approximately 1.22 Tb of high-quality data was generated, with a Q20 $\geq$ 96.71% and Q30 $\geq$ 94.71% (Table S9). The GC content was within the normal range. The maternal parent, 'Xinhai 21', and the paternal parent, '06E2062', yielded 83.2 Gb (sequencing depth of 48.48×) and 72.3 Gb (sequencing depth of 40.45×) of clean data, respectively. After aligning the parents to the reference genome, the coverage at 4.0× was determined to be 96.7% for the maternal parent and 97.1% for the paternal parent. The RIL population obtained an average of 10.1 Gb of clean data, with an average sequencing depth of 5.95×. The 1.0× coverage was $\geq$ 98.56%, and the average coverage at 4× was 75.7%. The high-depth sequencing of the parents provided more comprehensive and accurate variant data, which serves as the fundamental support for the reliability of the research results. After SNP calling, a VCF file was generated for each parent, resulting in a total of 1925,537 polymorphic SNP markers. Among these, the aa×bb genotype accounted for 48.31%, equivalent to 930,171 SNPs (Table S10). Subsequently, the filtered SNPs from the parents were extracted from the RIL population VCF file. Further filtering was performed within the population VCF data, resulting in a final set of 901,893 SNPs that were pure and polymorphic in the parents and exhibited no abnormal genotypes in the RIL population. These SNPs will be used for subsequent analysis. SNP density analysis revealed an average density of 0.45/kb and 0.38/kb for the A and D subgenomes, respectively. Although there were still gaps of varying sizes on chromosomes D10 and D11, the overall SNP quantity provided uniform coverage across the entire genome (Fig. 4A), making them suitable for further analysis.

3.6. Construction of genetic map

Bin markers, also known as segment markers, represent a segment of the genome rather than a single nucleotide change, effectively reducing the impact of single point variations. Firstly, the filtered SNP loci were used to analyze population structure and construct an evolutionary tree for the RIL population. Secondly, the heterozygosity rate of each family was calculated and filtered, revealing no abnormal branches or families with high heterozygosity rates (Fig. 4B). Then, a perl script was used to analyze the recombination breakpoints of the families using a sliding window approach, detecting a total of 72,545 recombination breakpoints (Fig. 4C). The offspring ranged from a minimum of 52 breakpoints to a maximum of 340 breakpoints, with an average of 98 breakpoints per offspring. Ultimately, 7789 bin markers were obtained for subsequent genetic map construction. The bin markers were input into the Joinmap 4.1 software to construct the genetic map, resulting in 5295 bin markers being mapped into 26 linkage groups (Table S11). The total genetic map distance was 2721.79 cM, with an average genetic map distance of 0.55 cM and a maximum gap of 15.11 cM. The percentage of gaps smaller than 5 cM averaged 99.33% (Table S11). The longest linkage group was on chromosome D08, with a genetic distance of 139.79 cM and containing 187 markers. The shortest linkage group was on chromosome D13, with a genetic distance of 77.10 cM and containing 145 bin markers. Chromosomes A06 and A11 had the most bin markers (349), while chromosome A04 had the fewest (116). The average recombination rate of the genetic map was 1.5, with each linkage group having 0–6 recombination hotspots, totaling 55 hotspots. Among them, five chromosomes had no recombination hotspots.

3.7. Quality assessment of the genetic map

Quality assessment of the genetic map (Fig. 4D) was conducted through analysis of monomer origin, collinearity comparison, and linkage analysis between adjacent markers. It was found that the majority of chromosome segments in the RIL population showed consistent origins. Most bin markers in each linkage group were ordered in accordance with the physical map order, and the strength of the linkage relationship weakened with increasing marker distance. Although there may be chromosome translocations in chromosomes A05, A06, A09, A11, and D03 compared to the reference genome, resulting in poorer collinearity, the correlation coefficient of chromosome A11 was the lowest, at only 0.214. However, 19 linkage groups had Spearman correlation coefficients greater than 0.5, and 15 linkage groups had correlation coefficients greater than 0.9. Among them, chromosome D10 had the highest correlation coefficient, at 1. Overall, these findings indicate that the constructed genetic linkage map exhibits good collinearity and high accuracy in calculating genetic recombination rates, making it

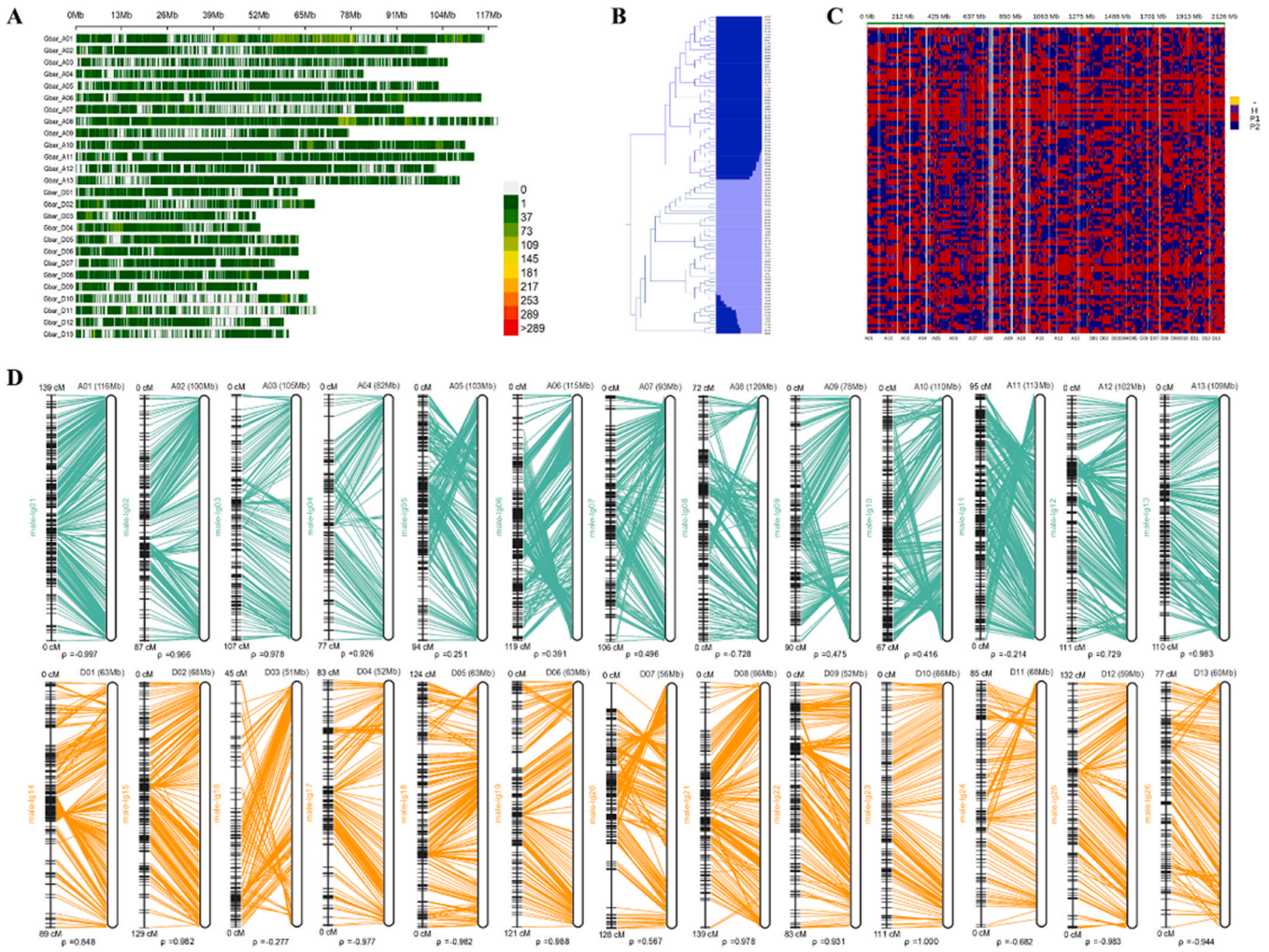


Fig. 4. High-density genetic map construction and collinearity analysis between the genetic map and the physical map graph of the RIL population. (A) Heatmap of SNP density in the RIL population. (B) Phylogenetic neighbor-joining tree and structure of the RIL population. (C) Recombination bin map generated from the RIL population. (D) Comparison of the genetic map and the *G. barbadense* 3–79 genome sequence.

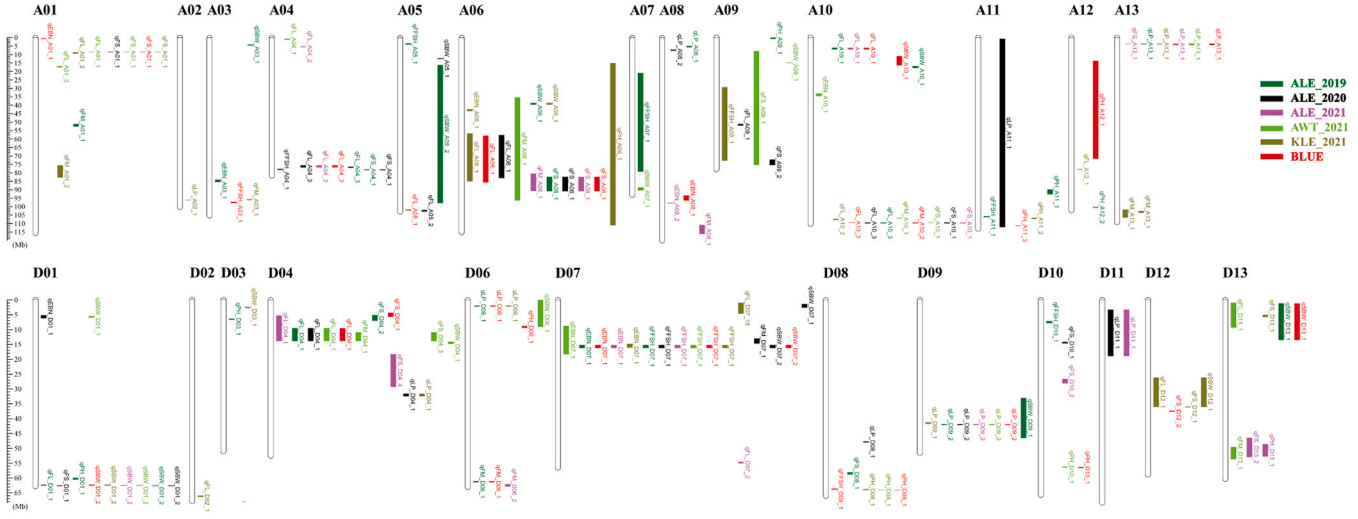


Fig. 5. Mapping of quantitative trait loci (QTL) for eight different traits on the physical map. Note: Length of segments represented physical intervals of corresponding QTLs. Different environments are represented by different colors.

suitable for subsequent mapping analyses.

3.8. Identification and analysis of important trait Loci in Sea Island Cotton

In order to identify the genetic loci related to cotton agronomic traits, yield, and fiber quality traits, a QTL mapping analysis was conducted using the ICIM (Interval Mapping) method in the QTL IciMapping software for eight phenotypic traits from 2019 to 2021 across five environments and one BLUE value (Table S12, Fig. 5). A total of 169 QTLs were detected, with a phenotypic variation explained ranging from 3.25% to 69.83% and LOD values ranging from 2.5 to 31.47. Among them, 78 QTLs were associated with fiber quality traits (34 QTLs for FL, 16 QTLs for FM, and 28 QTLs for FS), 61 QTLs were associated with yield or yield-related traits (12 QTLs for EBN, 23 QTLs for LP, and 26 QTLs for SBW), and 30 QTLs were associated with agronomic traits (14 QTLs for FFSH and 16 QTLs for PH). Except for chromosome D05, QTLs were distributed on the other 25 chromosomes, with chromosome D07 having the highest number of QTLs. Analysis of the QTL distribution on chromosomes revealed that the QTLs for fiber quality and yield-related traits were not randomly distributed in the At and Dt subgenomes. Among the QTLs associated with fiber quality, 49 were located in the At subgenome, while 29 were distributed in the Dt subgenome. In the identified QTLs, the additive effects ranged from −2.95–5.16, and both positive and negative additive effects were detected for the same trait. For example, the additive effects of *qLP_A08_1*, *qLP_D04_1*, *qLP_D09_1*, and *qLP_D09_2* were negative, indicating that the alleles contributing to higher LP were derived from the high-lint parent 06E2062. Although the LP of the female parent Xinhai 21 was lower than that of the male parent 06E2062, it still carried alleles for higher LP at the loci *qLP_A02_1*, *qLP_A08_2*, *qLP_A13_1*, and *qLP_D06_1*. The alleles for positive lint percentage QTLs were dispersed in the parents, and the QTL mapping results provided a reasonable explanation for the phenomenon of transgressive segregation in LP. Based on the phenotypic traits of the parents and the positive or negative values of the additive effects, it was found that 63 QTLs had alleles for increased effects derived from 'Xinhai 21', while 106 QTLs had alleles for increased effects derived from '06E2062' (Table S12). Additionally, 15 QTLs associated with fiber quality, 13 QTLs associated with yield, and 3 QTLs associated with agronomic traits were detected in at least two or more environments. Among them, *qFL-A04–3*, *qFS-A01–1*, and *qFS-A06–1* were detected in four environments, *qFL_D04_1* and *qEBN_D07_1* were detected in five environments, and *qLP-D09–1*, *qSBW-D01–2*, and *qFFSH_D07_1* were detected in all environments, indicating their stability. In the QTL mapping statistics table, each QTL was assigned a unique identifier, which also provided detailed information about the QTL location, including trait, environment, marker name, genetic position, LOD value, phenotypic variation explained (PVE), physical position, and the number of genes within the interval (Table S12). For example, *qLP-D09–1* was located at 51 cM on chromosome D09, with the left bin marker Gbar_D09_418 and the right bin marker Gbar_D09_421. It explained a PVE ranging from 11.47% to 22.50% and had LOD values ranging 4.91–11.63. The additive effects were −1.03 to −0.71, and the physical position was 41.85–42.11 Mb, with a fragment size of 0.26 Mb and a total of nine genes within the interval.

3.9. Identification of QTL clusters

The existence of QTL clusters implies the presence of multiple QTLs that are densely distributed on the chromosomes, and these loci may play a synergistic regulatory role in complex phenotypes. In the identification of QTL clusters, intervals with single QTLs larger than 20 Mb are not analyzed. A total of 17 QTL clusters were identified (Table S13), with 9 clusters in the At subgenome and 8 clusters in the Dt subgenome. Among them, there are 2 QTL clusters on chromosomes A06, D01, and D13, respectively. At least one stable QTL was identified in 11 QTL clusters, indicating that these regions may contain candidate genes related to important economic traits in cotton. Among the stable QTL clusters, *qClu-D07–1* cluster contains the most QTLs (4), while *qClu-A01–1*, *qClu-A06–2*, *qClu-A10–2*, and *qClu-D04–1* only contain QTLs related to fiber quality. *qClu-D01–2*, *qClu-D08–1*, *qClu-D12–1*, and *qClu-D13–2* are QTL clusters with mixed traits. When the additive effects within QTL clusters have the same direction, there may not be a problem of linkage drag, indicating the potential for marker-assisted selection breeding. Unfortunately, LP always has a different additive effect direction compared to other traits, and we did not identify potential QTL clusters related to LP. However, it is noteworthy that the results of QTL clusters largely explain the correlation between different phenotypic traits in the RIL population, which is crucial for understanding the genetic mechanisms and relationships between complex traits. It may serve as the basis for the genetic linkage between different traits.

3.10. Annotation of variation in RIL population

The variants in the RIL population, which are both homozygous and polymorphic, were annotated using the ANNOVAR software package. The results revealed a significant number of genomic structural variations occurring in intergenic regions, followed by upstream and downstream regions of genes. Specifically, the 5′ UTR and 3′ UTR regions were annotated 6616 and 4646 times, respectively (Table S14). Splicing regions were annotated 137 times, with 7 occurrences in exon splicing regions, 42,115 occurrences in intron regions, and 14,412 occurrences in exon regions. Further analysis of exon variants identified 490 frameshift insertion/deletion sites, 322 non-frameshift insertion/deletion sites, 270 stop codon loss/mutations, 4793 synonymous mutations, and 8518 non-synonymous mutations, involving a total of 6693 genes (Table S14).

3.11. Identification and analysis of candidate genes in stable QTL for agronomic traits

Based on the mapping results of agronomic traits, we selected two stable QTLs associated with PH and two stable QTLs associated with FFSH. By aligning the markers at both ends of the QTL intervals with the physical genome, we identified 81 genes within these intervals. After screening, we identified 16 genes carrying polymorphic variants (Table S15). For FFSH, we detected a QTL with a phenotypic variance ranging from 31.34% to 69.83%. Among the genes within this QTL, only one, *Gbar_D07G011870*, encoding the TFL1 protein, contained a polymorphic variant. This suggests that this gene may be the major functional gene regulating first fruit spur height in cotton. Gene functional annotation and description play a crucial role in providing valuable insights into gene function and related pathways, which helps us infer candidate genes related to the target traits. For plant height QTL mapping, we focused on genes related to cell wall formation and intracellular nutrient transport. Therefore, the candidate genes we selected include *Gbar_D08G025710* (*NAC070*), which has been proven to regulate cell wall maturation in *Arabidopsis* root hair growth (Bennett et al., 2010), *Gbar_D08G025730* (*DRP2B*), a member of the dynamin protein family involved in organelle formation and transport, and *Gbar_D10G021350* (*GnTL*), which participates in cellulose synthesis in *Arabidopsis*.

3.12. Identification and analysis of candidate genes in stable QTLs for yield-related traits

Based on the mapping results of yield-related traits, we selected 1 stable QTL associated with EBN, 5 stable QTLs associated with SBW, and 3 stable QTLs associated with LP. After comparing with the physical map, we identified 802 genes within the selected QTL regions, among which 99 genes carried polymorphic variation sites (Table S16). The QTL associated with EBN is a pleiotropic QTL that shares genetic basis with FFSH, and we have conducted further analysis on this. In the

mapping of SBW, there were no annotated polymorphic sites in the genes within the *qSBW_A06_1* interval, while 67 genes in the other QTL intervals were annotated with variations. The most appropriate genetic model for SBW is a minor effect polygenic model, and it is also the trait most affected by environmental effects. We speculate that these genes may collectively determine the formation of SBW. However, in genetic studies, strong mutations may lead to larger phenotypic differences and are therefore widely prioritized. We found that 3 genes had codons that changed to "stop codons" (stopgain), and 4 genes had non-frameshift deletions or insertions. Among them, the *Gbar_D01G023430* gene showed dominant expression during embryo sac development. It encodes the LAC4 protein, which has laccase activity and has been confirmed to be involved in lignin synthesis and the regulation of cell wall biosynthesis (Blaschek et al., 2023). It is speculated that this gene may be a key regulator of seed cell wall development. Transcriptomics can provide comprehensive gene expression information. We analyzed two transcriptomic datasets from SRA database with significant differences in LP (Duan et al., 2022; Zhao et al., 2022), and screened differentially expressed genes with a threshold of $|\log_2(\text{FoldChange})| > 1$ and P-value < 0.01 (Table S17, Fig. 6A-D). In the mapping of LP, the total physical length of the 3 QTL intervals associated with LP was 1.10 Mb, and we obtained 82 genes (Table S16), among which 11 differentially expressed genes were identified as candidate genes (Fig. 6E). The results showed that the *Gbar_A13G003180* and *Gbar_A13G003290* genes had the most significant differences (Fig. 6F). The former belongs to the SWI2/SNF2 protein plant-specific subfamily and encodes chromatin remodeling protein (*DRD1*), which regulates the methylation of plant promoters (Dong et al., 2011). The latter encodes ADP-glucose pyrophosphorylase (*APL3*), which is the first rate-limiting step in catalyzing starch biosynthesis (Veyres et al., 2008). In addition, the *Gbar_A13G003170* and *Gbar_A13G003350* genes showed differential expression in the 5 DPA sample (Fig. 6F). The former encodes DNA methyltransferase *CMT3*, and the latter encodes the RAD protein belonging to the *MYB* gene family. Both have been demonstrated to participate in plant growth and development. We speculated that these candidate genes may play similar biological roles in the formation of LP.

3.13. Stable QTL haplotype analysis and molecular marker-assisted breeding validation

Stable QTLs were important genetic loci that regulate target traits and serve as major-effect QTLs. The nearby markers of stable QTLs have high practical value in molecular marker-assisted breeding. We focused on studying stable QTLs across multiple environments, namely *qLP_A13_1*, *qLP_D06_1*, *qLP_D09_1*, and *qSBW_D01_2*. Haplotype analysis was performed using bin markers at the ends of the QTLs. The genotyping of bin markers revealed significant phenotypic differences among different haplotypes in the RIL population, which corresponded to the positive or negative additive effects of the QTLs. To validate whether the genotyping markers could be used for selecting target traits in other materials, we utilized a natural population of 287 high-depth resequenced sea island cotton accessions for validation (Table S18). The results showed that materials carrying favorable allelic variants had higher LP. Compared to materials without the favorable alleles, materials carrying *qLP_A13_1* exhibited a significant increase in LP by 5.48% and 5.33% (Fig. 6G, H), materials carrying *qLP_D06_1* showed a slight increase in LP by 1.23% and 0.53%, but the improvement effect was not significant (Fig. 6I, J), materials carrying *qLP_D09_2* had a highly significant increase in LP by 3.46% and 3.47% (Fig. 6K, L). Additionally, materials carrying both *qLP_A13_1* and *qLP_D09_2* showed a highly

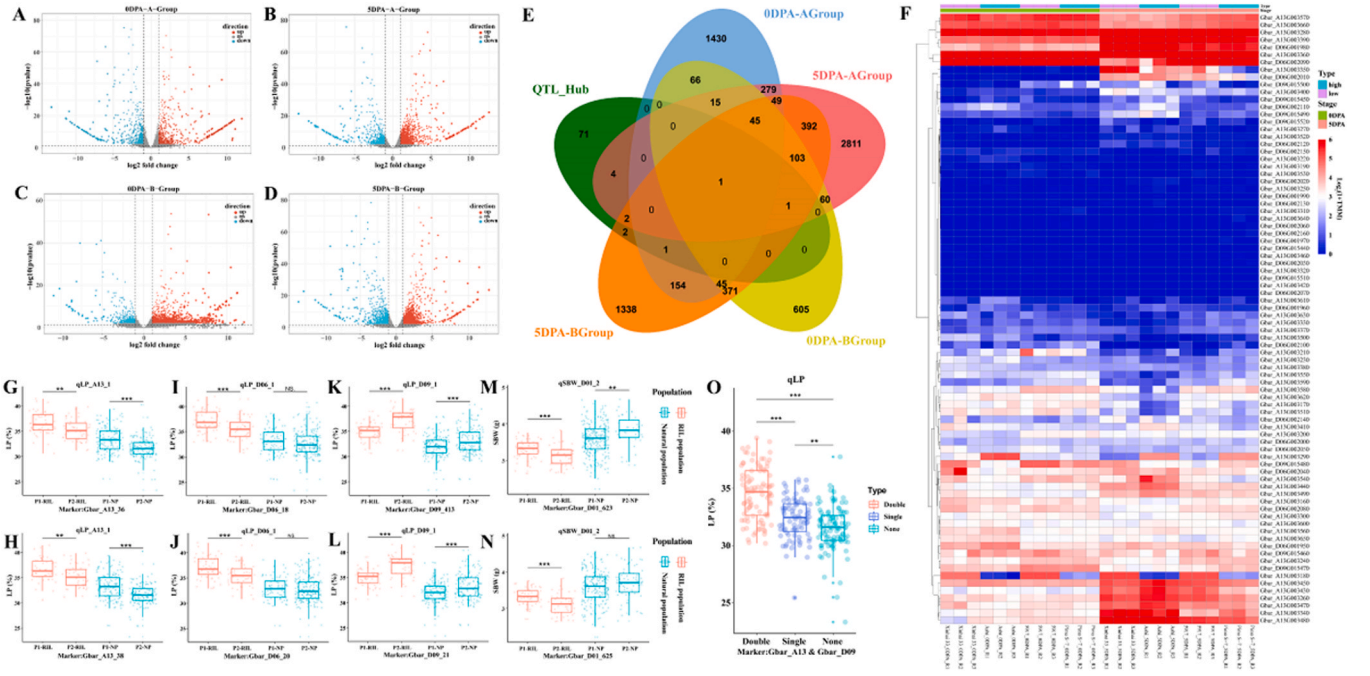


Fig. 6. Identification and analysis of stable QTLs for yield-related traits in sea island cotton. (A) Differentially expressed genes between the 5917 and Pima S-7 lines' ovules of 0 DPA are shown on the volcano plot. (B) Differentially expressed genes between the 5917 and Pima S-7 lines' fibers of 5 DPA are shown on the volcano plot. (C) Differentially expressed genes between the Xinhai 33 and Ashi lines' ovules of 0 DPA are shown on the volcano plot. (D) Differentially expressed genes between the Xinhai 33 and Ashi lines' fibers of 5 DPA are shown on the volcano plot. (E) Overlapping genes between candidates for LP and differentially expressed genes. (F) Heatmap of the expression level of candidate genes in stable QTLs for LP trait during cotton fiber development. (G, H) Comparison of the LP trait for the *qLP_A13_1* haplotype in both the RIL population and the natural population. (I, J) Comparison of the LP trait for the *qLP_D06_1* haplotype in both the RIL population and the natural population. (K, L) Comparison of the LP trait for the *qLP_D09_1* haplotype in both the RIL population and the natural population. (M, N) Comparison of the SBW trait for the *qSBW_D01_2* haplotype in both the RIL population and the natural population. (O) Comparison of the LP trait for the *qLP_A13_1* and q*LP_D09_1* haplotype in both the RIL population and the natural population. Note: *, **, *** indicates significant levels at $P < 0.05$, 0.01, 0.001, respectively, determined by Student's *t* test.

significant increase in LP by 6.99%, while materials carrying only *qLP_A13_1* or *qLP_D09_2* had a significant increase in LP by 2.73% (Fig. 6O). In terms of SBW, materials carrying *qLP_D01_2* exhibited an increase in SBW by 6.62% and 3.29%, but the improvement effect reached a significant level only with the Gbar_D01_623 marker (Fig. 6M, N). These results indicate that the bin markers at both ends of *qLP_A13_1* and *qLP_D09_2* can be designed for application in marker-assisted breeding. However, in QTLs with small additive effects, even stable QTLs may not be suitable for molecular marker-assisted breeding, as demonstrated by the unstable markers at the ends of *qLP_D06_1* and *qSBW_D01_2* in this study.

3.14. Differential analysis of cotton fiber transcriptomes at different developmental stages

Regulation of cotton fiber cell differentiation and development is controlled by highly programmed gene networks. To unravel the molecular mechanisms underlying cotton fiber development, comparative transcriptome analysis was performed on cotton fibers at different developmental stages from the RIL population parents. The statistics showed that the average number of clean reads obtained per sample was 23,993,599, with a Q30 ≥ 94.36%, an average overall alignment rate ≥ 95.60%, and an average unique alignment rate ≥ 86.21% (Table S19). qRT-PCR analysis of the relative expression levels of 10 genes demonstrated the reliability of the RNA-seq data, which could be used for subsequent analysis (Supplementary Figure 1 A, B-K). Clustering and correlation analysis revealed that the three biological replicates of the samples clustered together (Supplementary Figure 2 A, B). The principal component analysis (PCA) yielded the same conclusion as expected (Supplementary Figure 2 C). Furthermore, we found that the developmental stages of 0–10 DPA between the parents were difficult to distinguish, indicating that the gene regulatory networks involved in early fiber development were expressed similarly between the parents.

The immature cotton fibers were measured using the flow-through method. The rapid elongation phase of the paternal fibers lasted from 20 to 25 DPA (days post-anthesis), while the maternal fibers continued to elongate until 25–30 DPA, gradually slowing down and eventually ceasing elongation, indicating asynchronous elongation phases between the parents (Table S21). Differential gene expression analysis revealed 15,977 upregulated differentially expressed genes and 17,266 downregulated differentially expressed genes identified in different materials at the same developmental stage (Table S20, Fig. 7A). The dynamic changes in cotton fiber length helped us better understand the biological significance of differential analysis. In the differential analysis of different materials at the same developmental stage, a large number of genes were downregulated at 15 DPA (Fig. 7B), corresponding to the initiation stage of secondary wall synthesis during fiber growth and development. By querying the gene expression matrix, it was found that the differential expression was caused by increased expression of gene 06E2062, which aligns with the biological transition. Therefore, we speculate that the upregulated genes in fiber 06E2062 at 15 DPA may be key genes regulating the initiation of secondary wall development. In the differential analysis of the same material at different developmental times, a significant increase in differentially expressed genes was observed in the poor-quality fiber 06E2062 at 20–25 DPA, while the high-quality fiber Xinhai 21 exhibited a similar number of differentially expressed genes at 25–30 DPA (Fig. 7D). Therefore, we hypothesize that the developmental stage of 25 DPA is a critical period for exploring the transition mechanism from fiber elongation phase to secondary wall thickening phase. Through KEGG enrichment analysis, differentially expressed genes in 15 DPA fiber development were significantly enriched in pathways such as starch and sucrose metabolism, amino

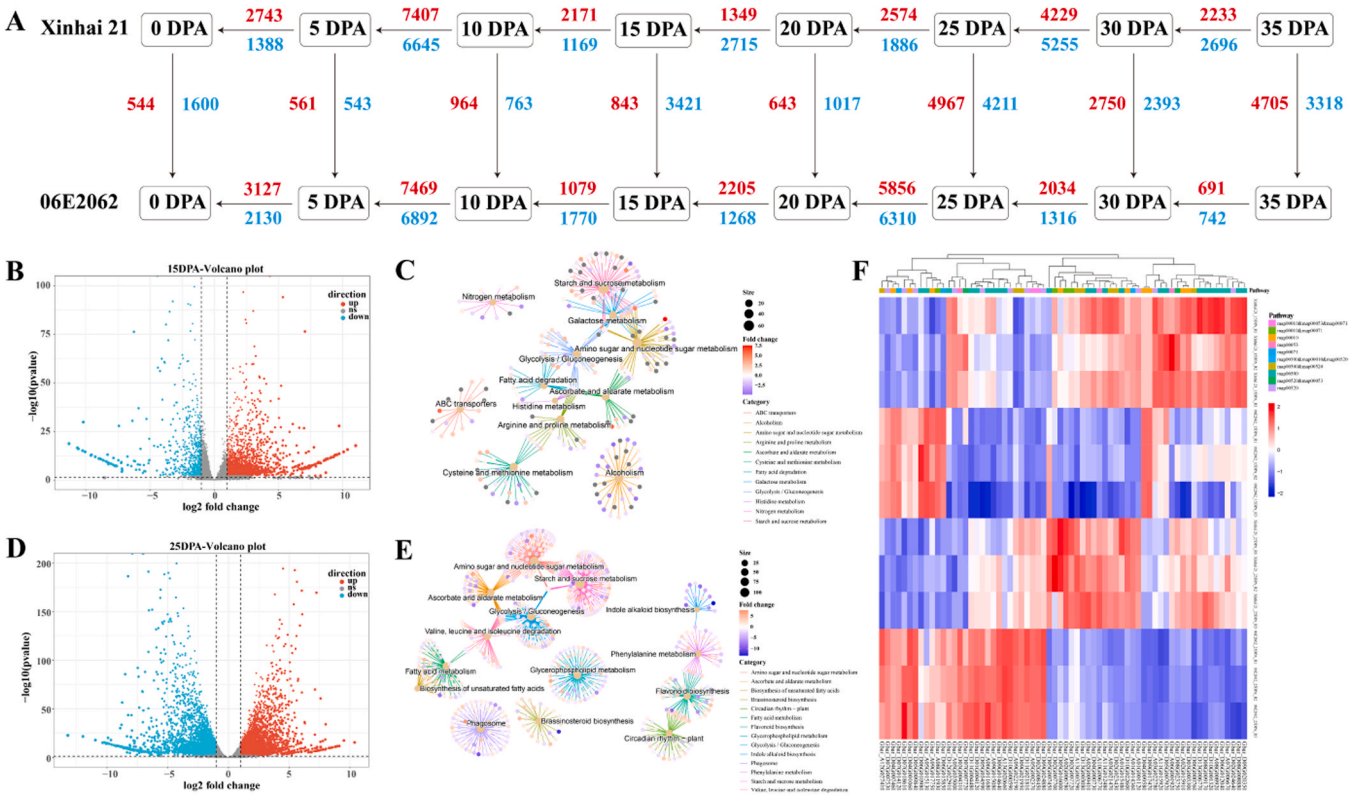


Fig. 7. Identification and analysis of candidate genes during secondary cell wall development in *G. barbadense*. (A) DEGs between the lines Xinhai 21 and 06E2062 at different stages of fiber development. Note: Around the arrows are the numbers indicating the number of genes differentially expressed for the respective comparison. Red: upregulation; Blue: downregulation. (B, D) Differentially expressed genes between the Xinhai 21 and 06E2062 lines' fibers of 15 or 25 DPA are shown on the volcano plot. (C, E) KEGG pathway functional classification of DEGs at 15 or 25 DPA between the lines Xinhai 21 and 06E2062. (F) Heatmap of the expression levels of candidate genes involved in secondary wall development during cotton fiber development.

sugar and nucleotide sugar metabolism, and glycolysis/gluconeogenesis (Table S22, Fig. 7C); differentially expressed genes in 25 DPA fiber development were significantly enriched in pathways such as starch and sucrose metabolism, flavonoid biosynthesis, phenylalanine metabolism, glycerophospholipid metabolism, and citric acid cycle (Table S22, Fig. 7E), which play important roles in plant secondary wall synthesis. By taking the intersection of the differentially enriched genes in the five common pathways between the two stages, we obtained 66 candidate genes, among which 11 genes were enriched in three different pathways, with the most enrichment in starch and sucrose metabolism (map00500), followed by amino sugar and nucleotide sugar metabolism (Fig. 7F). These metabolic pathways provide the basic materials and energy required for synthesizing secondary walls in fibers, directly or indirectly affecting the development process of secondary walls. Therefore, the 66 genes that were commonly enriched may play important roles in the metabolic network of fiber secondary wall development.

3.15. Identification of key genes related to fiber elongation through WGCNA analysis

In the weighted gene co-expression network analysis (WGCNA), multiple iterations indicated that a (soft-thresholding power) $\beta = 13$ helped to reduce noise and minimize the impact in the network (Fig. 8A). When the module correlation was $\geq 75\%$, they were merged into the same module, and a total of 13 distinct expression modules were identified in this study (Fig. 8B). Analysis using the LinkET package showed that the turquoise, green, and black expression modules had the highest and highly significant correlations with the phenotypes, with correlation coefficients of 0.82, 0.71, and 0.41, respectively. These modules contained 13,949, 1977, and 284 genes, respectively (Fig. 8C). GO enrichment analysis revealed that the MEturquoise module was enriched in 269 biological processes, cellular components, and molecular functions (Table S23, Fig. 8D). The MEgreen module was enriched in 41 biological processes, 9 molecular functions, and 25 cellular components (Table S23, Fig. 8E). The MEblack module did not show

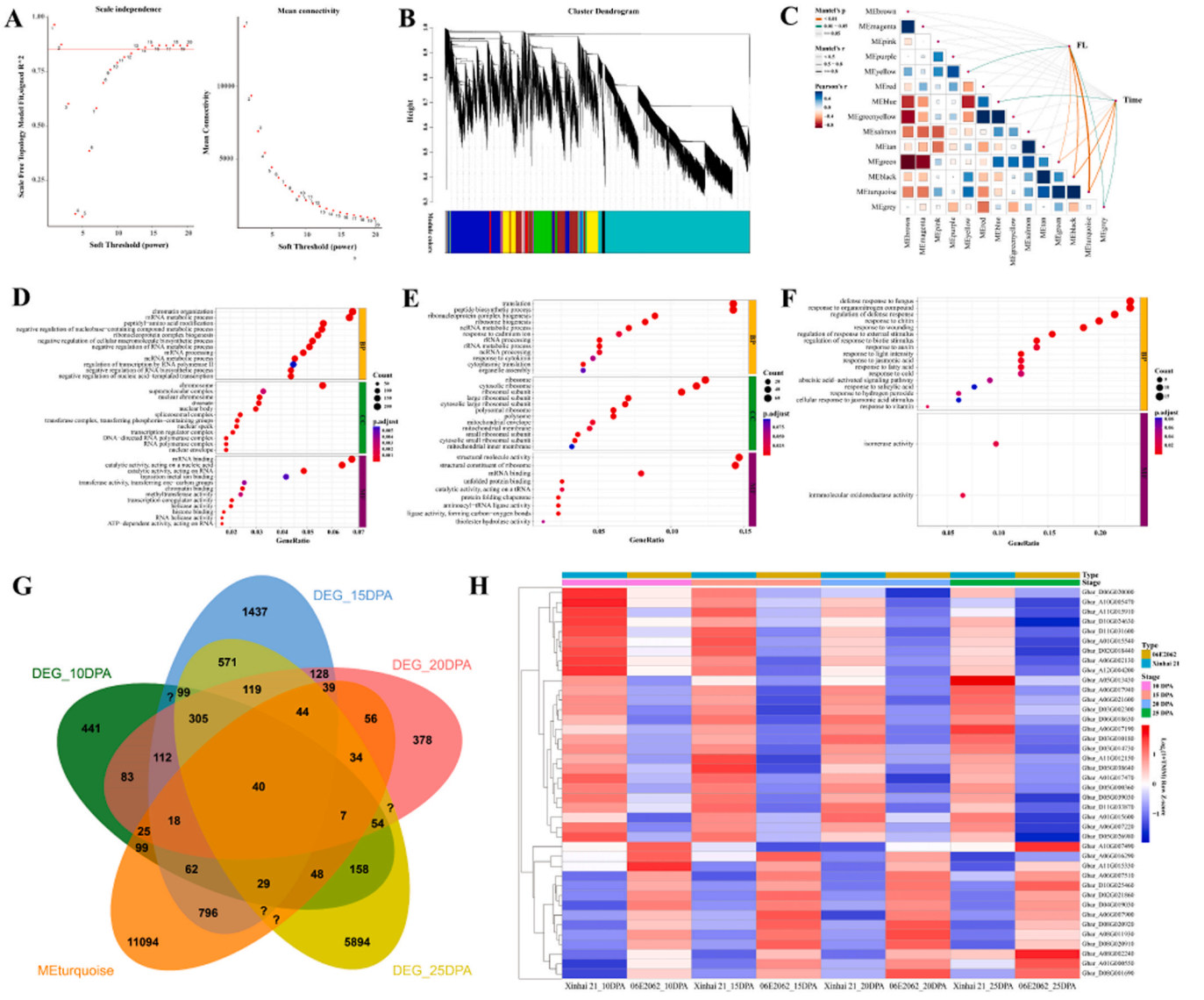


Fig. 8. WGCNA analysis of different developmental stages' fibers in *G. barbadense* to identify candidate genes involved in rapid elongation. (A) Determination of soft threshold of weighted gene co-expression network analysis. (B) Cluster tree showing WGCNA-identified co-expression modules. (C) Coexpression correlation analysis of co-expression modules in *G. barbadense*. (D) MEturquoise modules GO enrichment analysis. (E) MEgreen modules GO enrichment analysis. (F) MEblack modules GO enrichment analysis. (G) Overlapping genes between candidates for MEturquoise modules and differentially expressed genes in rapid elongation. (H) Heatmap of the expression levels of candidate genes involved in fiber rapid elongation during cotton fiber development.

enrichment in cellular components but was enriched in 75 processes related to plant hormone response, environmental or biotic stimulus response, and signal transduction and release in biological processes. It was also enriched in isomerase activity and oxidoreductase activity in molecular functions (Table S23, Fig. 8F). We speculate that this module may play an important role in the biological processes of fiber cell response to stimuli and hormone signaling.

To explore the genes related to fiber elongation, we selected DEGs during the rapid elongation stage of fiber development (10–25 DPA) for further analysis. A total of 345 genes were identified as DEGs at four different time points. Among them, 40 genes overlapped with the MEturquoise expression module identified in the WGCNA analysis, and were considered as hub genes for fiber elongation in sea island cotton (Fig. 8G). Based on the heatmap of gene expression and differential expression fold change (Fig. 8H), we identified well-known cotton fiber development-related genes, such as *Gbar_D02G021860* (Alpha-2,4 tubulin, *TUA2*), *Gbar_D08G001690* (AP2/ERF transcription factor family protein, *ERF* transcription factor), *Gbar_A06G007510* (Beta-galactosidase, *BGAL1*), and *Gbar_A10G007490* (Glycerophosphodiester phosphodiesterase, *GDPD1*), which have been cloned in upland cotton research (Wen et al., 2023). Importantly, we also identified two genes, *Gbar_D08G020910* and *Gbar_D08G020920*, that encode the SmD1a protein, which may play a minor role in RNA splicing and indirectly promote post-transcriptional gene silencing (PTGS) (Elvira-Matelot et al., 2017). Additionally, the *Gbar_A06G007900* gene, encoding a stretch-activated ion channel on the plasma membrane, was identified. In Arabidopsis, the *MSL10* gene, which is homologous to *Gbar_A06G007900*, is located in the plasma membrane and is mainly expressed in root tips, shoot tips, and vascular tissues. It enhances cell expansion and regulates cell apoptosis (Veley et al., 2014). These genes, which have not been functionally validated in cotton, are expected to provide new gene resources for improving fiber quality.

3.16. Identification and analysis of candidate genes in stable QTL for fiber quality traits

Based on the positioning results of fiber quality traits, we selected 5 stable QTLs associated with FL, 4 with FS, and 3 with FM for candidate gene mining. A total of 578 genes were found within the QTL intervals (Table S24). Combined with transcriptome analysis, 154 differentially expressed genes were identified, of which 6 genes showed significant differential expression at 8 different stages (Fig. 9A, B). To effectively exclude false positive differentially expressed genes, we analyzed the 6 significantly different genes together with another set of transcriptome data published in our project. Among them, 5917 materials had similar fiber quality to Xinhai 21, and Pima S-7 had similar fiber quality to 06E2062. Therefore, we excluded 3 genes that were likely false positives. *Gbar_A10G025470*, *Gbar_D04G007260*, and *Gbar_D04G007460* were selected as candidate genes for further analysis (Fig. 9C, D, E). *Gbar_A10G025470* had the highest differential multiple and encoded a tyrosine protein kinase (Pkinase_Tyr), which belonged to the protein kinase family and was homologous to the *Arabidopsis* gene *At2g40270*. It was an important cell membrane receptor that may participate in the regulation of programmed cell apoptosis (Ascencio-Ibánez et al., 2008). *Gbar_D04G007260* was homologous to the *Arabidopsis* gene *PLC2* (*AT3G08510*) and encoded a phospholipase specific for phosphatidylinositol, which catalyzed the hydrolysis of phosphatidylinositol 4, 5-bisphosphate into inositol 1,4,5-trisphosphate and diacylglycerol, participating in the biosynthesis and signal transduction of auxin, and regulating the development of male and female gametophytes (Li et al., 2015b). *Gbar_D04G007460* encoded a histidine acid phosphatase protein, serving as a 5-phosphatidylinositol and 6-phosphatidylinositol phosphatase, regulating the levels of inositol pentakisphosphate (InsP5) and inositol hexakisphosphate (InsP6) in cells. In addition, for FL, *Gbar_D04G006250*, *Gbar_D04G006950*, and *Gbar_A10G025320* were selected as candidate genes (Tables S20, 24, Fig. 9F, G, H). Among them, *Gbar_D04G006250* encoded a protein located in the cortex, which played a role in the stability of cortex cells and was necessary for the formation of cortex cell filaments (Froelich et al., 2011). From a biological perspective, we speculated that this gene might be a key gene for maintaining the stability and polar elongation of cotton fiber cells. *Gbar_D04G006950* encoded a lipid transfer protein (LTP), in upland cotton, Wangzhen Guo's team identified a lipid transfer protein GhLTP4 that could bind ceramides (Cers), confirming its involvement in Cers transport, which could increase Cers content and activate the auxin response pathway, thereby promoting cotton fiber elongation and

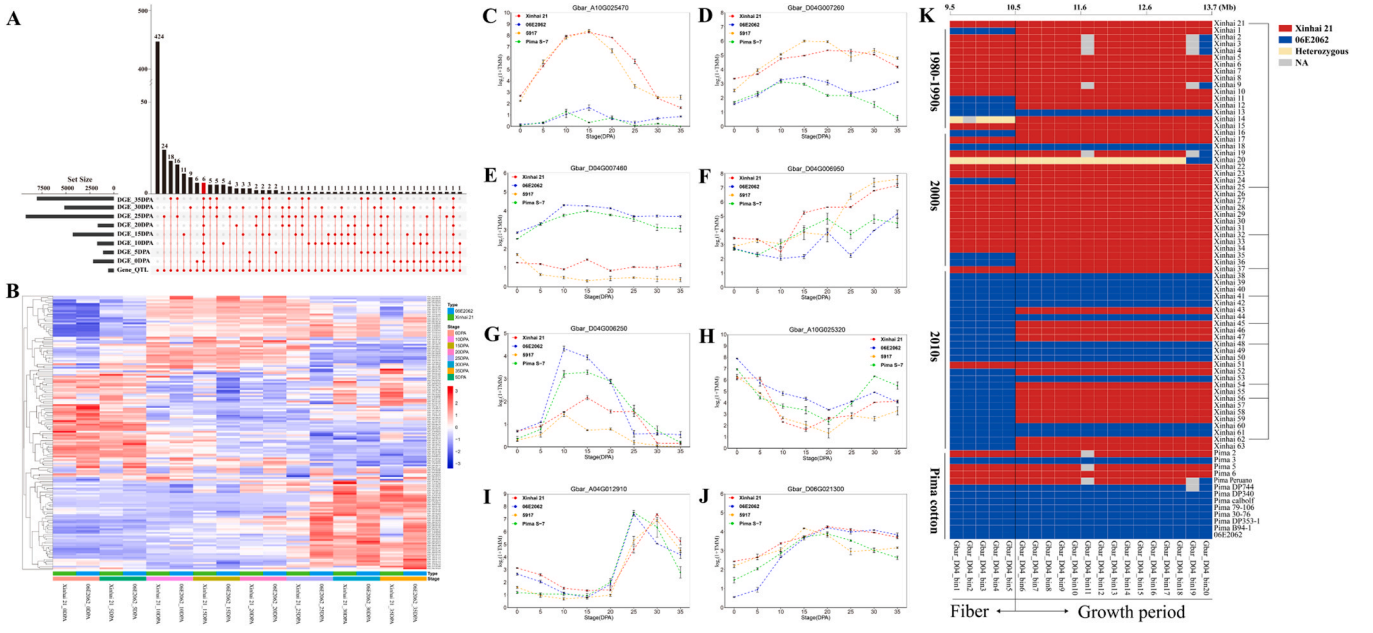


Fig. 9. Identification and analysis of stable QTLs for fiber quality traits in *G. barbadense*. (A) Overlapping genes between candidates for stable QTLs for fiber quality traits and differentially expressed genes. (B) Heatmap of differentially expressed genes within stable QTL intervals for fiber quality traits. (C-J) Expression pattern analysis of eight key candidate genes. (K) The genetic architecture of *qFL_D04_1* in *G. barbadense*.

significantly improving fiber quality (Duan et al., 2023). Whether this gene has a similar function in sea island cotton needs further verification. In terms of FS, *Gbar_A04G012910* was almost not expressed during the elongation period of fibers, but showed significant upregulation and differential expression at 25 DPA in the paternal parent 06E2062 and 30 DPA in the maternal parent sea island cotton (Fig. 9I). It encoded caffeoyl-CoA O-methyltransferase 1 (*CCoAOMT1*) in the lignin biosynthesis pathway. Previous studies have shown that *CCoAOMT1* catalyzes the synthesis of phenolic compounds, converting caffeoyl-CoA to feruloyl-CoA and 5-hydroxyferuloyl-CoA to sinapoyl-CoA, playing a critical role in cell wall formation (Vanholme et al., 2013). In addition, our other study on cotton fibers showed that when the *HCT* gene related to lignin synthesis was silenced, the expression of *CCoAOMT1* gene was downregulated and the number of trichomes on the plant epidermis was significantly reduced (Zheng et al., 2024). Therefore, we speculate that this gene may be involved in important processes of cotton fiber development. In terms of FM, we identified a gene, *Gbar_D06G021300*, which was homologous to *AtRHD3* in *Arabidopsis* and was essential for regulating cell expansion and normal root hair development. Additionally, we found that *Gbar_D06G021300* showed differential expression at 0 DPA and 5 DPA (Fig. 9J). Based on the developmental characteristics, fiber cells at 0–5 DPA form the final fiber diameter. Therefore, we speculate that this gene may be a candidate gene regulating cotton fiber thickness, thereby affecting the FM.

3.17. The genetic basis and breeding of qFL-D04-1

In the annotation of parental variation, 178 genes were annotated to the variant loci, with 8 different types of exonic variants annotated, including 299 nonsynonymous mutations, 17 frameshift deletions/insertions, 11 non-frameshift deletions/insertions, and 13 stop codon mutations (Table S24). It is worth noting that 75% of the annotations were from the confidence interval of *qFL-D04–1*, significantly higher than other QTLs. The loci were mapped onto the physical map, and the interval of 9.5–13.7 Mb on chromosome D04 was obtained as the confidence interval. Through comparison with the reference genome, we found that all variant loci were derived from the paternal parent 06E2062, indicating that the paternal parent has been subjected to sustained positive selection pressure at this locus, resulting in the preservation of the variant loci. Another interesting point is that the positive additive effect of *qFL-D04–1* suggests that the beneficial allele comes from the paternal parent 06E2062, while the maternal parent did not contribute to the aggregation of this excellent allele, indicating that the aggregation of this locus has an improvement effect on FL. We analyzed the genetic basis of the excellent allele of *qFL-D04–1* in different sea island cotton breeding processes using high-depth resequencing data of 76 accessions (63 Xinjiang approved varieties and 13 Pima cotton varieties) (Fig. 9K). By analyzing the genetic composition of different materials using 20 bin markers, the study showed that only a subset of early Pima cotton materials aggregated the excellent allele, but it was widely aggregated in subsequent breeding with different series of materials. In the history of Chinese sea island cotton breeding, only Xinhai 13 and Xinhai 18 aggregated the excellent allele before 2010, while 44.44% of the varieties aggregated the excellent allele after 2010, indicating that this excellent allele has gradually been utilized in modern breeding. Interestingly, Pima cotton retained the entire genetic basis of this locus, but the excellent allele genetic basis of 9.5–10.5 Mb has been more widely utilized in Chinese sea island cotton breeding. Approximately 25% of the varieties carried the excellent allele genetic basis of 9.5–10.5 Mb in early breeding, while about 92.6% of the approved varieties carried this excellent allele genetic basis in recent years. Combining the pedigree and phenotypic evolution of Xinhai 21, it was found that materials carrying the excellent allele genetic basis of 10.5–13.7 Mb have a longer growth period. Therefore, we speculate that fiber length and growth period share the genetic basis of *qFL-D04–1*, and there may be a linked inheritance between the two traits. This linkage was broken in the breeding process in China around 2010, and the interval of 9.5–10.5 Mb may have an independent effect on improving fiber length, while the interval of 10.5–13.7 Mb may still have a linked inheritance between fiber length and growth period that has not been broken.

4. Discussion

4.1. Fundamental elements in trait analysis: phenotypic variation, interrelationships, and genetic models

Genetic research heavily relies on suitable genetic populations. Parents with significant phenotypic differences provide a broader genetic background for the RIL population, and statistical analysis showed that the RIL population in this study exhibited abundant genetic diversity and phenotypic variation, providing more genetic information and statistical power for phenotypic analysis. The correlation analysis and estimated genetic parameters in this study were similar but not identical to previous research results. Strong correlations were observed among the same types of traits (Wang et al., 2020; Zhang et al., 2023, 2020), but there was no significant negative correlation between yield traits and quality traits. Agronomic and yield traits exhibited larger variation coefficients and were more susceptible to environmental influences, while LP and fiber quality traits were predominantly regulated by genotypes, showing higher broad-sense heritability (Fan et al., 2018; Fang et al., 2021), although fiber elongation lacks heritability. Phenotypic analysis reveals the interrelationships and genetic parameters among traits, which, although insufficient to explain deep-seated genetic questions, serve as important starting points and directions for dissecting genetic foundations and play crucial roles in genetic research. From a breeding perspective, strongly correlated traits may require combined selection to maintain or enhance the expression of these beneficial traits. Conversely, weakly correlated traits may need separate optimization to maximize the performance of each trait. From a genetic analysis perspective, the correlation analysis among traits indicates their interrelationships, suggesting shared genetic genes or regulatory pathways between highly correlated traits. In different environments, most traits showed strong positive correlations, indicating the close connection between the stability of these traits and their genetic basis, which was supported by higher broad-sense heritability. Importantly, the prediction of the most suitable genetic model provides reference information for the genetic basis of quantitative traits, helping extract true genetic information related to traits from complex data (Kearsey and Pooni, 2020). In summary, phenotypic analysis provided fundamental information about the RIL population, serving as a starting point for optimizing breeding strategies and dissecting genetic mechanisms, offering valuable clues and insights.

4.2. QTL localization and genetic basis analysis revealed by high-density genetic map of Sea Island Cotton

Next generation sequencing (NGS) technology has become an important tool in current crop genetic breeding research (Song et al., 2023), providing genome-wide SNP markers. The development of SNP markers relies on genome sequencing and assembly. With the rapid development of sequencing technologies, the cotton genome has undergone three iterations of assembly (Wen et al., 2023), significantly improving genome integrity and continuity (Chang et al., 2024). So far, multiple high-density genetic maps have been constructed for upland cotton within and between species, elucidating the genetic basis of various agronomic traits and identifying important candidate genes (Gu et al., 2020; Zhang et al., 2019, 2020). In contrast, the genetic analysis of sea island cotton has been limited. In 2018, our team utilized GBS technology for genotyping and constructed the first high-density genetic map for sea island cotton, which was used to locate QTLs related to fiber traits (Fan et al., 2018). Until the present time, Professor Tianzhen

Zhang's team investigated the structural variations contributing to the formation and differentiation of cultivated tetraploid cotton using Tanguis, Karnak, VIR29TV, and Xinhai 21 as parents. They constructed three high-density genetic maps within these populations using 4612, 3177, and 3488 bin markers, respectively, and identified 299 QTLs associated with agronomic traits, yield, and quality (Jin et al., 2023). Compared to the genetic maps in this study, the RIL population and parental sequencing in our study had higher depth, more pure and polymorphic loci, and a total of 5295 bin markers in the figure. This effectively increased the coverage and saturation of the genetic map, making it the highest-density genetic map known for sea island cotton within a single population. This lays a solid foundation for subsequent QTL mapping and gene discovery.

Compared with previous studies, most early research used interspecies or intra-species populations of upland cotton, and the QTL mapping results were not integrated into the physical genome map (Said et al., 2015). Therefore, it is difficult to compare the results of this study with previous studies. In this study, we collected QTL mapping results that have been integrated into the physical map as comprehensively as possible, with a focus on the association analysis results of six published natural populations of sea island cotton. A total of 57 genetic loci related to agronomic traits, yield, and fiber quality were found to overlap or be adjacent to the QTL regions reported by previous studies, including 9 related to agronomy, 20 related to yield traits, and 28 related to fiber quality (Table S25). At the same time, we identified 46 QTLs that contained loci obtained from previous association analysis, including 4 related to agronomy, 21 related to yield traits, and 21 related to fiber quality (Table S25) (Jin et al., 2023; Wang et al., 2022c; Yu et al., 2021; Zhao et al., 2022). For example, the *qFFSH_D07_1* locus associated with FFSH has been identified in multiple independent studies and has been well confirmed as a locus regulating cotton first fruit spur height. The protein proline to serine variation at position 113 of the *Gbar_D07G011870* gene's fourth exon is the genetic basis for the formation of different phenotypes (Si et al., 2018; Wen et al., 2021), which is consistent with our study. It is worth noting that we localized this locus within a range of 0.94 Mb, which is not as accurate as the association analysis (0.35 Mb), but it is significantly better than the BSA-seq localization results (2.3 Mb). Most of the QTLs obtained in this study overlap with previous studies, indicating a high level of confidence in the QTLs obtained in this study. Importantly, 17 QTL clusters discovered in this study provide a reasonable explanation for the phenotypic correlation results. Based on the additive effects of the QTLs forming the QTL clusters, it was found that the additive effects of QTL clusters related to FL and FS were consistent. *qClu-A06–2*, *qClu-A10–1*, *qClu-A10–2*, and *qClu-D04–1* may be the genetic basis for the highly significant negative correlation between FM and FL and FS. *qClu-D01–2*, *qClu-D12–1*, and *qClu-D13–2* may be the genetic basis for breaking the negative correlation between quality traits and yield traits. These results not only confirmed that phenotypic analysis provided clues for deciphering the genetic basis but also provided another perspective to prove the reliability of the QTLs obtained in this study. Previous studies have shown that cotton fiber QTLs were unevenly distributed in the At and Dt subgenomes, with more loci regulating fiber development in the Dt subgenome than in the At subgenome (He et al., 2021). In this study, the distribution of QTLs related to FS in the Dt subgenome was about twice that in the At subgenome, which was similar to previous studies. However, the distribution of QTLs related to FL in the At subgenome was about twice that in the Dt subgenome, which was opposite to previous studies. Interestingly, this result was consistent with the breeding history of Xinjiang, where it was generally believed that sea island cotton in Xinjiang had a problem of excessive FL and insufficient FS, indicating that early breeding focused on FL and overlooked FS (Zheng et al., 2022). We speculated that this may be fundamentally due to breeders focusing on breeding in the At subgenome and overlooking the loci regulating FS in the Dt subgenome.

4.3. Molecular mechanisms and key genes involved in the development of excellent traits

Identification of functional genes within QTL intervals is crucial as they are considered key factors that directly regulate or influence target traits. In crops, the presence of genetic linkage drag is one of the important factors restricting simultaneous improvement of important economic traits. In rice, studies based on rice QTL maps have shown that approximately 25% of the rice genome carries potential genetic linkage drag (Wei et al., 2021). In cotton, genetic linkage drag often leads to negative correlations between fiber quality and yield-related traits, with the negative correlation between FL, FS, and LP being the most typical (Huang et al., 2021). This severely restricts the simultaneous improvement of cotton fiber quality and yield. The advancement of modern molecular biology techniques has provided researchers with high-precision genome editing tools, allowing for precise modification of alleles and effectively reducing adverse effects. This breakthrough in overcoming genetic linkage drag enables researchers to design ideal crops. Therefore, the identification of functional genes within QTL intervals is of paramount importance.

In agronomic traits, the *Gbar_D07G011870* gene, encoding the TFL1 protein, was a functional gene that regulateed the height of the first fruit spur branch. However, in a recent study by the Wang lab, it was found that a single nucleotide mutation in the *GaTFL1* gene led to the loss of gene function due to the deletion of amino acids at the RNA 5′ splice site, revealing a novel mechanism by which *TFL1* gene mutations altered cotton structure (Liu et al., 2024). The plant cell wall played a critical role in providing support and protection within the plant body, and its synthesis rate directly affected plant cell elongation. The *Gbar_D08G025710* gene, which promoted cell wall formation in *Arabidopsis*, was involved in regulating root cap cell maturation and shedding (Bennett et al., 2010). Therefore, we speculated that this gene may be closely associated with plant height development. In yield-related traits, two *LAC* laccase genes (*Gbar_A13G003320* and *Gbar_D01G023430*) had been identified as candidate genes for SBW. Among them, *Gbar_D01G023430* encoded the LAC4 protein, and a study had shown that *GhLAC4* was involved in G-lignin biosynthesis. *GhmiR397b* was involved in the formation of lignin in cotton fibers, and *GhLAC4* was a target gene of *miR397b* that regulated cotton fiber development (Ding et al., 2017). Another study proposed a model in which *GhLAC4* regulated the triacylglycerols (TAG) biosynthetic pathway, elucidating the functional role of *GhLAC4* in lipid metabolism in cotton seeds (Zhong et al., 2023). We speculated that *Gbar_D01G023430* might play a similar role in sea island cotton, promoting the formation of SBW. Four genes had been identified as candidate genes for LP, and we proposed two different regulatory pathways. The first pathway involved *MYB* transcription factors, and there was ample evidence for the regulation of cotton fiber development by *MYB* transcription factors (Wen et al., 2023). Among them, *Gbar_A13G003350* encoded the *MYB2* transcription factor, and the mechanism of *GaMYB2* gene regulation in fiber initiation development was validated in a study by Xiaoya Chen in 2004, making it one of the earliest transcription factors identified to participate in fiber initiation development (Wang et al., 2004). The second pathway involved epigenetic regulation. In upland cotton, Guangyu Chen used Xu142 and its lintless mutant *Xu142fl* to systematically analyze the DNA 5mC and 6 mA changes at the genomic level in ovules and fibers during the fiber initiation stage, demonstrating the abundant DNA methylation in ovules during fiber initiation and providing a preliminary understanding of the mechanisms of DNA methylation and related epigenetic modifications in cotton fiber initiation development (Qin et al., 2022). *Gbar_A13G003170* and *Gbar_A13G003180* encoded different DNA methyltransferases with methylation-related epigenetic regulatory functions. We speculated that they might participate in the fiber initiation and differentiation process by regulating DNA methylation levels.

The molecular mechanism underlying the excellent fiber formation

of cotton had been extensively studied. The cotton genetic improvement team at Huazhong Agricultural University was highly representative in this field. The team first discovered that asymmetric domestication selection of the upland cotton subgenome affects the formation of high-quality fibers (Wang et al., 2017), and elucidated the genetic basis of the evolution of cotton fibers from nonexistence to existence and the formation of superior fiber quality in sea island cotton (Wang et al., 2022b, 2019). Recently, the team had also elucidated the dynamic co-regulation mechanism of fiber development and quality formation by two subgenomes of the allopolyploid upland cotton, explaining the potential of homologous genes from the subgenomes for improving fiber quality (You et al., 2023). Candidate genes related to fiber development, such as *TUA2* (*Gbar_D02G021860*), *BGAL1* (*Gbar_A06G007510*), *ERF* (*Gbar_D08G001690*), *PLC2* (*Gbar_D04G007260*), *LTP* (*Gbar_D04G006950*), and *CCoAOMT1* (*Gbar_A04G012910*), had been screened and preliminarily analyzed and confirmed in previous studies on cotton fiber development. For example, the *Gbar_D04G007260* gene, which was associated with cotton fiber cell membrane signal transduction, encodes the PLC2 protein. In 2021, the team led by Guanghui Xiao revealed that *GhPIPLC2D* promotes ethylene biosynthesis by generating IP3, further promoting cotton fiber elongation development (Zhu et al., 2021). The *Gbar_D08G001690* gene, which was related to secondary cell wall development in fibers, encodes the *ERF* gene. The team led by Xuebao Li from Central China Normal University revealed that the *AP2/ERF* transcription factor *GhERF108*, together with the auxin response factors *GhARF7–1* and *GhARF7–2*, regulates secondary cell wall development in cotton fibers through the ethylene-auxin signaling crosstalk pathway (Wang et al., 2023). Although most of the candidate genes identified in this study have already been elucidated in terms of their functions and regulatory networks in fiber development, it also demonstrates the reliability of this research work. It is worth noting that we had also identified a group of genes that had not been reported in cotton fiber development, such as the *Gbar_D04G007460* gene, which was associated with cell membrane signal transduction or fluidity in functional annotation. It had been found to promote cell expansion in *Arabidopsis* and showed significant differential expression in fiber development. The *Gbar_D04G006250* gene was a protein necessary for the formation of bast fibers and played an important role in maintaining cell morphology stability. The *Gbar_D06G021300* gene was required for regulating cell expansion and normal root hair development in *Arabidopsis*. The functional roles of these genes in fibers also need to be elucidated and their mechanisms of action clarified. In addition, most of the reported genes were derived from upland cotton or Asiatic cotton, and although they are orthologous genes of sea island cotton, they cannot be equated with having the same biological functions in sea island cotton. Cotton development is regulated by complex networks, and the same gene may play completely different roles in different cotton species. This is another important significance of conducting this study.

4.4. Genetic analysis and prospects for trait improvement of Sea Island cotton in different Cotton Regions

Due to climate limitations, the cultivation range of sea island cotton is relatively narrow. Modern sea island cotton varieties are mainly derived from three gene pools: Egyptian, American, and Central Asian types (Abdullaev et al., 2017). Xinjiang is the only production area for sea island Cotton in China. Since its first introduction in the late 1940 s, Central Asian Egyptian sea island cotton had been selected as the core variety, incorporating the unique ecological environment of Xinjiang. This had resulted in the development of unique Xinjiang-bred sea island cotton germplasm resources (Zhang et al., 2023; Zheng et al., 2022). Represented by American Pima cotton, American-type sea island cotton breeding began in 1910. To meet market demand and agricultural conditions, Pima cotton varieties such as Pima S, Pima DP, Pima PHY, and Pima HA have been developed (Zhang and Abdelraheem, 2017). These varieties not only have strong advantages in variety breeding but also stand out in the cotton industry and textile manufacturing. Although China had collected a large number of Pima germplasm resources, the utilization efficiency had been relatively low due to the limited understanding of the genetic basis of dominant traits (Fang et al., 2021; Zhao et al., 2022). In this study, the genetic composition and locus segregation patterns of the breeding backbone parents from the two major cotton regions were revealed using high-density SNP markers. Eight important trait loci with excellent alleles were identified, indicating that a large number of outstanding Pima cotton alleles are not present in the breeding backbone parents in China, providing greater utilization potential. The development of bin markers based on SNP markers not only covers a wider range but also effectively eliminates false positive SNPs, providing strong support for revealing the transmission patterns of genetic material. In the natural population of sea island cotton, we verified that bin markers at both ends of *qLP_A13_1* and *qLP_D09_1* can be designed as bin markers for breeding purposes. To reveal the genetic basis of the excellent allele locus *qFL-D04–1* in sea island cotton breeding, we analyzed the genetic composition and pedigree of approved varieties in different cotton regions. We found that the excellent allele variation of *qFL-D04–1* had undergone strong positive selection. This locus was overlooked in the early stages of variety breeding for some reason, but in the past decade, the excellent allele variation of *qFL-D04–1* had been effectively utilized. The 1 Mb region (9.5–10.5 Mb) was carried by almost all new varieties, demonstrating the improvement effect of this locus. Although this study provides preliminary guidance for production practices, it is not feasible to rely solely on individual or a few SNPs for molecular marker-assisted breeding. It is imperative to utilize the American-type gene pool resources of sea island cotton to improve the breeding process in China. Further work is needed to construct larger populations and analyze the genetic basis of excellent agronomic traits, laying the foundation for a better understanding of the adaptive evolution and trait improvement of sea island cotton.

5. Conclusions

The present study constructed a RIL population using the female parent Xinhai 21 and the male parent 06E2062. Twelve different phenotypic traits were investigated in five environments, systematically analyzing the correlation and genetic characteristics among these traits. The study provided preliminary insights into the combined effects of genetics and environment on the phenotype of cotton. Through whole-genome resequencing, a high-density intraspecific linkage map of sea island cotton was constructed. Used the ICIM interval mapping method, 169 QTLs were detected for eight traits. Through cotton fiber RNA-seq data, the dynamic changes and differentially expressed genes in fiber gene expression during different developmental stages were revealed. Importantly, 26 candidate genes potentially involved in regulating target traits were identified from stable QTLs, and their potential biological processes were preliminarily elucidated. Furthermore, two bin markers related to LP were validated as potential breeding markers, and the aggregation process of the excellent allele site *qFL_D04_1* in the history of Chinese sea island cotton breeding was discussed. These research findings further elucidated the genetic basis of important traits in sea island cotton, providing a theoretical foundation for future gene function validation and bio-breeding of sea island cotton.

CRediT authorship contribution statement

Dongcai Guo: Supervision. **Yajie Duan:** Validation, Data curation. **Ying Li:** Visualization. **Qiang Zhou:** Visualization. **Peng Huo:** Investigation. **Jun Liu:** Investigation. **Long Yang:** Investigation, Data curation. **kai Zheng:** Writing – review & editing, Resources, Funding acquisition, Data curation. **Quanjia Chen:** Writing – original draft, Validation, Resources, Funding acquisition. **Yanying Qu:** Writing – original draft, Supervision, Funding acquisition. **Yongsheng Cai:** Writing – review &

editing, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare that the work described is original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. I confirmed that no conflict of interest exists in the submission of this manuscript, and is approved by all authors for publication in your journal.

Data Availability

Data will be made available on request.

Acknowledgments

The present study was funded by the National Natural Science Foundation of China (Grant no. 32301867), the "Tianshan Talents Training Program" (Grant no. 2023TSYCLJ0012), the National Breeding Joint Research Project - Joint Research Project on Breeding of Important and Characteristic Species (Grant no. 2022MH-01), the "Revealing the Leadership" Project of Xinjiang Uygur Autonomous Region - Development of New Varieties and Key Core Technology Breakthroughs for Mechanized Harvesting of Extra-long Staple Cotton (Grant no. GMLM2023–09).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.indcrop.2024.118663.

References

Abdullaev, A.A., Salakhutdinov, I.B., Egamberdiev, S.S., Khurshut, E.E., Rizaeva, S.M., Ulloa, M., Abdurakhmonov, I.Y., 2017. Genetic diversity, linkage disequilibrium, and association mapping analyses of *Gossypium barbadense* L. germplasm. PLoS One 12, e0188125. https://doi.org/10.1371/journal.pone.0188125.

Ascencio-Ibánez, J.T., Sozzani, R., Lee, T.J., Chu, T.M., Wolfinger, R.D., Cella, R., Hanley-Bowdoin, L., 2008. Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant Physiol. 148, 436–454. https://doi.org/10.1104/pp.108.121038.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 64, 1–48. https://doi.org/10.48550/arXiv.1406.5823.

Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V., Snel, B., Scheres, B., 2010. SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in *Arabidopsis*. Plant Cell 22, 640–654. https://doi.org/10.1105/TPC.109.072272.

Blaschek, L., Murozuka, E., Serk, H., Ménard, D., Pesquet, E., 2023. Different combinations of laccase paralogs nonredundantly control the amount and composition of lignin in specific cell types and cell wall layers in *Arabidopsis*. Plant Cell 35, 889–909. https://doi.org/10.1093/plcell/koac344.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. 38, 5825–5829. https://doi.org/10.1093/molbev/msab293.

Chang, X., He, X., Li, J., Liu, Z., Pi, R., Luo, X., Wang, R., Hu, X., Lu, S., Zhang, X., Wang, M., 2024. High-quality *Gossypium hirsutum* and *Gossypium barbadense* genome assemblies reveal the landscape and evolution of centromeres. Plant Commun. 5, 100722 https://doi.org/10.1016/j.xplc.2023.100722.

Chen, R., Deng, Y., Ding, Y., Guo, J., Qiu, J., Wang, B., Wang, C., Xie, Y., Zhang, Z., Chen, J., 2022. Rice functional genomics: decades' efforts and roads ahead. Sci. China Life Sci. 65, 33–92. https://doi.org/10.1007/s11427-021-2024-0.

Deng, J., Fang, L., Zhu, X., Zhou, B., Zhang, T., 2019. A *CC-NBS-LRR* gene induces hybrid lethality in cotton. J. Exp. Bot. 70, 5145–5156. https://doi.org/10.1093/jxb/erz312.

Ding, Y., Wang, Y., Liu, J., 2017. Identification of the biochemical function of *miR397b* in *Gossypium hirsutum*. Acta Agric. Shanghai 33, 10–14.

Dong, L., Liu, M., Fang, Y.Y., Zhao, J.H., He, X.F., Ying, X.B., Zhang, Y.Y., Xie, Q., Chua, N.H., Guo, H.S., 2011. DRD1-Pol V-dependent self-silencing of an exogenous silencer restricts the non-cell autonomous silencing of an endogenous target gene. Plant J. 68, 633–645. https://doi.org/10.1111/j.1365-313X.2011.04714.x.

Duan, Y., Chen, Q., Chen, Q., Zheng, K., Cai, Y., Long, Y., Zhao, J., Guo, Y., Sun, F., Qu, Y., 2022. Analysis of transcriptome data and quantitative trait loci enables the identification of candidate genes responsible for fiber strength in Gossypium barbadense. G3. 12, 167. https://doi.org/10.1093/g3journal/jkac167.

Duan, Y., Shang, X., He, Q., Zhu, L., Li, W., Song, X., Guo, W., 2023. LIPID TRANSFER PROTEIN4 regulates cotton ceramide content and activates fiber cell elongation. Plant Physiol. 193, 1816–1833. https://doi.org/10.1093/plphys/kiad431.

Elvira-Matelot, E., Bardou, F., Ariel, F., Jauvion, V., Bouteiller, N., Le Masson, I., Cao, J., Crespi, M.D., Vaucheret, H., 2017. The nuclear ribonucleoprotein SmD1 interplays with splicing, RNA quality control, and posttranscriptional gene silencing in *Arabidopsis*. Plant Cell 28, 426–438. https://doi.org/10.1105/tpc.17.00066.

Fan, L., Wang, L., Wang, X., Zhang, H., Zhu, Y., Guo, J., Gao, W., Geng, H., Chen, Q., Qu, Y., 2018. A high-density genetic map of extra-long staple cotton (*Gossypium barbadense*) constructed using genotyping-by-sequencing based single nucleotide polymorphic markers and identification of fiber traits-related QTL in a recombinant inbred line population. BMC Genom. 19, 1–12. https://doi.org/10.1186/s12864-018-4890-8.

Fang, L., Zhao, T., Hu, Y., Si, Z., Zhu, X., Han, Z., Liu, G., Wang, S., Ju, L., Guo, M., 2021. Divergent improvement of two cultivated allotetraploid cotton species. Plant Biotechnol. J. 19, 1325–1336. https://doi.org/10.1111/pbi.13547.

Froelich, D.R., Mullendore, D.L., Jensen, K.H., Ross-Elliott, T.J., Anstead, J.A., Thompson, G.A., Pélissier, H.C., Knoblauch, M., 2011. Phloem ultrastructure and pressure flow: sieve-element-occlusion-related agglomerations do not affect translocation. Plant Cell 23, 4428–4445. https://doi.org/10.1105/tpc.111.093179.

Gu, Q., Ke, H., Liu, Z., Lv, X., Sun, Z., Zhang, M., Chen, L., Yang, J., Zhang, Y., Wu, L., 2020. A high-density genetic map and multiple environmental tests reveal novel quantitative trait loci and candidate genes for fibre quality and yield in cotton. Theor. Appl. Genet. 133, 3395–3408. https://doi.org/10.1007/s00122-020-03676-z.

He, S., Sun, G., Geng, X., Gong, W., Dai, P., Jia, Y., Shi, W., Pan, Z., Wang, J., Wang, L., 2021. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. Nat. Genet. 53, 916–924. https://doi.org/10.1038/s41588-021-00844-9.

Hickey, L.T., N. Hafeez, A., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C., Tester, M., Gao, C., Godwin, I.D., Hayes, B.J., Wulff, B.B., 2019. Breeding crops to feed 10 billion. Nat. Biotechnol. 37, 744–754. https://doi.org/10.1038/s41587-019-0152-9.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., 2019. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat. Genet. 51, 739–748. https://doi.org/10.1038/s41588-019-0371-5.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., 2009. High-throughput genotyping by whole-genome resequencing. Genome Res. 19, 1068–1076. https://doi.org/10.1101/gr.089516.108.

Huang, G., Huang, J.-Q., Chen, X.-Y., Zhu, Y.-X., 2021. Recent advances and future perspectives in cotton research. Annu. Rev. Plant Biol. 72, 437–462. https://doi.org/10.1146/annurev-arplant-080720-113241.

Jin, S., Han, Z., Hu, Y., Si, Z., Dai, F., He, L., Cheng, Y., Li, Y., Zhao, T., Fang, L., 2023. Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. Mol. Plant 16, 678–693. https://doi.org/10.1016/j.molp.2023.02.00.

Kearsey, M., Pooni, H., 2020. Genetical analysis of quantitative traits. Garland Science.

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

Lalitha, S., 2000. Primer premier 5. Biotech. Softw. 1, 270–272. https://doi.org/10.1089/152791600459894.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma. 9, 1–13. https://doi.org/10.1186/1471-2105-9-559.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R.J., Ma, Z., Shang, H., Ma, X., Wu, J., 2015a. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat. Biotechnol. 33, 524–530. https://doi.org/10.1038/nbt.3208.

Li, L., He, Y., Wang, Y., Zhao, S., Chen, X., Ye, T., Wu, Y., Wu, Y., 2015b. Arabidopsis PLC2 is involved in auxin-modulated reproductive development. Plant J. 84, 504–515. https://doi.org/10.1111/pce.13492.

Li, Y., Si, Z., Wang, G., Shi, Z., Chen, J., Qi, G., Jin, S., Han, Z., Gao, W., Tian, Y., 2023. Genomic insights into the genetic basis of cotton breeding in China. Mol. Plant 16, 662–677. https://doi.org/10.1016/j.molp.2023.01.012.

Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., Si, H., Xu, Z., Ma, Y., Zhang, B., 2021. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. Genome Biol. 22, 1–26. https://doi.org/10.1186/s13059-021-02351-w.

Liao, Y., Smyth, G.K., Shi, W., 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 41, 108. https://doi.org/10.1093/nar/gkt214.

Liu, J., Miao, P., Qin, W., Hu, W., Wei, Z., Ding, W., Zhang, H., Wang, Z., 2024. A novel single nucleotide mutation of TFL1 alters the plant architecture of *Gossypium arboreum* through changing the pre-mRNA splicing. Plant Cell Rep. 43, 1–15. https://doi.org/10.1007/s00299-023-03086-7.

Ma, Z., Zhang, Y., Wu, L., Zhang, G., Sun, Z., Li, Z., Jiang, Y., Ke, H., Chen, B., Liu, Z., 2021. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. Nat. Genet. 53, 1385–1391. https://doi.org/10.1038/s41588-021-00910-2.

McCouch, S.R., 1997. Report on QTL nomenclature. Rice Genet 14, 11–13.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., 2010. The Genome Analysis

Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
Meng, L., Li, H., Zhang, L., Wang, J., 2015. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J. 3, 269–283. https://doi.org/10.1016/j.cj.2015.01.001.
Peterson, B.G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., Lestel, M., Balkissoon, K., Wuertz, D., 2014. PerformanceAnalytics: Econometric tools for performance and risk analysis. 1. https://doi.org/http://CRAN.R-project.org/package=PerformanceAnalytics.
Qin, Y., Sun, M., Li, W., Xu, M., Shao, L., Liu, Y., Zhao, G., Liu, Z., Xu, Z., You, J., 2022. Single-cell RNA-seq reveals fate determination control of an individual fibre cell initiation in cotton (*Gossypium hirsutum*). Plant Biotechnol. J. 20, 2372–2388. https://doi.org/10.2225/vol16-issue4-fulltext-1.
Rao, J., Peng, L., Liang, X., Jiang, H., Geng, C., Zhao, X., Liu, X., Fan, G., Chen, F., Mu, F., 2020. Performance of copy number variants detection based on whole-genome sequencing by DNBSEQ platforms. BMC Bioinforma. 21, 518. https://doi.org/10.1186/s12859-020-03859-x.
Reinisch, A.J., Dong, J.-M., Brubaker, C.L., Stelly, D.M., Wendel, J.F., Paterson, A.H., 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138, 829–847. https://doi.org/10.1093/genetics/138.3.829.
Said, J.I., Knapka, J.A., Song, M., Zhang, J., 2015. Cotton QTLdb: a cotton QTL database for QTL analysis, visualization, and comparison between *Gossypium hirsutum* and *G. hirsutum* × *G. barbadense* populations. Mol. Genet. Genom. 290, 1615–1625. https://doi.org/10.1007/s00438-015-1021-y.
Shappley, Z.W., Jenkins, J.N., Meredith, W.R., McCarty Jr, J.C., 1998. An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor. Appl. Genet. 97, 756–761. https://doi.org/10.1007/s001220050952.
Si, Z., Liu, H., Zhu, J., Chen, J., Wang, Q., Fang, L., Gao, F., Tian, Y., Chen, Y., Chang, L., 2018. Mutation of SELF-PRUNING homologs in cotton promotes short-branching plant architecture. J. Exp. Bot. 69, 2543–2553. https://doi.org/10.1093/jxb/ery093.
Song, B., Ning, W., Wei, D., Jiang, M., Zhu, K., Wang, X., Edwards, D., Odeny, D.A., Cheng, S., 2023. Plant genome resequencing and population genomics: Current status and future prospects. Mol. Plant 16, 1252–1268. https://doi.org/10.1016/j.molp.2023.07.009.
Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J. 3, 739–744. https://doi.org/10.1111/j.1365-313X.1993.00739.x.
Su, X., Zhu, G., Song, X., Xu, H., Li, W., Ning, X., Chen, Q., Guo, W., 2020. Genome-wide association analysis reveals loci and candidate genes involved in fiber quality traits in sea island cotton (*Gossypium barbadense*). BMC Plant Biol. 20, 289. https://doi.org/10.1186/s12870-020-02502-4.
Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J.C., Schnable, P.S., Lyons, E., Lu, J., 2015. ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. 16, 1–15. https://doi.org/10.1186/s13059-014-0573-1.
Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P., 2013. Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. Science 341, 1103–1106. https://doi.org/10.1126/science.1241602.
Veley, K.M., Maksaev, G., Frick, E.M., January, E., Kloepper, S.C., Haswell, E.S., 2014. Arabidopsis MSL10 has a regulated cell death signaling activity that is separable from its mechanosensitive ion channel activity. Plant Cell 26, 3115–3131. https://doi.org/10.1105/tpc.114.128082.
Veyres, N., Danon, A., Aono, M., Galliot, S., Karibasappa, Y.B., Diet, A., Grandmottet, F., Tamaoki, M., Lesur, D., Pilard, S., 2008. The Arabidopsis sweetie mutant is affected in carbohydrate metabolism and defective in the control of growth, development and senescence. Plant J. 55, 665–686. https://doi.org/10.1111/j.1365-313X.2008.03541.x.
Voorrips, R., 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93, 77–85. https://doi.org/10.1093/jhered/93.1.77.
Wang, P., Dong, N., Wang, M., Sun, G., Jia, Y., Geng, X., Liu, M., Wang, W., Pan, Z., Yang, Q., 2022c. Introgression from *Gossypium hirsutum* is a driver for population divergence and genetic diversity in *Gossypium barbadense*. Plant J. 110, 764–780. https://doi.org/10.1111/tpj.15702.
Wang, L., Feng, Z., Wang, X., Wang, X., Zhang, X., 2010b. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136–138. https://doi.org/10.1093/bioinformatics/btp612.
Wang, K., Li, M., Hakonarson, H., 2010a. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 https://doi.org/10.1093/nar/gkq603.
Wang, Y., Li, Y., He, S.-P., Xu, S.-W., Li, L., Zheng, Y., Li, X.-B., 2023. The transcription factor *ERF108* interacts with AUXIN RESPONSE FACTORs to mediate cotton fiber secondary cell wall biosynthesis. Plant Cell 35, 4133–4154. https://doi.org/10.1093/plcell/koad214.
Wang, M., Li, J., Qi, Z., Long, Y., Pei, L., Huang, X., Grover, C.E., Du, X., Xia, C., Wang, P., 2022b. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. Nat. Genet. 54, 1959–1971. https://doi.org/10.1038/s41588-022-01237-2.
Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., 2017. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat. Genet. 49, 579–587. https://doi.org/10.1038/ng.3807.
Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G., 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat. Genet. 51, 224–229. https://doi.org/10.1111/pbi.13547.
Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., Yue, Z., Cong, L., Shang, H., Zhu, S., 2012. The draft genome of a diploid cotton *Gossypium raimondii*. Nat. Genet. 44, 1098–1103. https://doi.org/10.1038/ng.2371.
Wang, S., Wang, J.-W., Yu, N., Li, C.-H., Luo, B., Gou, J.-Y., Wang, L.-J., Chen, X.-Y., 2004. Control of plant trichome development by a cotton fiber *MYB* gene. Plant Cell 16, 2323–2334. https://doi.org/10.1105/tpc.104.024844.
Wang, F., Zhang, J., Chen, Y., Zhang, C., Gong, J., Song, Z., Zhou, J., Wang, J., Zhao, C., Jiao, M., 2020. Identification of candidate genes for key fibre-related QTL s and derivation of favourable alleles in *Gossypium hirsutum* recombinant inbred lines with *G. barbadense* introgressions. Plant Biotechnol. J. 18, 707–720. https://doi.org/10.1111/pbi.13237.
Wang, J., Zhang, Y., Du, Y., Ren, W., Li, H., Sun, W., Ge, C., Zhang, Y., 2022a. SEA v2.0: an R software package for mixed major genes plus polygenes inheritance analysis of quantitative traits. Acta Agron. Sin. 48, 1416–1424. https://doi.org/10.3724/SP.J.1006.2022.14088.
Watson, J.D., Crick, F.H., 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 171, 737–738. https://doi.org/10.1038/171737a0.
Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., Liu, J., Wang, Q., Olsen, K.M., Han, B., 2021. A quantitative genomics map of rice provides genetic insights and guides breeding. Nat. Genet. 53, 243–253. https://doi.org/10.1038/s41588-020-00769-9.
Wei, H., Zhao, H.-m, Yang, X.-w, Zhang, R., Wang, J.-j, 2019. Patent analysis provides insights into the history of cotton molecular breeding worldwide over the last 50 years. J. Integr. Agric. 18, 539–552. https://doi.org/10.1016/S2095-3119(18)62012-X.
Wen, X., Chen, Z., Yang, Z., Wang, M., Jin, S., Wang, G., Zhang, L., Wang, L., Li, J., Sumbul, S., He, S., Wang, z, Wang, K., Kong, Z., Li, F., Zhang, X., Chen, X., Zhu, Y., 2023. A comprehensive overview of cotton genomics, biotechnology and molecular biological studies. Sci. China Life Sci. 66, 2214–2256. https://doi.org/10.1007/s11427-022-2278-0.
Wen, T., Liu, C., Wang, T., Wang, M., Tang, F., He, L., 2021. Genomic mapping and identification of candidate genes encoding nulliplex-branch trait in sea-island cotton (*Gossypium barbadense* L.) by multi-omics analysis. Mol. Breed. 41, 34. https://doi.org/10.1007/s11032-021-01229-w.
Wendel, J.F., Grover, C.E., 2015. Taxonomy and evolution of the cotton genus, Gossypium. Cotton. 57, 25-44. https://doi.org/10.2134/agronmonogr57.2013.0020.
Wickham, H., 2009. Ggplot2: Elegant Graphics for Data Analysis. Springer Publishing Company.
Yang, Z., Wang, J., Huang, Y., Wang, S., Wei, L., Liu, D., Weng, Y., Xiang, J., Zhu, Q., Yang, Z., 2023. CottonMD: a multi-omics database for cotton biological study. Nucleic Acids Res. 51, D1446–D1456. https://doi.org/10.1093/nar/gkac863.
You, J., Liu, Z., Qi, Z., Ma, Y., Sun, M., Su, L., Niu, H., Peng, Y., Luo, X., Zhu, M., 2023. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. Nat. Genet. 55, 1987–1997. https://doi.org/10.1038/s41588-023-01530-8.
Yu, J., Hui, Y., Chen, J., Yu, H., Gao, X., Zhang, Z., Li, Q., Zhu, S., Zhao, T., 2021. Whole-genome resequencing of 240 *Gossypium barbadense* accessions reveals genetic variation and genes associated with fiber strength and lint percentage. Theor. Appl. Genet. 134, 3249–3261. https://doi.org/10.1007/s00122-021-03889-w.
Yu, J., Jung, S., Cheng, C.-H., Ficklin, S.P., Lee, T., Zheng, P., Jones, D., Percy, R.G., Main, D., 2014. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 42, 1229–1236. https://doi.org/10.1093/nar/gkt1064.
Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a J. Integr. Biol. 16, 284–287. https://doi.org/10.1089/omi.2011.0118.
Yuan, Y., Wei, X., Mao, S., Pan, J., Fang, H., Lv, H., Deng, X., Ge, Y., Wei, X., Yang, W., 2018. Genetic and Breeding Progress of Cotton. J. Plant Genet. Resour. 19, 455–463.
Zhang, J., Abdelraheem, A., 2017. Combining ability, heterosis, and genetic distance among nine elite American Pima cotton genotypes (*Gossypium barbadense*). Euphytica 213, 240. https://doi.org/10.1007/s10681-017-2036-8.
Zhang, L., Fan, A., Hong, M., Ma, Z., Chen, J., Zhao, S., Zheng, K., Tu'er, hong, T.e, x, 2023. Genetic Diversity Analysis of 647 Sea Island Cotton Germplasm Resources. J. Plant Genet. Resour. 24, 307–324. https://doi.org/10.13430/j.cnki.jpgr.20220815003.
Zhang, K., Kuraparthy, V., Fang, H., Zhu, L., Sood, S., Jones, D.C., 2019. High-density linkage map construction and QTL analyses for fiber quality, yield and morphological traits using CottonSNP63K array in upland cotton (*Gossypium hirsutum L.*). BMC Genom. 20, 889–912. https://doi.org/10.1186/s12864-019-6214-z.
Zhang, Z., Li, J., Jamshed, M., Shi, Y., Liu, A., Gong, J., Wang, S., Zhang, J., Sun, F., Jia, F., 2020. 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. 18, 239–253. https://doi.org/10.1111/pbi.13191.
Zhang, J., Percy, R.G., McCarty, J.C., 2014. Introgression genetics and breeding between Upland and Pima cotton: a review. Euphytica 198, 1–12. https://doi.org/10.1007/s10681-014-1094-4.
Zhao, N., Wang, W., Grover, C.E., Jiang, K., Pan, Z., Guo, B., Zhu, J., Su, Y., Wang, M., Nie, H., Xiao, L., Guo, A., Yang, J., Cheng, C., Ning, X., Li, B., Xu, H., Adjibolosoo, D., Aierxi, A., Li, P., Geng, J., Wendel, J., Kong, J., Hua, J., 2022. Genomic and GWAS analyses demonstrate phylogenomic relationships of *Gossypium barbadense* in China and selection for fibre length, lint percentage and Fusarium wilt resistance. Plant Biotechnol. J. 20, 691–710. https://doi.org/10.1111/pbi.13747.
Zheng, K., Cai, Y., Qu, Y., Teng, L., Wang, C., Gao, J., Chen, Q., 2024. Effect of the *HCT* gene on lignin synthesis and fiber development in *Gossypium barbadense*. Plant Sci. 338, 111914 https://doi.org/10.1016/j.plantsci.2023.111914.

Zheng, Z., Xu, H., Lin, T., Guo, R., Wang, L., Cui, J., Zhang, D., Wei, X., 2022. Evolution tendency and comprehensive evaluation of long-staple cotton bred varieties in Xinjiang. J. China Agric. Univ. 27, 55–70.

Zhong, Y., Wang, Y., Li, P., Gong, W., Wang, X., Yan, H., Ge, Q., Liu, A., Shi, Y., Shang, H., 2023. Genome-Wide Analysis and Functional Characterization of LACS Gene Family Associated with Lipid Synthesis in Cotton (*Gossypium spp.*). Int. J. Mol. Sci. 24, 8530. https://doi.org/10.3390/ijms24108530.

Zhu, L., Dou, L., Shang, H., Li, H., Yu, J., Xiao, G., 2021. GhPIPLC2D promotes cotton fiber elongation by enhancing ethylene biosynthesis. Iscience 24, 102199. https://doi.org/10.1016/j.isci.2021.102737.

Zhu, T., Liang, C., Meng, Z., Sun, G., Meng, Z., Guo, S., Zhang, R., 2017. CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol. 17, 101–110. https://doi.org/10.1186/s12870-017-1039-x.