

Construction of a Genetic Map by Resequencing and Specific Locus Amplified Fragment sequencing (SLAF-seq) for QTL fine mapping of plant height in upland cotton (*Gossypium hirsutum* L.)

Wenqing Qiao

Cotton Research Institute of Chinese Academy of Agricultural Sciences

Gentu Yan

Cotton Research Institute of Chinese Academy of Agricultural Sciences

Ning Wang

Cotton Research Institute of Chinese Academy of Agricultural Sciences

Qun Huang (✉ huangqun@caas.cn)

Cotton Research Institute of Chinese Academy of Agricultural Sciences

Research

Keywords: Upland cotton (*Gossypium hirsutum* L.); SLAF-seq; Plant height; QTL identification; candidate genes

Posted Date: August 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-64347/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: *Gossypium hirsutum* (upland cotton), the most widely cultivated cotton species in the world, is an important raw material for the textile industry. Using high-throughput sequencing to construct high-density genetic maps can be widely used in quantitative trait locus (QTL) mapping and molecular marker-assisted breeding.

Results: In this study, an F_2 population was used to construct a genetic map by parental resequencing and progeny SLAF-seq. The F_2 population consisted of *Gossypium hirsutum* L. cultivars: CCRI 49 and 396289. The genetic map contained 4,607 single nucleotide polymorphisms markers, which overlapped a total length of 3,063.4 cM with an average genetic distance of 0.898 cM between adjacent markers. A high-density genetic map was used to map the QTLs of plant height in four environments. 16 QTLs were obtained, which could explain 2.07%–19.04% of the phenotypic variation. A total of 1,028 candidate genes were identified in the confidence interval and were categorized according to function through cluster of orthologous groups analysis, gene ontology analysis, and Kyoto Encyclopedia of Genes and Genomes analysis. Within the QTLs confidence interval, the D05 chromosome(ChrD05) was finely mapped using Mutmap-like strategy, and the reliability was validated by qRT-PCR of 18 candidate genes. Finally, we obtained 14 candidate genes that are most likely to be related to plant height.

Conclusions: This study provides a successful application of parental sequencing and progeny SLAF-seq strategy in the genetic map of upland cotton in an F_2 population. This study provides theoretical support for molecular marker-assisted breeding and plant height-heterosis in upland cotton and provides a fine mapping scheme for QTLs. In addition, the combination of multiple analytical methods also provided a solution for QTL fine mapping.

Background

Upland cotton (*Gossypium hirsutum* L.) is the largest cultivated species. It is characterized by high yield and good fiber quality, thus is an important raw material in the textile industry and also closely related to people's livelihoods (Dass et al. 2016; Chen et al. 2007; Li et al. 2013). Higher labor costs promote the continuous improvement of the mechanization of cotton; therefore, there is a demand for breeding "mechanized picking cotton". Plant type is an important index to evaluate whether cotton is suitable for mechanical harvesting, and plant height is an important component of plant type. For example, if plants are too short, the spacing between cotton bolls is smaller, affecting the cotton picking rate; if plants are too tall, bolls ripen late and the vigorous vegetative growth is not conducive to cotton bolting (Qi et al. 2017). One way to breed plants with desirable traits is through molecular marker technology, which is a strategy used to construct genetic linkage maps (Luo et al. 2001; Chen and Zhang 2009). Based on the construction of genetic linkage maps, functional genes can be identified by locating the quantitative trait loci (QTLs) of the target trait and preparing for molecular polymerization breeding (Zhang et al. 2016c). Through molecular marker-assisted selection (MAS), the genotypes of target traits can be screened directly to improve the breeding efficiency and speed up the breeding process (Guo et al. 2005).

With the progress of high-throughput sequencing technology and the reduction of sequencing cost, the genome sequences of Raymond cotton (*Gossypium Raimondii*) (Wang et al. 2012), Asian cotton (*Gossypium arboreum*) (Li et al. 2014), and upland cotton (*Gossypium hirsutum*) (Li et al. 2015; Zhang et al. 2015b) have been published. To identify a large number of polymorphic markers, genotyping and construction of high-density genetic mapping and other detailed sequencing work (Davey et al. 2011) should be conducted. Based on the availability of advanced recombinant inbred lines (RIL) or F_2 populations, genotyping has been used to construct high-density genetic maps (Huang et al. 2009; Xu et al. 2013; Wang et al. 2015b; Mun et al. 2015). Re-seq is a technique that can be used to find a large number of genetic markers. In addition, digestion-based simplified genome sequencing has also been used, including restriction site associated DNA sequencing (RAD-seq) (Baird et al. 2008), genotyping-by-sequencing (GBS-seq) (Elshire et al. 2011), and specific-locus amplified fragment sequencing (SLAF-seq) (Sun et al. 2013). SLAF-seq has the advantage of avoids repetitive genome sequences and highly accurate, low cost, and rapid (Sun et al. 2013), and widely used in the construction of high-density genetic maps and has been applied in many species. Jiang et al. (Jiang et al. 2015) constructed a high-density genetic map of wax gourd (*Benincasa hispida* (Thunb.) Cogn) by using SLAF-seq, the map spanned 2,172.86 cM with 4,607 polymorphic markers and an average distance between adjacent markers for 0.49 cM. Mao et al. (Mao et al. 2015) constructed a high-density genetic map of rice (*Oryza rufipogon*) by using SLAF-seq, the

total length of map was 1537.1 cM with 4740 SLAFs and the average distance between adjacent markers was 0.32 cM. Zhang et al. (Zhang et al. 2016a) constructed a high-density genetic map of soybean (*Glycine max* (L.) Merrill), the map, spanning 3020.59 cM in length, contained 6159 markers and with an average distance of 0.49 cM between adjacent markers.

In this study, an F_2 population was obtained by crossing cultivars CCRI 49 and 396289. We attempted to use this population to construct high-density genetic maps and refine the mapping of plant height QTLs by using linkage analysis and correlation analysis. Hoping to find candidate genes related to plant height. This study will enrich the genetic map of upland cotton, will provide theoretical support for molecular marker-assisted breeding, and will provide a successful application of the parental resequencing and progeny SLAF-seq strategy in an F_2 population of upland cotton. At the same time, the combination of multiple analytical methods will provide a solution for fine QTL mapping.

Methods

Plant material and phenotypic data analysis

An F_2 population was obtained from an intra-specific cross between two cultivars (lines): CCRI 49 and 396289. CCRI 49 is a cultivar with shorter, excellent fiber quality, while cultivar 396289 has the characteristics of higher, stress resistance and good adaptability, and 396289 is significantly higher than CCRI 49 on plant height. In the summer of 2015, CCRI 49 was crossed with 396289 to get F_1 seeds at Anyang, Henan Province. The F_1 seeds were plant during winter in Sanya, Hainan Province, and self-pollinated to produce F_2 seeds. The F_2 seeds were plant and self-pollinated to get $F_{2:3}$ seeds in 2016. 160 $F_{2:3}$ families were selected and planted in three environments in 2017. A survey of phenotypic data was conducted in late September at 2016 and 2017. The plant height of the F_2 population was taken as the measured value, while the plant height of the $F_{2:3}$ pedigree was taken as the average value of the plant line. In 2016, the F_2 population was planted in Anyang, Henan Province, China. In 2017, an $F_{2:3}$ population was planted in three locations: Anyang, Henan Province; Anqing, Anhui Province, China; and Alaer, Xinjiang Autonomous Region, China. All four environments were hereinafter referred to as 16AY, 17AY, 17AQ, 17ALE, respectively. In 16AY, plants were grown in single-row plots, with 0.8-m row distances and 7-m row lengths. In 17AY, plants were grown in with 0.8-m row distances and 3-m row lengths. In 17AQ, plants were grown in 1.2-m row distances and 5-m row lengths. In 17ALE, plants were grown in 0.6-m row distances and 5-m row lengths. One-way analysis of variance (ANOVA) was used to test the significance of the height of the two parents. Descriptive statistics of the F_2 population and $F_{2:3}$ population were performed using the program Microsoft EXCEL 2010.

DNA Extractions

Young leaves of parents and the F_2 populations were sampled in July 2016 and stored at -80°C . Genomic DNA was extracted following the CTAB method (Paterson et al. 1993). Agarose gel electrophoresis (1%) was used to check DNA integrity and a Nanodrop 2000 UV-vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) was used to measure DNA purity.

Library Construction And High Throughput Sequencing

Paired-end (PE) and paired Solexa libraries were prepared according to the sequencing requirements of the resequencing platform (Illumina, Hayward, CA, USA). Genomic DNA from the CCRI 49 and 396289 cultivars were randomly interrupted. The acquisition and purification of DNA fragments, as well as the specific process of DNA library construction was conducted following the methods described by Zhang et al. (Zhang et al. 2016b). The resulting DNA library was Solexa sequenced using the Illumina PE150 (Illumina) platform at Beijing Biomarker Technologies Co., Ltd. (Beijing, China; <http://Biomarker.com.cn/>). Low-quality reads (< 30), reads with adaptor sequences, and duplicate reads were filtered out, which was done using the Burrows-Wheeler Alignment (BWA) software (Zhang et al. 2016b; Li and Durbin 2010). SNPs were detected using the GATK software (Xu et al. 2012) toolkit. To ensure the accuracy of SNP detection based on the results of clean reads of the alignment of upland cotton reference genomes (<http://mascotton.njau.edu.cn/info/1054/1118.htm>), picard was used for duplication, and GATK for local realignment and base recalibration. After a single nucleotide polymorphism (SNP) was detected using GATK, it was filtered and the final SNP site was obtained. We followed the best practice methods described by the official GATK website (<https://www.broadinstitute.org/gatk/guide/best-practices.php>).

One of the advantages of SLAF sequencing is that it is possible to use the bioinformatics method to mimic the enzymatic cleavage process according to published genomic sequences to obtain a suitable enzymatic cleavage protocol (Sun et al. 2013; Zhang et al. 2015a). For the upland cotton F₂ population, two enzymes (RsaI and HaeIII, NEB, USA) were used to digest the genomic DNA, the length of the digested fragment was 314 bp to 344 bp. The details of the SLAF-seq strategy was described by Sun et al. (Sun et al. 2013) and Zhang et al. (Zhang et al. 2015a). Pair-end sequencing was conducted using the Illumina platform at the Beijing Biomarker Technologies Corporation.

SNP Grouping And Genotyping

Specific procedures for the identification and genotyping of progeny SLAF markers are as follows. In short, the low-quality reads were filtered out (quality score < 30e), and the remaining reads were classified into the progeny by diploid barcode sequencing. Each high-quality read was then trimmed at the 5-bp terminal location. Then, 150 bp pair-end clean reads were obtained from the same sample, and the BWA software (Li and Durbin 2010) was used to map the read on the published genome of *G. hirsutum* (Zhang et al. 2015b). Sequences with greater than 95% similarity were defined as one SLAF locus (Zhang et al. 2015a). Alleles of each SLAF were defined using the minor allele frequency (MAF) assessment (Mei et al. 2017). Upland cotton is a tetraploid species; therefore, it can only have a maximum of eight alleles at one locus. If the SLAF group contained more than eight alleles or only one allele, it was considered repetitive or not polymorphic and filtered out. Therefore, SLAFs with 2–8 alleles were identified as polymorphic and divided into eight separate modes as following: ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc, and cc × ab. The F₂ population used here was derived from two homozygous inbred lines, so only the SLAF tags that had segregation patterns of aa × bb were used for subsequent analysis.

Genetic Map Construction And Evaluation

Genetic maps were constructed using Highmap mapping software (Liu et al. 2014) and referenced to the complete upland cotton genome (Zhang et al. 2015b). Highmap mapping software was used to conduct a two-point chain analysis and calculate the reorganization rate and LOD score between two markers. A cluster analysis was performed using the shortest distance method, and all of the markers were divided into multiple linkage groups. Using linkage groups as a unit, the two-step mapping method (Zhang et al. 2015a) based on an optimized high density map was used to obtain the linear arrangement of markers in the linkage group, and the genetic distance between adjacent markers was estimated and finally mapped onto a genetic map. Segregation distortion markers (SDM) are common, and some SDMs (χ^2 test, $p < 0.001$) can be selected for genetic map construction. Genetic map construction followed the methods described by Zhang et al. (Zhang et al. 2015a) and Zhang et al. (Zhang et al. 2016c). In order to construct the genetic map accurately, a haplotype map, heat map, and collinearity were used to assess the quality of the genetic map (Zhang et al. 2016c; Liu et al. 2017a).

QTL Mapping And Mutmap-like Analysis

The R/QTL package (Broman and Sen 2009) was used to identify QTLs for plant height together with the F₂ linkage map with F₂ and F_{2:3} phenotypic data from four environmental treatments (16AY, 17AY, 17AQ, and 17ALE). Plant height QTL mapping was conducted using composite interval mapping (CIM) (Zeng 1994). The permutation test (PT) significance threshold was set to 1,000 times and the positions were kept in the default mode. QTLs were named as described by Sun et al. (Sun et al. 2012). According to the standard described by Stuber et al. (Stuber et al. 1987), the QTL mode of action was determined. When the D/A absolute value was less than 0.20, it was an additive effect; 0.21–0.80 was a partially dominant effect, 0.81–1.20 was a dominant effect; and > 1.20 was an overdominance effect.

To further the finer mapping of the QTLs controlling plant height, a Mutmap-like strategy was used to identify SLAF-tagged polymorphisms. Mutmap was exploited on the basis of bulked segregation analysis (BSA) and the contribution of the allele to the mutant phenotype was determined by calculating the parameters (SNP-index) (Takagi et al. 2015). According to the Mutmap-like strategy, sequencing data were grouped and analyzed. On the basis of the segregation of plant height of the F_{2:3} population, individuals with extreme plant height were selected from the F₂ population. To get high quality SLAF markers for each individual, these markers were used to build two pools (the tallest pool and shortest pool) according to plant height. By

calculating the Δ (SNP-index) value of each marker, the QTLs of the plant height were further specified. The details of the Mutmap-like strategy are given in Zhang et al. (Zhang et al. 2015a).

Screening And Annotation Of Candidate Genes

After confirming the identity of the QTLs, the sequence information from the parents in the QTL region was extracted and the open reading frame (ORF) was predicted to obtain the gene information of the candidate region. A bioinformatics analysis was conducted to determine the non-synonymous mutation gene between parents as the final candidate gene. Using BLASTp (Camacho et al. 2009) and BLAST2GO (A et al. 2005), the data was aligned to the previously published sequence of the upland cotton genome (Zhang et al. 2015b), and annotation using multiple databases was performed on the coding genes in the confidence intervals. Candidate gene products were orthologically classified using the Cluster of Orthologous Groups of proteins (COG) database. Using the Gene Ontology consortium (GO) database, candidate genes were enriched and analyzed from the aspect of biological processes, molecular functions, and cellular components. Pathway enrichment analysis of candidate genes was conducted using Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Real-time Fluorescence Quantitative PCR(qrt-pcr)

Total RNA from the cultivars CCRI49 and 396289 was extracted using an RNAPrep Pure Plant Kit (Polysaccharides & Polyphenolics-rich, Tiangen, China). Total RNA was reverse transcribed using PrimerScript™ RT reagent Kit (Perfect Real Time, TaKaRa, Japan) to obtain cDNA for qRT-PCR. Real-time PCR was performed using an UltraSYBR Mixture (Low ROX, CWbio, China) according to the manufacturer's protocols. Specific primers for these candidate genes were designed using the Primer-BLAST NCBI online tool (Table S1). Real-time fluorescence quantitative PCR (qRT-PCR) was performed on an CFX96™ PCR instrument (Bio-Rad, USA). Three biological and technical replicates were performed in the qRT-PCR test to validate the results. Relative gene expression levels were quantified by the $2^{-\Delta\Delta Ct}$ method (KJ and TD 2001).

Results

Statistics of plant height of parents and offspring

Results from the ANOVA showed that the P -value was 0.0062, indicating that there was a significant difference in plant height among the parents. The descriptive statistics of the F_2 population and $F_{2:3}$ pedigree and their parents in the four tested environments are shown in Table 1. The skewness of plant height in all four environments was < 1 , indicating a nearly normal distribution of the phenotype.

Table 1
Statistical analysis of plant height of parents and offspring

Env	Parents				Population							
	CCRI49	396289	Rang	P -value	Min	Max	Rang	Average	Var	Skew	Kurt	
16AY	73.0	125.0	52.0	0.0062	80.0	140.0	60.0	111.6	174.3	-0.26	-0.49	
17AY	75.4	124.0	48.6		69.0	127.0	58.0	94.6	90.8	0.05	0.55	
17AQ	100.0	160.9	60.9		98.0	165.6	67.6	141.1	141.3	-0.48	0.22	
17ALE	64.7	113.9	49.2		63.4	115.1	51.7	87.1	85.6	0.17	0.19	

High-throughput Sequencing And Genotyping

A total of 355.23 Gb of high quality raw data was obtained from high-throughput sequencing, including 1,448 M pair-end reads (clear reads) with each read 150 bp long. The average GC content was 38.54%, and the average Q30 (a quality score of 30, the probability of base sequencing error is 0.1%, thus 99.9% confidence) ratio was 92.90% (Table 2). According to the sequencing

results, 343.84 M, 314.02 M, and 790.82 M reads were obtained from CCRI49 (paternal parent), 396289 (maternal parent), and the 160 F₂ offspring, respectively (Table 2). The reads from the 160 offspring ranged from 2.71 M to 9.81 M, and the average read was 4.94 M (Table S2). The average sequencing depths from CCRI49, 396289, and the offspring were 30.10×, 27.54×, and 16.93×, respectively (Table 2). The number of SNPs obtained from CCRI49, 396289, and the offspring were 1,619,902, 1,613,499, and 105,168, respectively. The developed parents and offspring SNPs were compared, the SNPs that had no polymorphism in the progeny and the SNPs that lack more and show partial segregation in the population were filtered out, with the remaining 2,207,531 SNPs used for genotyping. According to the gene coding rules (Mei et al. 2017), 1,103,837 of the 2,207,531 SNPs were successfully coded and divided into five separation modes (ef × eg, lm × ll, nn × np, aa × bb, and hk × hk) (Fig. 1). Since the parents (CCRI49 and 396289) were homozygous inbred lines, only 620,378 SNPs (aa × bb) were used to construct the genetic map.

Table 2
The sequencing data from the F₂ population

Sample	Total reads	Total bases	Q30 (%)	GC content (%)	Average depth
CCRI49	343,845,291	103,043,996,378	91.16	35.18	30.1
396289	314,016,396	94,103,385,042	90.41	34.90	27.54
Offspring	790,816,675	158,081,050,758	92.92	38.59	16.93
Total	1,448,678,362	355,228,432,178	92.90	38.54	—

Construction Of The Genetic Map With SNPs

In order to ensure the quality of the genetic map and make the markers in the linkage group as uniform as possible, the polymorphic SNP markers were filtered out according to the following rules: (1) the markers with a depth of less than 4 × were omitted; (2) the selected SNPs cover at least 50% of offspring individuals; (3) segregation distortion markers ($P < 0.001$) were omitted. Therefore, a total of 5,182 SNPs were used for mapping. These SNPs were divided into 26 linkage groups by comparison with the reference genome (Zhang et al. 2015b), and each chromosome was a linkage group. The admixture LOD (MLOD) values between every two markers were calculated. The markers with MLOD values lower than three were filtered out, and 4,607 upper mapping SNPs were obtained (Table S3). The upper mapped rate was 88.90%. A genetic map with a total distance of 3,063.4 cM was obtained, and the average genetic distance was 0.898 cM between adjacent markers (Fig. 2, Table 3). The number of markers on each chromosome ranged from 22 (ChrA06) to 650 (ChrD06), the genetic distance ranged from 25.31 cM (ChrD05) to 192.16 cM (ChrA13), and the average genetic distance ranged from 0.18 cM (ChrD06) to 1.66 cM (ChrA06). The largest gap on the map was 18.85 cM, located on ChrA09. The proportion of the gap was < 5 cM, which was more than 90%, except the gap in ChrA06 (80.95%), while ChrD05 and ChrD11 had no gap more than 5 cM (Table 3).

Quality Assessment Of The Genetic Map

Segregation distortion markers are common and can affect map construction and QTL mapping (Zhang et al. 2016c). Among the 4,607 mapped SNPs, five of them were segregation distortion markers ($P < 0.001$). The segregation distortion markers are located on ChrA09 and ChrA10, and the numbers of the two chromosomes were three and two, respectively (Table 3).

A haplotype map can be used to display recombination events and missing data for each individual (Liu et al. 2017a). In this study, haplotype maps were obtained from each of the 160 offspring using 4,607 SNPs (Fig. S1). From these haplotype maps, the percent of double crossover sites was less than 0.30%, of which the maximum was only 0.28% at ChrD11, and for ChrA03, ChrA06, and ChrD08 there were no double crossover sites. There was no missing data from the 26 chromosomes (Table 3). This indicates that the source of the larger segments of each individual in the progeny is consistent and the genetic map is of high quality.

A heat map can reflect the reorganization value of the marker, and potential sequencing problems can be found by the reorganization analysis between the markers (Liu et al. 2017a). The heat map produced by this study illustrates that the linkage between the adjacent markers on each linkage group is very strong, and with the increase of distance, the linkage between the markers gradually weakened (Fig. S2). This indicates that the sequence of the markers is correct.

Collinearity analysis was performed by locating the genomic position and genetic map of the mapped markers. Spearman correlation coefficients of each linkage group and physical map were calculated and this data is shown in Table 3. The closer the Spearman coefficient is to 1, the better the collinearity between the map and the physical map (Table 3 and Fig. 3). Most of the linkage groups are consistent with the genome, indicating that the collinearity is good and the calculation accuracy of the genetic recombination rate is high.

Table 3
Characteristics of the genetic map

Chr. ID	Total Marker	Total Distance (cM)	Average Distance (cM)	Max Gap (cM)	Gap < 5 cM (%)	Total BS	Singleton Percent (%)	Missing Percent (%)	Spearman
A01	134	142.49	1.06	15.42	95.49	0	0.13	0	0.98
A02	177	124.67	0.70	12.77	99.43	0	0.13	0	0.96
A03	318	150.92	0.47	11.65	98.74	0	0	0	1.00
A04	83	92.57	1.12	9.75	95.12	0	0.08	0	0.99
A05	106	108.78	1.03	8.15	98.10	0	0.1	0	1.00
A06	22	36.42	1.66	7.83	80.95	0	0	0	0.98
A07	238	148.95	0.63	10.27	97.47	0	0.01	0	0.99
A08	136	122.34	0.90	14.49	96.30	0	0.08	0	0.99
A09	190	109.31	0.58	18.85	97.88	3	0.03	0	0.99
A10	182	145.75	0.80	8.76	97.24	2	0.17	0	0.97
A11	115	116.72	1.01	5.59	98.25	0	0.2	0	0.95
A12	112	124.96	1.12	12.24	96.40	0	0.01	0	1.00
A13	374	192.16	0.51	13.62	98.66	0	0.04	0	0.98
D01	299	151.12	0.51	12.31	98.32	0	0.09	0	0.87
D02	362	182.02	0.50	9.17	98.34	0	0.01	0	0.98
D03	30	31.01	1.03	9.47	93.10	0	0.04	0	0.94
D04	41	43.29	1.06	14.69	95.00	0	0.05	0	0.94
D05	37	25.31	0.68	3.32	100.00	0	0.03	0	0.60
D06	650	116.77	0.18	5.91	99.69	0	0.04	0	0.88
D07	347	158.39	0.46	13.43	99.13	0	0.05	0	0.90
D08	77	125.17	1.63	12.71	92.11	0	0	0	0.99
D09	87	107.51	1.24	17.82	94.19	0	0.04	0	0.99
D10	167	175.36	1.05	13.17	96.39	0	0.12	0	1.00
D11	36	53.37	1.48	4.85	100.00	0	0.28	0	0.95
D12	164	151.24	0.92	12.37	95.09	0	0.09	0	0.94
D13	123	126.80	1.03	14.34	95.08	0	0.09	0	0.99
Total	4,607	3,063.40	0.90	18.85	96.4	5	—	—	—

QTL Fine Mapping Of Plant Height

A total of 16 QTLs related to plant height were distributed on nine chromosomes. Among the nine chromosomes, there are six QTLs located on ChrA13, two QTLs located on ChrD01 and ChrD05, and only one QTL located on ChrA03, ChrD04, ChrD06, ChrD07, ChrD08 and ChrD09. Two QTLs were detected in the F₂ population and the remaining 14 QTLs were detected in the F_{2:3}

progeny. The genetic distance of the 16 QTLs ranged from 0 to 3.13 cM, while the physical distance of the alignment to the reference genome ranged from 0 to 7.72 Mb. The 16 QTLs explained 2.07–19.04% of the phenotypic variation. Of the 16 QTLs, three QTLs, qPH-D05-1, qPH-D05-2, and qPH-D06-1, explained more than 10% of the phenotypic variation. Their LOD scores were all greater than five, 7.50, 5.37 and 5.13, respectively. Since the additive effects of qPH-D01-1, qPH-A01-1, and qPH-D09-1 are negative, they have a negative additive effect, and their favorable alleles are derived from the male parent, CCRI49, which reduced plant height. The other QTLs had positive additive effects, with favorable alleles from the female parent, cultivar 396289, which had the effect of increasing plant height. qPH-A03-1, qPH-D01-1, qPH-A01-1, qPH-D04-1, qPH-D06-1, and qPH-D09-1 all had a negative dominant effect, and all other QTLs had a positive effect. The negative dominance effect of qPH-D01-1 was the highest, at -7.62, while the positive dominance effect of qPH-D07-1 was 4.71. The effect of the QTL was judged according to the ratio of dominant effect to additive effect: qPH-D04-1 had a negative super-dominant effect, qPH-A03-1 and qPH-D06-1 had a negative partial dominant effect, qPH-D05-1 and qPH-D05-2 had a partially positive dominant effect, qPH-A13-4 and qPH-A13-5 had a positive dominant effect, while the rest of the QTLs had a positive super-dominant effect (Table 4).

Table 4
Details of QTLs affecting the plant height of upland cotton.

Env.	QTL	Chr.ID	Genetic Position(cM)		Physical Position		LOD	ADD	DOM	S(D/A)	PV (%)
			Start	End	Start	End					
16AY	qPH-A03-1	A03	104.920	105.545	73,850,187	75,320,385	2.36	4.13	-2.39	-0.58	5.62
	qPH-D01-1	D01	126.454	126.454	55,808,353	55,808,532	3.08	-0.05	-7.62	152.47	8.30
17AY	qPH-D05-1	D05	0.000	0.312	4,447,784	12,169,784	7.50	5.46	1.75	0.32	19.04
17AQ	qPH-A13-1	A13	65.117	65.117	47,857,367	48,168,698	2.38	2.03	2.69	1.33	3.10
	qPH-A13-2	A13	71.055	72.617	48,800,047	50,224,986	2.49	1.82	2.94	1.61	3.03
	qPH-A13-3	A13	74.619	75.059	48,335,655	49,997,642	2.13	1.79	2.38	1.33	2.41
	qPH-A13-4	A13	81.564	83.254	49,774,062	49,985,932	2.33	2.36	2.67	1.13	3.67
	qPH-A13-5	A13	88.070	89.945	49,415,357	53,839,609	2.51	2.39	2.45	1.03	3.49
	qPH-A13-6	A13	99.008	99.008	54,961,845	55,060,905	2.23	1.43	2.53	1.77	2.07
	qPH-D05-2	D05	1.563	2.002	59,560,690	60,650,745	5.37	5.22	2.08	0.40	11.57
17ALE	qPH-A01-1	A01	85.908	86.533	90,293,490	90,293,512	3.25	-1.42	-4.14	2.90	6.71
	qPH-D01-2	D01	143.464	143.464	60,197,342	60,197,550	2.88	1.35	4.41	3.28	7.25
	qPH-D04-1	D04	42.044	43.294	50,105,284	50,110,017	2.14	1.09	-4.67	-4.28	6.67
	qPH-D06-1	D06	99.046	99.359	51,905,395	51,952,666	5.13	4.16	-1.55	-0.37	10.92
	qPH-D07-1	D07	136.393	138.083	51,718,028	51,718,233	2.31	1.62	4.71	2.91	8.70

Env Envirment, ADD Additive effect, DOM Dominance effect, S(D/A) Degree of dominance, PV Phenotypic variation

Env.	QTL	Chr.ID	Genetic Position(cM)		Physical Position		LOD	ADD	DOM	S(D/A)	PV (%)
			Start	End	Start	End					
	qPH-D09-1	D09	56.419	59.544	34,262,910	35,026,327	2.64	-1.19	-4.85	4.08	8.17

Env Environment, *ADD* Additive effect, *DOM* Dominance effect, *S(D/A)* Degree of dominance, *PV* Phenotypic variation

Table 4 Details of QTLs affecting the plant height of upland cotton.

To further refine the map of QTLs controlling plant height, a Mutmap-like strategy was used to identify SLAF-tagged polymorphisms. Mutmap was developed on the basis of bulked separation analysis (BSA) to determine the contribution of alleles to the tall mutant by calculating parameters (SNP-index) (Takagi et al. 2015). Here, we considered plant height a mutant trait by calculating the Δ (SLAF-index) for each SLAF marker on the genome. The figure shows the relationship between Δ (SLAF index) and the SLAF position (Fig. 4). Two significant peaks were screened while the Δ (SLAF index) threshold was set to 0.6, and the QTL was located on ChrA07 and ChrD05, respectively (Table S4). A significant peak was observed on ChrD05 with six markers in the region (marker 561423, marker 561425, marker 562264, marker 562265, marker 562458, and marker 562459). This was considered a tight linkage marker for plant height, ranging from 0.312 cM to 0.312 cM on ChrD05 (Fig. 5A). These results strongly suggest that the major gene controlling plant height is on ChrD05. Based on the QTL mapping results, we narrowed the QTL controlling plant height to a fragment of essentially 0, and its physical distance on the genome was shortened from 7.72 Mb to 0.72 Mb. In addition, using the Mutmap-like strategy, the region not mapped by the QTL was also obtained. On ChrA07, the genome has a physical distance of 0.1 Mb in this region.

Functional Annotation Of Candidate Genes

In total, 1,028 candidate genes were obtained in the confidence intervals of the QTLs for plant height. The number of candidate genes in the QTL confidence intervals ranged from 1 (qPH-D06-1) to 839 (qPH-D05-1), except for the seven QTLs (qPH-D01-1, qPH-D01-2, qPH-D07-1, qPH-D04-1, qPH-A01-1, and qPH-A13-4) that have no candidate genes in the confidence intervals. We annotated 1,012 out of 1,028 candidate genes. In the COG classification, a total of 365 genes were annotated, of which 105 genes were involved in general function prediction only (R), 67 genes were involved in transcription (K), 62 genes were involved in replication, recombination and repair (L), and 62 genes were involved in signal transduction mechanisms (T). Among these genes, 45 genes have four functions of RTKL at the same time, and seven genes have four functions of GEPR (Table S5). In addition, no genes involved in cell motility (N), extracellular structures (W), and nuclear structures were found (Y) (Fig. 6A). In the GO analysis, a total of 628 genes were annotated; 1,270 genes were identified as cellular components, 765 genes were identified as molecular functions, 1,739 genes were identified as biological processes, and some had multiple functions and were included into two or three categories. In the cellular component classification, there are 285 genes related to the cell, 271 genes related to cell parts, and 203 genes related to organelles. In the molecular function classification, 317 genes are related to catalytic activity, 287 genes are related to binding, and 71 genes are related to transporter activity. In the biological processes classification, 421 genes are related to metabolic processes, 377 genes are related to cellular processes, and 288 genes are related to single-organism processes (Fig. 6B, Table S6). In the KEGG analysis, 214 genes were identified in 97 pathways. Among them, 14 genes were related to plant hormone signal transduction pathways, 12 genes were related to protein processing in the endoplasmic reticulum pathway, and 12 genes were related to ribosome pathways (Table S7).

A total of 128 candidate genes were obtained by Mutmap-like correlation analysis, including 122 on Chr.D05 and six on Chr.A07 (Table S4, Table S8). 120 of 122 candidate genes were annotated. In comparison with the mapping results of the QTLs of the genetic map, the results were mapped to 18 candidate genes that may be related to plant height at 0.312 cM on ChrD05 (Fig. 5B). Among the six candidate genes obtained from ChrA07, three candidate genes (Gh_A07G1465, Gh_A07G1466, and Gh_A07G1468) were annotated in the COG classification, which were respectively assigned to coenzyme transport and metabolism [H], intracellular trafficking, and secretion. Among the three categories of GO classification, five candidate genes (Gh_A07G1464, Gh_A07G1465, Gh_A07G1466, Gh_A07G1467, and Gh_A07G1468) were annotated in the GO classification, of

which Gh_A07G1467 was annotated as having cellular component functions. Among the three categories of vesicular transport [U] and general function prediction only [R] Gh_A07G1464 and Gh_A07G1466 are mainly classified into two groups: cell component and biological process; Gh_A07G1465 and Gh_A07G1468 have three kinds of cell components, biological processes, and molecular functions. In the KEGG pathway analysis, Gh_A07G1465 is involved in the pathway of pantothenate and CoA biosynthesis and beta-alanine metabolism, Gh_A07G1466 is involved in SNARE interactions in the vesicular transport pathway, and Gh_A07G1468 is involved in the development of eukaryotes.

3.7 Identification of candidate genes related to plant height

According to the results of QTL fine mapping, 18 candidate genes were subjected to qRT-PCR (Fig. 7). The results of qRT-PCR analysis showed that the expression level of candidate genes Gh-D05G1321 and Gh-D05G1335 in cultivar 396289 was significantly lower than their expression in cultivar CCRI49. However, the expression levels of Gh-D05G1315, Gh-D05G1334, Gh-D05G1339, and Gh-D05G1342 in cultivar 396289 and CCRI49 were not significantly different. The expression levels of the 12 remaining candidate genes in cultivar 396289 were higher than their expression levels in CCRI49, especially Gh-D05G1297, Gh-D05G1310, and Gh-D05G1351, where the expression level in 396289 was more than twice that of CCRI49.

Discussion

Characteristics of Re-seq and SLAF-seq strategy for large-scale marker development

With the vast amount of plant genomic data published, whole genome resequencing and simplified genome sequencing techniques have been applied widely in different species (Causse et al. 2013; Takagi et al. 2015; Liang et al. 2016). Sequence variation of SNPs is commonly used in plant genetic analysis due to its richness and universality in the genome (Wang et al. 2015b). High-throughput next-generation sequencing (NGS) technology enables rapid and cost-effective access to massive amounts of SNP markers. A large number of SNPs can be detected comprehensively and accurately using genome-wide sequencing technology, but the sequencing cost is higher. Simplified genome sequencing can reduce the complexity of the genome by using the enzymatic capture technique obtaining the target data and can greatly reduce the experimental cost, but the number of labeled SNPs obtained is far fewer than re-sequencing techniques. Compared with traditional simplified genome sequencing, SLAF-seq has the ability to develop personalized enzyme digestion schemes based on different reference genomes, and simultaneously exploits dual-index sequencing to develop high-quality markers and dynamic suspicious markers to obtain the filtering process (Zhang et al. 2016c; Sun et al. 2013; Liu et al. 2014). Zhang et al. (Zhang et al. 2016c) developed the first high-density genetic map of upland cotton and sequenced the RIL population using SLAF-seq, obtaining 443.65 Mb of clean reads. Qi et al. [6] used the GBS-seq method to obtain a total of 908 Mb of clean reads for an F₂ population of upland cotton. In this study, a total of 1,448 Mb clean reads were obtained for an F₂ population of upland cotton using a combination of re-sequencing and SLAF sequencing, demonstrating the robust data development capabilities of this strategy. This economical and viable method of obtaining mass data is the basis for developing a large number of polymorphic markers, ensuring the accuracy of genetic map construction. This is also the first successful application of this strategy in an F₂ population of upland cotton.

Characteristics Of Genetic Map Construction

Compared with traditional RFLP, RAPD, AFLP, and SSR markers, more SNP polymorphism markers can be obtained by using high-throughput sequencing. Sun et al. (Sun et al. 2012) constructed a genetic map of upland cotton using SSR markers, including 34 linkage groups with an average distance of 3.69 cM and a genetic linkage map of 700.9 cM. Because of the narrowed genetic background about *G. hirsutum* (Rungis et al. 2005), SSR markers are still commonly used to construct population maps. However, the number of markers obtained by SSR markers is still not ideal. Chen et al. (Chen et al. 2015) detected 2,763 markers in 26 linkage groups (chromosomes) in an F₂ population (*G. hirsutum* × *G. darwinii*), covering the genome length of 4,176.7 cM and the average distance between genes was 1.5 cM. In the present study, 4,607 upland cotton genotypes were obtained using high-throughput sequencing. The total genetic distance was 3,063 cM, with an average distance

of 0.898 cM. Compared with the previously produced genetic maps (Chen et al. 2015; Sun et al. 2012), our study obtained a larger number of markers and the genetic distance of the genetic map was finer. Producing genetic maps from high-throughput sequencing data is superior to other techniques.

QTL Mapping For Plant Height

Previous studies on QTL mapping of upland cotton have focused on fiber quality traits (Sun et al. 2012; Zhang et al. 2009; Tang et al. 2015; Nie et al. 2016) and yield traits (Wang et al. 2015a; Cong et al. 2016; Liu et al. 2017b), with little focus on morphological traits. Plant height is an important morphological trait that effects the economic output of cotton as well as the mechanization of cotton harvesting. Eighty-nine QTLs related to the plant height of upland cotton can be found in the QTL database (<http://www2.cottonqtl.org:8081/index>), and 75 QTLs were obtained after removing 14 repetitive QTLs. These QTLs are distributed on 26 chromosomes except Chr04, Chr12, and Chr18. The LOD values of these QTLs ranged from 1.41 to 18.61, accounting for 0.05–73.9% of the phenotypic variation. The confidence intervals of four of the QTLs could not be obtained, but the remaining QTLs had confidence intervals of 0–48 cM. In addition to the QTL database, Qi et al. (Qi et al. 2017) detected seven QTLs related to plants in a single environment using an $F_{2:3}$ family of upland cotton; one QTLs was located on Chr05, and the others were located on Chr25. These QTLs had a confidence interval of 0.7–15 cM and explained 3.49–21.51% of the phenotypic variation. Li et al. (Li et al. 2017) detected five QTLs related to plant height in a single environment in an F_2 population of upland cotton. Those QTLs were distributed on four chromosomes, ChrA01, ChrA08, ChrD06, and ChrD08, which can explain 4.92–10.75% of the phenotypic variation. In the present study, 16 QTLs were detected in a total of four environments from F_2 and $F_{2:3}$ populations, explaining 2.07–19.04% of the phenotypic variation. In addition, QTLs related to plant height were isolated to ChrD05. This result filled gaps in the QTL database and added to the work previously published by Qi et al. (Qi et al. 2017) and Li et al. (Li et al. 2017). The QTLs in this study have a confidence interval of 0–3.13 cM, which greatly reduces the confidence intervals of QTLs compared with previous studies and is conducive to follow-on functional verification and molecular marker-assisted breeding. Because of heterosis studies are rarely involved in the construction of genetic maps and QTL mapping using high-throughput sequencing, this study provided theoretical support creatively for the study of heterosis by studying the additive effect, dominant effect, and degree of dominance of QTLs. In addition, using the Mutmap-like strategy, plant height-related fragments were obtained at ChrA07 and ChrD05. Through the combination of QTL mapping with a genetic map, we obtained a fine mapping of ChrD05 and identified nine candidate genes related to plant height. Compared with previous studies (Qi et al. 2017; Li et al. 2017; Zhang et al. 2016c), there are some innovations in the analysis methods, which provide a case for the fine mapping of QTLs.

Functional Annotation Analysis Of Candidate Genes

714 candidate genes were annotated using at least one of the three databases (COG, GO, and KEGG), and some genes may be related to plant height. In *Arabidopsis* (Wang and Li 2008) and rice (Sun et al. 2010), the biosynthesis and signaling pathways of some hormones (e.g., auxin and brassinolide) regulate plant height. In this study, the COG analysis found more genes involved in general function prediction only [R], transcription [K], replication, recombination and repair [L], signal transduction mechanisms [T], and posttranslational modification, protein turnover, and chaperones [O], which are related to hormone biosynthesis and signal transduction pathways, and may affect the height of cotton. Qi et al. (Qi et al. 2017) found that cytochrome P450 is involved in plant height, and in this study we found that 11 genes (Gh_D05G0646, Gh_D05G0686, Gh_D05G0830, Gh_D05G0879, Gh_D05G0951, Gh_D05G1038, Gh_D05G1098, Gh_D05G1281, Gh_A13G0977, Gh_A13G0992, and Gh_D09G0920) were associated with cytochrome P450. In the GO analysis, more genes were involved in auxin and brassinosteroid related processes, such as Gh_D05G0636, Gh_D05G0661, Gh_D05G1056, and Gh_D05G1254, which are involved in the response to auxin (GO: 0009733). Specifically, Gh_D05G0636 is involved in the auxin metabolic process (GO: 0009850), Gh_D05G1123, and Gh_D05G1357 are associated with auxin homeostasis (GO: 0010252), Gh_D05G0606 is associated with basipetal auxin transport (GO: 0010540), Gh_D05G0837, Gh_D05G0755, Gh_D05G1216, and Gh_D05G0765 are related to the auxin-activated signaling pathway (GO: 0009734), and Gh_D09G0919 is related to auxin polar transport (GO: 0009926). Gh_D05G1281 and Gh_D05G0615 participated in the process of brassinosteroid biosynthetic processes (GO: 0016132) and Gh_D05G1055 is involved in response to brassinosteroids (GO: 0009741). In addition, Silvia et al. (Fornalé et al. 2010) found that lignin impacts the composition of cell polysaccharides and the composition of the cell wall, thereby affecting the height of maize. In the

present study, the lignin biosynthetic process (GO: 0009809) and the lignin catabolic process (GO: 0046274) were Gh_D05G1123 and Gh_D05G0888, respectively. These genes may be related to the plant height of cotton. In the KEGG pathway analysis, the three pathways that contain the most genes are hormonal signal transduction, the processing of proteins in the endoplasmic reticulum, and ribosomal pathways. The results of this study are similar to the results described in Zhang et al. (Zhang et al. 2016c), which suggested that ribosomes are related to protein synthesis and plant hormones, which can in turn affect plant height.

The Mutmap-like correlation analysis was used to fine map the QTLs of the candidate region on ChrD05, and 18 candidate genes related to plant height were obtained. The results of qRT-PCR showed that there was no significant difference in the relative expression level of four genes between the two parents, and the relative expression levels of 14 candidate genes were significantly different between the two parents. In future studies, these 14 genes can be used for functional verification.

Deficiencies In The Genetic Map

The strategy of parental re-sequencing can help find more polymorphic sites and contribute to the construction of high-density genetic maps. However, according to the results of genetic map construction in this study, chromosomes A06, D03, D04, and D05 were found to be short in length. By examining the physical length of each chromosome in the *G. hirsutum* genome (Zhang et al. 2015b) and counting the SNP markers of each chromosome during development of the marker [Table S9], we found that: 1) the genome length of chromosomes D03 and D04 was shorter than the other chromosomes; 2) parents had fewer polymorphic SNPs, polymorphisms, and homozygous SNPs in the A06, D03, and D04 chromosomes. Therefore, the reason for the short A06, D03, and D04 chromosomes may be related to the reference genome or may be more related to the parental conserved regions on these chromosomes. Qi et al. (Qi et al. 2017) also found that ChrD05 is short, which may be related to the mapping population used, and requires further investigation.

Conclusions

In this study, an F₂ population of upland cotton was used to construct a high-density genetic map by parental re-sequencing and progeny SLAF sequencing. Finally, 4,607 of upmapping markers were obtained with a total genetic distance of 3,063.4 cM. The average genetic distance of markers was 0.898 cM. Sixteen QTLs related to plant height were obtained from upland cotton grown in four environments, and the QTLs were genetically analyzed and functional annotations were identified for candidate genes within confidence intervals. In addition, the fine mapping of QTLs on ChrD05 by the Mutmap-like strategy greatly reduced the genetic distance and physical distance of the QTLs. This study not only enriched the genetic map of upland cotton, which provided theoretical support for molecular marker-assisted breeding, but also was a successful application of the parental resequencing and progeny SLAF-seq strategy in the F₂ population of upland cotton. At the same time, the combination of multiple analytical methods provided a solution for the fine mapping of QTLs. A total of 18 candidate genes were obtained through fine mapping, and we obtained 14 candidate genes that may be related to plant height.

Abbreviations

SLAF-seq: Specific Locus Amplified Fragment sequencing; **QTLs**: Quantitative trait locus; **SNP**: single nucleotide polymorphisms; **PH**: plant height; **SDM**: segregation distortion markers; **qRT-PCR**: quantitative real-time polymerase chain reaction.

Declarations

Availability of data and materials

Review of data is available and there are no competing interests. All data presented within is the corresponding authors' data and is available upon request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have declared that no competing interests exist

Funding

The special academic-level coordinating work task of the basic scientific research business expenses of the Chinese Academy of Agricultural Sciences "CCIA high-quality cotton industry chain key technology research and demonstration" (Y2020LM01); Chinese Academy of Agricultural Sciences Science and Technology Innovation Project Agricultural Science and Technology go out collaborative innovation and integration demonstration research task "Central Asia Screening of high-quality cotton varieties and integrated R&D and application of supporting cultivation techniques" (CAAS-XTCX2018020-4); "Belt and Road" agricultural resources and technology cooperation project of the Ministry of Agriculture and Rural Affairs (125161016000180005)

Authors' contributions

HQ conceived and designed the experiments. QW performed the experiments and analyzed the data. YG and WN contributed reagents and materials. QW wrote the article. All the authors read and approved the final manuscript.

Acknowledgements

We thank Huang Qun for their guidance and help in the design of this experiment, and all our lab mates and friends for their helpful suggestion and useful comments on the manuscript.

References

Baird NA, Etter PD, Atwood TS, et al. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One*. 2008; 3 (10):e33376. <http://doi.org/10.1371/journal.pone.0003376>.

Broman KW, Sen S (2009) A Guide to QTL Mapping with R/qtl. *Journal of Statistical Software*. 2009; 66 (32):658-659. http://doi.org/10.1007/978-0-387-92125-9_9.

Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *Bmc Bioinformatics*. 2009; 10 (1):421. <https://doi.org/10.1186/1471-2105-10-421>.

Causse M, Desplat N, Pascual L, et al. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *Bmc Genomics*. 2013; 14 (1):791. <http://doi.org/10.1186/1471-2164-14-791>.

Chen H, Khan MK, Zhou Z, et al. A high-density SSR genetic map constructed from a F2 population of *Gossypium hirsutum* and *Gossypium darwinii*. *Gene*. 2015; 574 (2):273-286. <http://doi.org/10.1016/j.gene.2015.08.022>.

Chen Q, Zhang Y. Distorted segregation and construction of a molecular linkage map of SSR markers in indica rice. *Molecular Plant Breeding*. 2009; 7 (4):685-689. <http://doi.org/10.3969/mpb.007.000685>.

Chen ZJ, Scheffler BE, Dennis E, et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiology*. 2007; 145 (4):1303. <http://doi.org/10.1104/pp.107.107672>

Conesa A, Gotz S, Garcia-Gomez J, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)* 2005;21(18):3674 <https://doi.org/10.1093/bioinformatics/bti610>

Cong L, Dong Y, Zhao T, et al. Genome-Wide SNP Linkage Mapping and QTL Analysis for Fiber Quality and Yield Traits in the Upland Cotton Recombinant Inbred Lines Population. *Front Plant Sci.* 2016; 7:1356. <http://doi.org/10.3389/fpls.2016.01356>.

Dass A, Abdin MZ, Reddy VS, et al. Isolation and characterization of the dehydration stress-inducible GhRDL1 promoter from the cultivated upland cotton (*Gossypium hirsutum*). *Journal of Plant Biochemistry and Biotechnology.* 2016; 26 (1):113-119. <http://doi.org/10.1007/s13562-016-0369-3>.

Davey JW, Hohenlohe PA, Etter PD, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics.* 2011; 12 (7):499. <http://doi.org/10.1038/nrg3012>.

Elshire RJ, Glaubitz JC, Sun Q, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One.* 2011; 6 (5):e19379. <https://doi.org/10.1371/journal.pone.0019379>.

Fornalé S, Shi X, Chai C, et al. ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant Journal.* 2010; 64 (4):633. <http://doi.org/10.1111/j.1365-313X.2010.04363.x>.

Guo W, Zhang T, Zhu X, et al. Modified backcross pyramiding breeding with molecular marker-assisted selection and its applications in cotton. *Acta Agronomica Sinica.* 2005; 31:963-970.

Huang X, Feng Q, Qian Q, et al. High-throughput genotyping by whole-genome resequencing. *Genome Research.* 2009; 19 (6):1068. <http://doi.org/10.1101/gr.089516.108>.

Jiang B, Liu W, Xie D, et al. High-density genetic map construction and gene mapping of pericarp color in wax gourd using specific-locus amplified fragment (SLAF) sequencing. *Bmc Genomics.* 2015; 16 (1):1035. <http://doi.org/10.1186/s12864-015-2220-y>.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ Method. *Methods.* 2001; 25 (4):402-408. <http://doi.org/10.1006/meth.2001.1262>.

Li C, Wang X, Dong N, et al. QTL analysis for early-maturing traits in cotton using two upland cotton (*Gossypium hirsutum* L.) crosses. *Breeding Science.* 2013; 63 (2):154. <http://doi.org/10.1270/jsbbs.63.154>.

Li F, Fan G, Lu C, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nature Biotechnology.* 2015; 33 (5):524. <http://doi.org/10.1038/nbt.3208>.

Li F, Fan G, Wang K, et al. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature genetics.* 2014; 46 (6):567. <http://doi.org/10.1038/ng.2987>.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2010; 26 (5):589-595. <http://doi.org/10.1093/bioinformatics/btp698>.

Li L, Zhao S, Su J, et al. High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). *Plos One.* 2017; 12 (8):e0182918. <http://doi.org/10.1371/journal.pone.0182918>.

Liang D, Chen M, Qi X, et al. QTL Mapping by SLAF-seq and Expression Analysis of Candidate Genes for Aphid Resistance in Cucumber. *Front Plant Sci.* 2016; 7:1000. <http://doi.org/10.3389/fpls.2016.01000>.

Liu D, Ma C, Hong W, et al. Construction and Analysis of High-Density Linkage Map Using High-Throughput Sequencing Data. *Plos One.* 2014; 9 (6):e98855. <http://doi.org/10.1371/journal.pone.0098855>.

Liu H, Cao F, Yin T, et al. A Highly Dense Genetic Map for *Ginkgo biloba* Constructed Using Sequence-Based Markers. *Front Plant Sci.* 2017a; 8. <http://doi.org/10.3389/fpls.2017.01041>.

Liu X, Teng Z, Wang J, et al. Enriching an intraspecific genetic map and identifying QTL for fiber quality and yield component traits across multiple environments in Upland cotton (*Gossypium hirsutum* L.). *Molecular genetics and genomics* : MGG. 2017b; 292 (6):1281-1306. <http://doi.org/10.1007/s00438-017-1347-8>.

Luo ZW, Hackett CA, Bradshaw JE, et al. Construction of a genetic linkage map in tetraploid species using molecular markers. *Genetics*. 2001; 157 (3):1369-1385.<http://doi.org/10.1089/109065701750168806>.

Mao D, Yu L, Chen D, et al. Multiple cold resistance loci confer the high cold tolerance adaptation of Dongxiang wild rice (*Oryza rufipogon*) to its high-latitude habitat. *Theor Appl Genet* 128, 1359–1371 (2015). <https://doi.org/10.1007/s00122-015-2511-3>

Mei H, Liu Y, Du Z, et al. High-Density Genetic Map Construction and Gene Mapping of Basal Branching Habit and Flowers per Leaf Axil in Sesame. *Front Plant Sci*. 2017; 8:636. <http://doi.org/10.3389/fpls.2017.00636>.

Mun JH, Chung H, Chung WH, et al. Construction of a reference genetic map of *Raphanus sativus* based on genotyping by whole-genome resequencing. *Theoretical & Applied Genetics*. 2015; 128 (2):259-272. <http://doi.org/10.1007/s00122-014-2426-4>.

Nie X, Huang C, You C, et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *Bmc Genomics*. 2016; 17 (1):352. <http://doi.org/10.1186/s12864-016-2662-x>.

Paterson AH, Brubaker CL, Wendel JF. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter*. 1993; 11 (2):122-127.<http://doi.org/10.1007/BF02670470>.

Qi H, Wang N, Qiao W, et al. Construction of a high-density genetic map using genotyping by sequencing (GBS) for quantitative trait loci (QTL) analysis of three plant morphological traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2017; 213 (4):83. <http://doi.org/10.1007/s10681-017-1867-7>.

Rungis D, Llewellyn D, Dennis ES, et al. Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Australian Journal of Agricultural Research*. 2005; 56 (3):301-307. <http://doi.org/10.1071/AR04190>.

Stuber CW, Edwards MD, Wendel JF. Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits. *Crop Science*. 1987; 27 (4):639-648.<http://doi.org/10.2135/cropsci1987.0011183x002700040006x>.

Sun FD, Zhang JH, Wang SF, et al. QTL mapping for fiber quality traits across multiple generations and environments in upland cotton. *Molecular Breeding*. 2012; 30 (1):569-582. <http://doi.org/10.1007/s11032-011-9645-z>.

Sun X, Liu D, Zhang X, et al. SLAF-seq: An Efficient Method of Large-Scale De Novo SNP Discovery and Genotyping Using High-Throughput Sequencing. *Plos One*. 2013; 8 (3):e58700. <http://doi.org/10.1371/journal.pone.0058700>.

Sun Y, Fan XY, Cao DM, et al. Integration of Brassinosteroid Signal Transduction with the Transcription Network for Plant Growth Regulation in *Arabidopsis*. *Developmental Cell*. 2010; 19(5):765. <http://doi.org/10.1016/j.devcel.2010.10.010>.

Takagi H, Tamiru M, Abe A, et al. MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nature Biotechnology*. 2015; 33 (5):445. <http://doi.org/10.1038/nbt.3188>.

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

Wang B, Lu C, Gou C, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nature Genetics*. 2012; 44 (10):1098-1103<http://doi.org/10.1038/ng.2371>.

Wang H, Huang C, Guo H, et al. QTL Mapping for Fiber and Yield Traits in Upland Cotton under Multiple Environments. *Plos One*. 2015a; 10 (6):e0130742. <https://doi.org/10.1371/journal.pone.0130742>.

Wang S, Chen J, Zhang W, et al. Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *Genome Biology*. 2015b; 16 (1):108. <http://doi.org/10.1186/s13059-015-0678-1>.

Wang Y, Li J. Molecular basis of plant architecture. *Annual Review of Plant Biology*. 2008; 59 (59):253-279.<http://doi.org/10.1109/83.760334>.

Xu F, Wang W, Wang P, et al. A fast and accurate SNP detection algorithm for next-generation sequencing data. *Nature Communications*. 2012; 3 (1):1258.<http://doi.org/10.1038/ncomms2256>.

Xu X, Zeng L, Tao Y, et al. Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110 (33):13469-13474.<http://doi.org/10.1073/pnas.1222368110>.

Zeng ZB. Precision mapping of quantitative trait loci. *Genetics*. 1994; 136 (4):1457-1468. <http://doi.org/10.1007/s00122-012-2032-2>.

Zhang D, Li H, Wang J, et al. High-Density Genetic Mapping Identifies New Major Loci for Tolerance to Low-Phosphorus Stress in Soybean. *Front Plant Sci*. 2016a; 7:372. <http://doi.org/10.3389/fpls.2016.00372>.

Zhang H, Yi H, Wu M, et al. Mapping the Flavor Contributing Traits on "Fengwei Melon" (*Cucumis melo* L.) Chromosomes Using Parent Resequencing and Super Bulked-Segregant Analysis. *Plos One*. 2016b; 11 (2):e0148150. <http://doi.org/10.1371/journal.pone.0148150>.

Zhang J, Zhang Q, Cheng T, et al. High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). *Dna Research*. 2015a; 22 (3):183. <http://doi.org/10.1093/dnares/dsv003>

Zhang T, Hu Y, Jiang W, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nature Biotechnology*. 2015b; 33 (5):531. <http://doi.org/10.1038/nbt.3207>.

Zhang Z, Shang H, Shi Y, et al. Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (*Gossypium hirsutum*.). *Bmc Plant Biology*. 2016c; 16 (1):79. <http://doi.org/10.1186/s12870-016-0741-4>.

Zhang ZS, Hu MC, Zhang J, et al. Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding*. 2009; 24 (1):49-61. <http://doi.org/10.1111/j.1600-0722.2008.00560.x>.

Figures

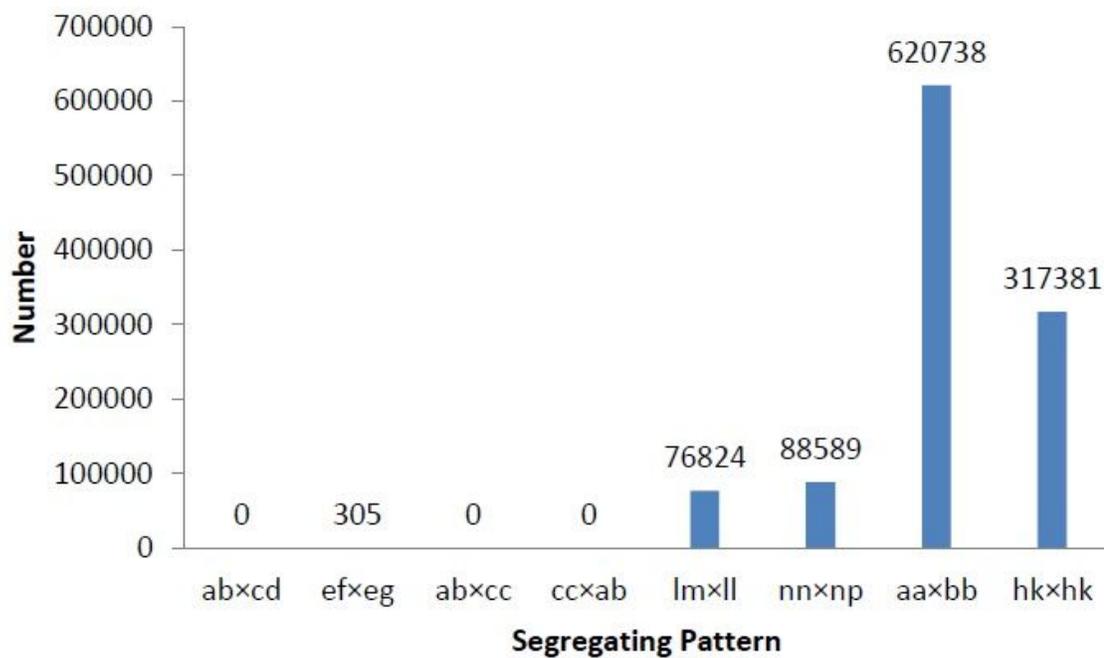


Figure 1

The number of SNPs in eight segregation patterns. The x-axis indicates eight segregation patterns grouped by the genotype encoding rule, and the y-axis represents the number of SNPs.

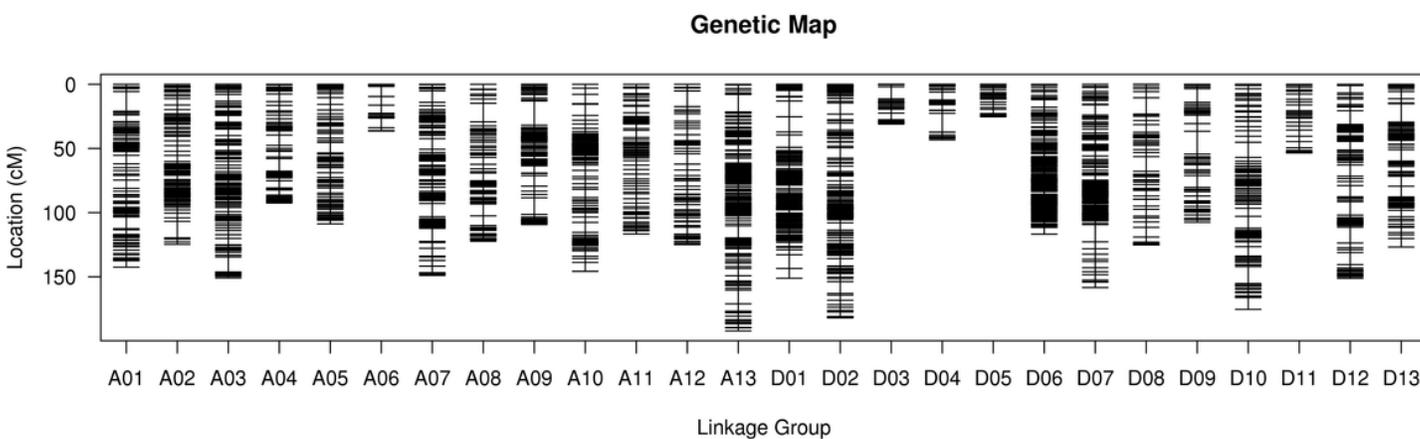


Figure 2

The genetic map constructed from the obtained SNPs. The black bar indicates a marker. The x-axis indicates linkage group (LG) and the y-axis represents genetic distance (cM as the unit).

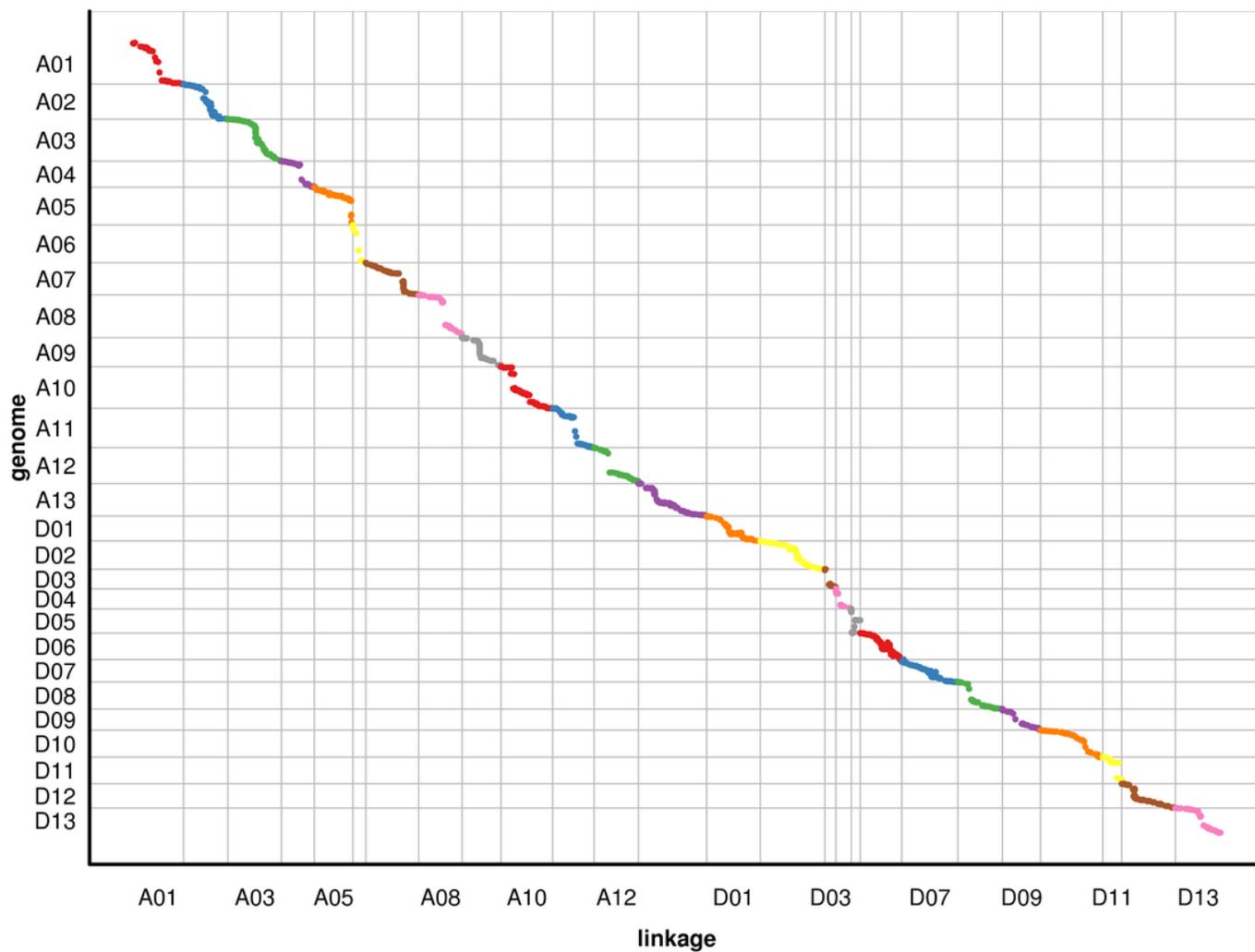


Figure 3

Collinearity analysis of 26 chromosomes of upland cotton. The x-axis represents the genetic distance of 26 chromosomes and the y-axis represents the physical distance of 26 chromosomes of the reference genome.

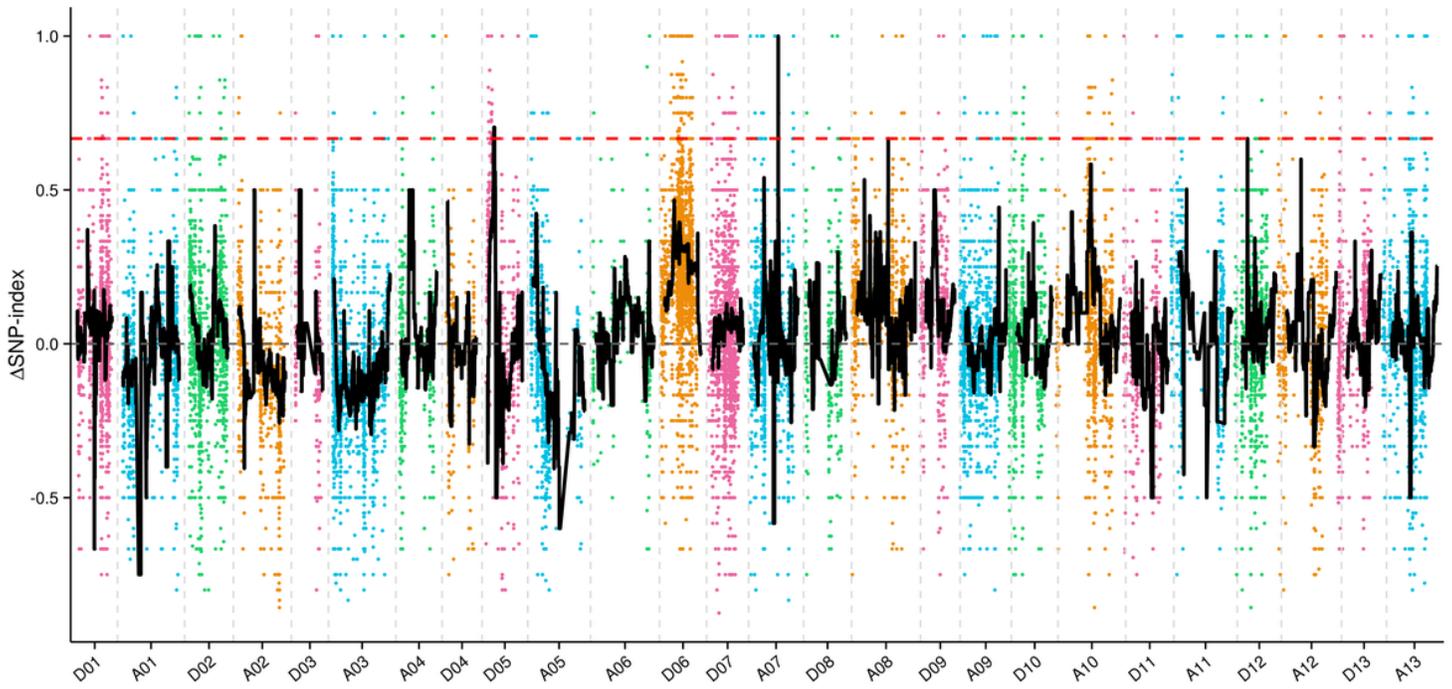


Figure 4

Mutmap-like analysis of plant height in upland cotton. The x-axis represents the position of 26 chromosomes, whereas the y-axis shows the value of Δ (SLAF-index). The dashed line represents the threshold value.

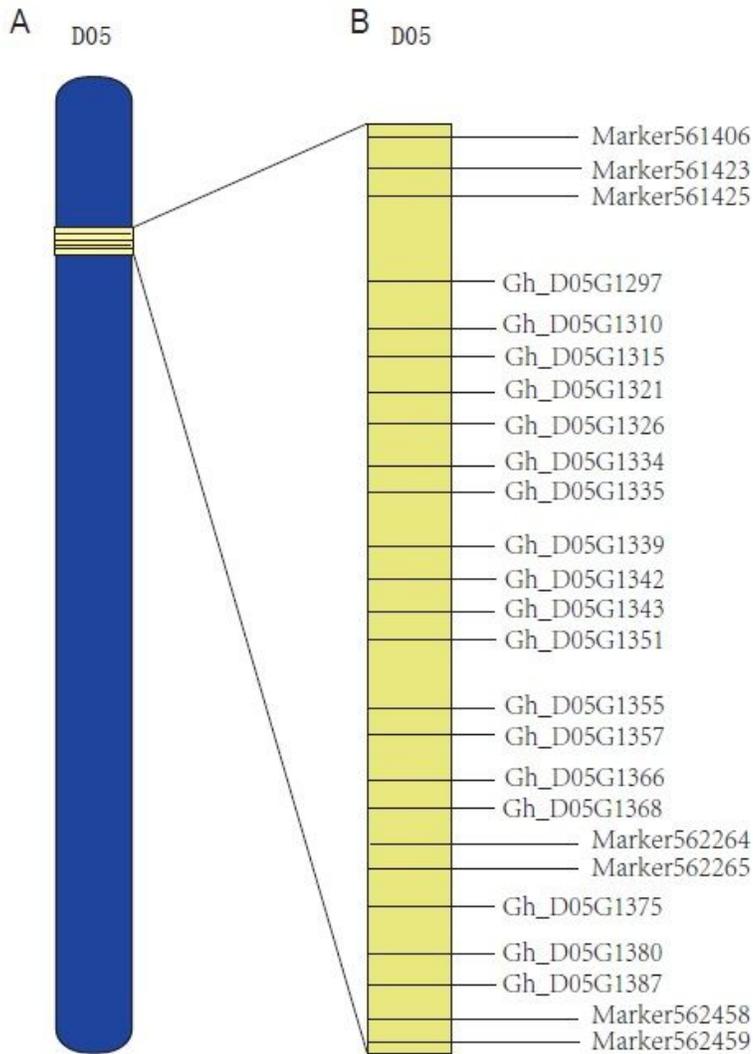


Figure 5

The distribution of tightly linked markers and candidate genes of plant height on ChrD05. (A) Candidate genomic region on ChrD05. (B) Schematic diagram of the order of the seven tightly linked markers and the 18 candidate genes from plant height based on gene annotation.

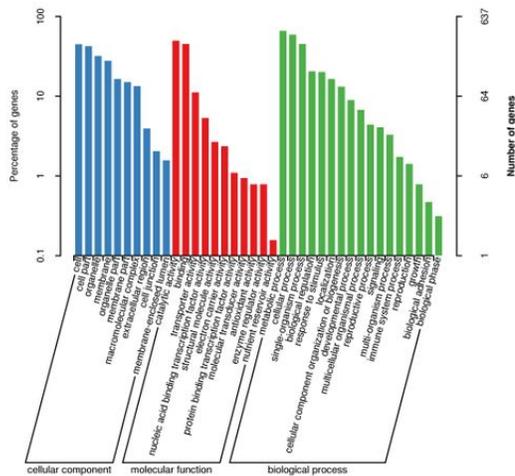
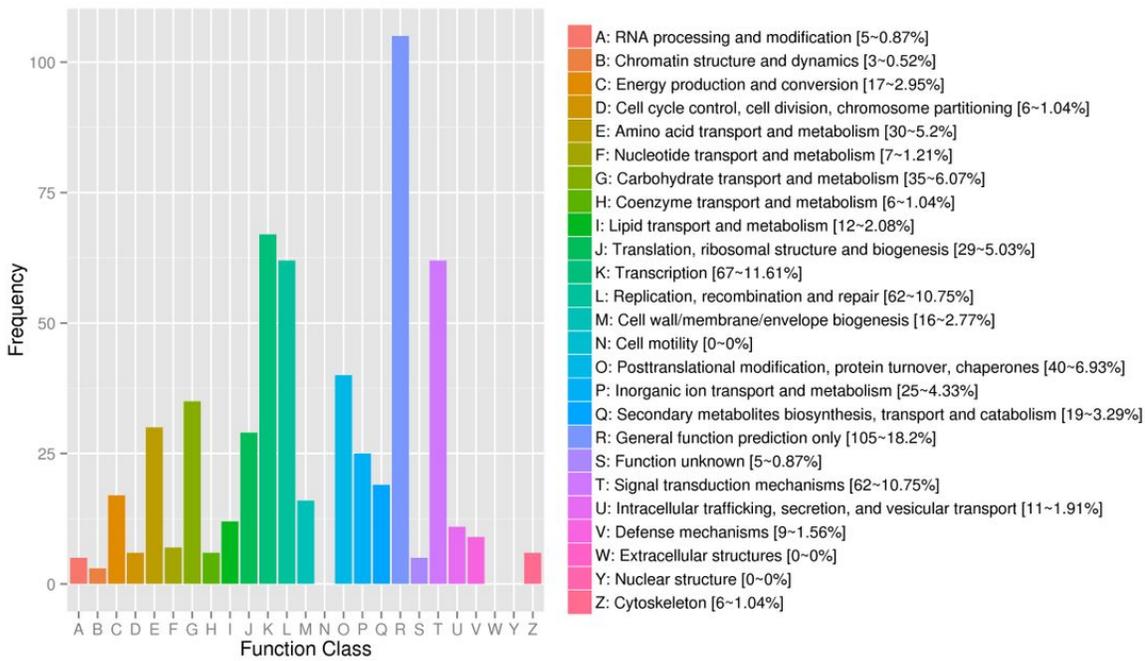


Figure 6

Annotation of the candidate genes in the confidence intervals of QTLs. (A) The annotation of candidate genes in the confidence intervals of QTLs through COG analysis. (B) The annotation of candidate genes in the confidence intervals of QTLs through GO analysis.

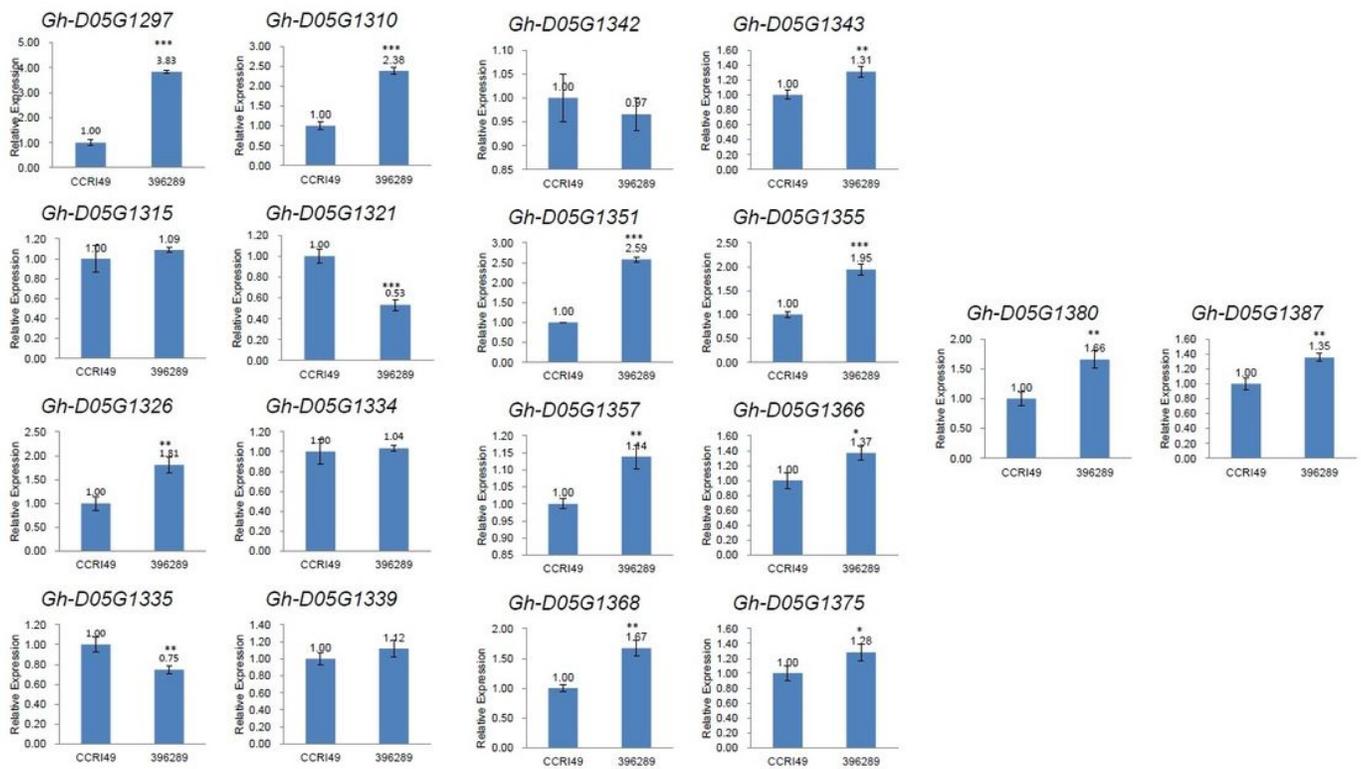


Figure 7

Expression of the genes by qRT-PCR. In these figures, the x-axis represents cultivars CCRI49 and 396289, while the y-axis shows the relative expression level of the gene and the expression level of 396289 relative to CCRI49. * indicate different levels of significance (Student's t-test), *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.