Association Mapping for Epistasis and Environmental Interaction of Yield Traits in 323 Cotton Cultivars under 9 Different Environments

Yinhua Jia1*, Xiwei Sun2*, Junling Sun1*, Zhaoe Pan1, Xiwen Wang1, Shoupu He1, Songhua Xiao3, Weijun Shi4, Zhongli Zhou1, Baoyin Pang1, Liru Wang1, Jianguang Liu3, Jun Ma4, Xiongming Du1*, Jun Zhu2*

1 Institute of Cotton Research of Chinese Academy of Agricultural Sciences (ICR, CAAS), State Key Laboratory of Cotton Biology, Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang, China, 2 Key Laboratory of Crop Germplasm Resource of Zhejiang Province, Zhejiang University, Hangzhou, China, 3 Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, 4 Economic Crop Research Institute, Xinjiang Academy of Agricultural Science, Urumqi1, China

Abstract

Improving yield is a major objective for cotton breeding schemes, and lint yield and its three component traits (boll number, boll weight, and lint percentage) are complex traits controlled by multiple genes and various environments. Association mapping was performed to detect markers associated with these four traits using 651 simple sequence repeats (SSRs). A mixed linear model including epistasis and environmental interaction was used to screen the loci associated with these four yield traits by 323 accessions of Gossypium hirsutum L. evaluated in nine different environments. 251 significant loci were detected to be associated with lint yield and its three components, including 69 loci with individual effects and all involved in epistasis interactions. These significant loci explain ~ 62.05% of the phenotypic variance (ranging from 49.06% ~ 72.29% for these four traits). It was indicated by high contribution of the genetic environment interaction to the phenotypic variance for lint yield and boll numbers, that genetic effects of SSR loci were susceptible to environment factors. Shared loci were also observed among these four traits, which may be used for simultaneous improvement in cotton breeding for yield traits. Furthermore, consistent and elite loci were screened with -Log10 (P-value) > 8.0 based on predicted effects of loci detected in different environments. There was one locus and 6 pairs of epistasis for lint yield, 4 loci and 10 epistasis for boll number, 15 loci and 2 epistasis for boll weight, and 2 loci and 5 epistasis for lint percentage, respectively. These results provided insights into the genetic basis of lint yield and its components and may be useful for marker-assisted breeding to improve cotton production.

Citation: Jia Y, Sun X, Sun J, Pan Z, Wang X, et al. (2014) Association Mapping for Epistasis and Environmental Interaction of Yield Traits in 323 Cotton Cultivars under 9 Different Environments. PLoS ONE 9(5): e95882. doi:10.1371/journal.pone.0095882

Editor: Baohong Zhang, East Carolina University, United States of America

Received October 31, 2013; Accepted March 31, 2014; Published May 8, 2014

Copyright: © 2014 Jia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work has been funded by the National Science and Technology Support Program of China (2006BAD13B04), and National Basic Research Program of China (2010CB12600, 2011CB109306). The funders had no rule in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: jzhu@zju.edu.cn (JZ); duxm@cricaas.com.cn (XDI)

† These authors contributed equally to this work.

Introduction

Cotton is one of the most important economic crops in the world and provides large amounts of raw materials for textile industry. High yield is always the primary focus of cotton breeding programs. Boll number, boll weight, and lint percentage are three yield component traits, which are complex traits controlled by multiple genes with gene x gene interaction and gene x environment interaction [1–2]. Understanding genetic architecture of lint yield and its component traits at the molecular level is very important for efficiently improving cotton yield traits. Identifying and charactering the key QTLs and genes affecting yield traits in cotton has been a research frontier over the past few decades. On the basis of linkage analysis in the QTL mapping populations, several QTLs controlling lint yield and its component traits have been obtained [1–8]. However, these detected QTLs involved in large regions of the chromosome, where candidate genes or genetic variants are difficult to distinguish. In addition, the mapping results have only limited utilization because of lower genetic diversity in the special population used for QTL mapping and can not fully elucidate the genetic basis of lint yield and its component traits.

Genome-wide association studies (GWAS), which are based on the linkage disequilibrium (LD), provide the opportunity to systematically identify the genetic components of complex traits of crops in recent years [9–14]. It has the potential to detect single nucleotide polymorphisms (SNPs) or other types of molecular markers such as single sequence repeats (SSRs) within or nearby a gene. It can extensively exploit historical recombination and natural genetic diversity in natural population, and overcome the limitations of conventional linkage mapping for crop breeding. Recently, Zhang et al performed association mapping using 81
Plant Materials

Materials and Methods

However, current GWAS predominately focus on examination of with additive effects significantly associated with the traits [14]. Due to a heavy computational burden, current methods of GWAS have not been used for detecting gene × gene interaction and gene × environment interaction for breeding populations in multiple environments. In this study, we used mixed linear model approach and a newly developed mapping software of QTNNet- work based on GPU parallel computation to identify individual and epistasis loci and their environmental interactions for lint yield and its component traits. We also identified candidate genes related to markers of these target traits. These results and genetic architecture identified in this study could provide more precise and reliable information for marker-assisted selection in crop breeding.

DNA Extraction and Marker Analysis

The genomic DNA was extracted from fresh, young leaves of the 323 cultivars using CTAB method. SSR primers sequences including BNL, JESPw, TMB, CIR, HAU, DPL, GH, NAU, MGHES etc. can be available from Cotton Microsatellite Database (CMD, http://www.cottonssr.org). We selected 20 cultivars of large morphological variations from the research natural population to screen for polymorphisms primers. In total, 5,600 SSR primers were chosen to survey the polymorphisms of the 323 cultivars. Only 651 polymorphisms SSR alleles were detected, which were used for association mapping (Table S2).

Genetic Models and Statistical Methods

Association mapping were performed using the mixed linear model, including environment (e) as fixed effects, SSR loci effects (a, aa) and loci by environment interaction (ae, aae) as random effects. The genetic model for phenotypic value of the k-th genotypes in the h-th environment (yhk) can be expressed by the following mixed linear model,

\[
y_{hk} = \mu + a_h + \sum_{i=1}^{4} a_i u_{ih} + \sum_{i<j} a_{ij} u_{ih} u_{jk} + \sum_{i} a_{eih} u_{ih} + \varepsilon_{hk}
\]

where \( \mu \) is the population mean; \( a_h \) is the fixed effect of the h-th environment; \( a_i \) is the i-th additive effect with coefficient \( u_{ih} \); \( a_{ij} \) is the i-th additive by j-th additive epistasis effect with coefficient \( u_{ih} u_{jk} \); \( a_{eih} \) is the i-th additive by the h-th environment interaction effect with coefficient \( u_{ih} \); \( a_{eih} \) is the fixed effect of the h-th environment interaction effect with coefficient \( u_{ih} u_{jk} \); and \( \varepsilon_{hk} \) is the random residual effect of the k-th breeding line in the h-th environment.

The mixed linear model can be presented in matrix notation,

\[
y = Xb + U_{AA} e_A + U_{AE} e_{AE} + U_{AAE} e_{AAE} + e_c
\]

\[
\sim MVN(Xb, \sum_{i=1}^{4} \sigma_i^2 U_i U_i^T + I\sigma_e^2)
\]

where \( y \) is an \( n \times 1 \) column vector of phenotypic values and \( n \) is the sample size of observations; \( b \) is a column vector of \( u \) and environments in the experiment; \( X \) is the known incidence matrix relating to the fixed effects; \( U_i \) is the known coefficient matrix relating to the \( i \)-th random vector \( e_i \); \( e_c \sim MVN(0, I\sigma_e^2) \) is an \( n \times 1 \) column vector of residual effects.

The phenotypic variance \( V_p \) is considered as the sum of additive variance \( V_A \), epistasis variance \( V_{AA} \), additive by environment interaction variance \( V_{AE} \), epistasis by environment interaction variance \( V_{AAE} \), and residual variance \( V_c \).

\[
V_p = V_A + V_{AA} + V_{AE} + V_{AAE} + V_c
\]

Heritability is defined as the relative contribution of genetic variance to phenotypic variance with the following estimation model:

\[
h_{G+GE}^2 = h_A^2 + h_{AA}^2 + h_{AE}^2 + h_{AAE}^2
\]

\[
= h_A^2 + \sum h_{ai}^2 + \sum h_{ae}^2 + \sum h_{aee}^2
\]
where $h^2_{GE}$ = total heritability; $h^2_G$ = heritability contributed by sum of individual locus, $h^2_A$ = heritability contributed by sum of pair-wise epistasis, $h^2_{AE}$ = additive by environment interaction heritability contributed by sum of individual additive by environment interaction effects, $h^2_{AAE}$ = epistasis by environment interaction heritability contributed by sum of pair-wise epistasis by environment interaction effects, $h^2_a$ = additive heritability of individual locus, $h^2_{aa}$ = epistasis heritability of pair-wise loci, $h^2_{ae}$ = heritability of additive by environment interaction effect, $h^2_{aae}$ = heritability of epistasis by environment interaction effect.

Association Mapping

A GPU parallel computing software QTXNetwork (http://ibi.zju.edu.cn/software/QTXNetwork/) was used to dissect the genetic architecture of lint yield and its component traits. There were 651 SSR markers used for association mapping for yield traits of 323 representative Upland cotton cultivars in nine different environments. Significant SSR markers were screened and the estimation of fixed effects ($\phi$) and prediction of random effects ($a$, $aa$, $ae$, and $aae$) of loci were obtained.

Gene Locations Based on the Associated Markers by Reference D Genome

Based on the experiment-wise type I error ($\alpha_{EW} < 0.05$) setting by QTXNetwork, the SSR loci associated with the fiber yield traits were selected. The primer information was acquired from cotton marker database (http://www.cottonmarker.org/cmd_downloads/ssr_project_data/CMD_PRIMER_ALL.xls). Only one SSR location for each marker on the genome D (ftp://ftp.ncbi.nlm.nih.gov/genbankigenomes/Eukaryotes/plants/Gossypium_raimondii/) was acquired by bowtie method which 2 bases not matched was permitted (bowtie -a -v 2–fr cottonPD -f -1 primerForward.fasta -2 primerReverse.fasta -best –strata bowtieResult_PD.txt). The SSR motifs of the searched SSR sequences on the D genome were scanned by MISA. The location between SSR locus and the related gene was decided based on the star location of the motif on the genome.

Results

In total, there were 94 significant SSRs (14 loci with individual effects and 93 loci involved in epistasis interactions) associated with lint yield, 95 SSRs (28 individual loci) associated with boll number, 91 SSRs (21 individual loci) associated with boll weight, and 88 SSRs (19 individual loci) associated with lint percentage. All identified loci associated with lint yield components were involved in epistasis interactions.

As to the lint yield, the total heritability of genotype × environment interaction effects ($h^2_{GE}$ = 38.26%) was mainly contributed due to epistasis × environment interaction ($h^2_{AAE}$ = 34.05%), and was much larger than the total genotype heritability ($h^2_G$ = 10.80%) (Table 1). We listed highly significant genetic effects of loci ($-\log_{10}(P)>8.0$ and $h^2_{GE}>1.0\%$, $h^2_G>0.5\%$) and consistent loci with environment interaction which were stable in a particular ecological area across at least two years for lint yield and three yield traits (Figure 1, Table 2–5). Among the loci associated with lint yield (Table 2), 1 locus (NAU3011-2) and 6 pairs of epistasis loci were highly significant with environmental interaction at Nanjing in 2007 and 2009. It was suggested that improving lint yield could be expected by selecting major-allele
Table 1. Estimated heritability for lint yield and yield components.

| Trait            | $h^2_A$ (%) | $h^2_{AE}$ (%) | $h^2_{G+GE}$ (%) | $h^2_{AA+AE}$ (%) | $h^2_{A+GE}$ (%) |
|------------------|-------------|----------------|------------------|--------------------|-----------------|
| Lint Yield       | 0.20        | 10.6           | 4.21             | 34.05              | 49.06           |
| Boll Number      | 3.87        | 9.28           | 8.89             | 41.93              | 63.98           |
| Boll Weight      | 21.48       | 22.01          | 0.63             | 18.73              | 62.85           |
| Lint Percentage  | 6.58        | 57.78          | 0.29             | 7.64               | 72.29           |

Note: $h^2_A$ = heritability for additive effects, $h^2_{AE}$ = heritability for additive epistasis effects, $h^2_{G+GE}$ = heritability for additive by environment interaction effects, $h^2_{AA+AE}$ = heritability for additive epistasis by environment effects, $h^2_{A+GE}$ = total heritability.
The heritability of complex traits would be missed in the case of reduced models [20–21]. The heritability of lint yield and yield components decreased in the additive model (Table 1), the total heritability of lint yield and yield components. As compared with the full model (Table 1), the total heritability of lint yield and yield components decreased in the additive model ($h^2 = 18.35\%$ for lint yield, $6.53\%$ for boll number, $28.65\%$ for boll weight, and $22.95\%$ for lint percentage), and also in the additive with environmental interaction ($h^2 = 17.31\%$ for lint yield, $23.76\%$ for boll number, $32.25\%$ for boll weight, and $25.08\%$ for lint percentage). The results also showed steep decrease in numbers and effects of loci associated lint yield and yield components between reduced models and full model. This may explain the reasons for missing heritability reported recently [21].

Because of low LD associated with higher frequencies of cross-pollination, cotton species are amenable for association mapping of agronomic traits with a relatively large numbers of markers. Previous studies involving nearly 1000 polymorphic markers would ensure the coverage of whole cotton genome [14]. A total of 651 SSR markers and 323 Upland cotton cultivars under 9 environments in this study can significantly improve the power of association mapping. However, with the rapid development of next generation sequencing, a substantial number of single nucleotide polymorphisms (SNPs) will facilitate the accurate and fine association analysis.

It is shown by the results of GWAS that lint yield, boll number and lint percentage are inherited in a mainly epistasis manner, whereas boll weight is predominately controlled by both additive and epistasis effects. In addition, lint yield and boll number are particularly susceptible to a broad range of environments, as compared to boll weight and lint percentage, which has also been observed in previous studies [5,22–23]. Both lint yield and component traits were complex traits controlled by many genes and gene-network cascades interacting with each other and also with the environment factors (Figure 1). Many loci (ten or more in the present study) associated with lint yield, boll number, and boll weight showed much larger effects and heritability (additive or epistasis effects), while few loci associated with lint percentage could have large effects. This was indicated that improving cotton lint yield could be difficult through increasing lint percentage. All these information obtained from GWAS could provide useful reference for MAS in cotton breeding programs.

Although the effects of many loci appeared to be consistent across various environments, the magnitude of effects and even direction of loci may vary depending on environmental factors because of interactions between loci and environment [3,10,24]. We can classify loci identified in the present study into three types. The first category of loci is called constitutive loci with large major effects ($a, a^d$) and can be consistently detected across different environments. This type of loci is very useful for improving cotton traits across different ecological locations. Another type of loci is called environment-specific loci with only environment-special effects identified in a special ecological region but stable in different years. For example, TMB312–212 × GH132-180 was stably found in the Northwestward region and NAU3013–245 × NAU5163–216 was stably scanned in the Yangtze River valley region. This type of loci could be used for maker-assisted breeding in a particular ecological area. The third type of loci is called environment-sensitive loci that could not be consistently detected in various grown regions. This kind of loci could only be used in specific conditions.

SSRs widely used in major crop were highly reliable, polymorphic, simple and cheap. The SSRs associated with nearby genes were extremely useful for MAS. In the present study, we successfully detected 251 significant loci associated with lint yield and its three components, including 69 loci with individual effects. Most of these significant loci generally had small impacts on cotton

### Table 2. Predicted genetic effects with significance and heritability of lint yield for cotton regional trials in three locations and three years.

| Locus          | Effect | Predict | $-\log_{10} P$ | $h^2$ (%) | Candidate Gene |
|----------------|--------|---------|---------------|-----------|----------------|
| NAU3011–197    | $a_{y1}$ | 1.01    | 11.32         | 1.04      | IPR000679      |
| NAU1794–320 × TMB10–375 | $a_{y1}$ | −1.58   | 26.17         | 0.96      | − × Gorai.004G126800 |
| MUCS101–330 × MGHE518–228 | $a_{y1}$ | 0.98    | 10.97         | 0.65      | Cotton_A_29394 × IPR001128 |
| NAU1362–225 × NAU3774–248 | $a_{y1}$ | −0.84   | 8.05          | 0.79      | IPR001764 × IPR002913 |
| NAU3325–238 × HAU423–175 | $a_{y1}$ | −1.04   | 12.03         |           |                 |
| NAU3325–238 × TMB1989–255 | $a_{y1}$ | −1.25   | 17.12         |           |                 |
| NAU3608–245 × HAU773–170 | $a_{y1}$ | 1.15    | 14.23         |           |                 |

**Note:** $a_{y1}$ = additive by environment interaction effect at Nanjing in 2007; $a_{y2}$ = additive by environment interaction effect at Nanjing in 2009; $a_{y3}$ = epistasis by environment interaction effect at Nanjing in 2007; $a_{y4}$ = epistasis by environment interaction effect at Nanjing in 2008; $a_{y5}$ = epistasis by environment interaction effect at Nanjing in 2009; $-\log_{10} P =$ minus log10 ($P$-value), $h^2$ (%) = heritability (%).

doi:10.1371/journal.pone.0095882.t002
breeding due to small effects and easily affected by environment [23]. In order to develop effective MAS schemes, reliable and elite loci associated with lint yield and yield components have been screened. These loci, which consist of constituted loci and environment-specific loci, are reliable to predict corresponding phenotype. Constituted loci (such as NAU3588−315 and NAU3332−230) can be applied for MAS in various cotton-growing region, whereas application of environment-specific loci (examples are NAU1102−241 and NAU1362−225 × NAU3774−248) are limited to specific cotton-growing region. Both of them have substantial potential to improve the efficiency and precision of conventional cotton breeding.

With the information from Cotton Microsatellite Database (CMD) and published report, several markers identified in our present study had been located within or nearby the reported QTL in previous linkage analysis and association mapping [1–8,14]. For example, Zhang et al [14] independently detected QTLs called qLP-A1-1 (between JESPR101 and BNL3590) using linkage method and a locus called JESPR101 using association method, which affected lint percentage. We also identified this marker ($-\log_{10}P$-value = 15.09) associated with the same (Table 5). In addition, most of these consistent loci were involved in interaction in our present results (examples are NAU5433−330 for numbers of boll in Table 5). It indicated that these loci were stably transferred and can be further used in MAS.

The traditional QTL was mainly located between two markers with the scale of cM (centi-Morgan). Based on the physical mapping of reference genome, the fine mapping of the markers and the genes related to the fiber yield traits become possible. In this study, candidate gene locations of yield fiber traits were identified based on the associated SSR markers analyzed by the QTXNetwork. For the lint yield, one major-allele genotype $QQ$ (NAU3011−197) related to zinc finger (GATA-type) gene (IPR000679) was identified. 3 epistasis loci with positive effects ($QQ\times QQ$ genotype) could be related to the interactions with the genes of cytochrome P450 (IPR001128), Cytochrome b245 (IPR000778), zinc finger with $C2H2$-type (IPR007087), and protein kinase (IPR000719) etc. 3 minor-allele with negative effects ($qq\times qq$ genotype) was linked to the interactions with the genes of glycoside hydrolase (IPR001764), lipid-binding

### Table 3. Predicted genetic effects with significance and heritability of boll number for cotton regional trials in three locations and three years.

| Locus            | Effect | Predict | $-\log_{10}P$ | $h^2$ (%) | Candidate Gene          |
|------------------|--------|---------|---------------|-----------|-------------------------|
| HAU385−150       | $a$    | 0.3     | 42.29         | 0.68      | IPR003245               |
| MGHE541−330      | $a$    | −0.27   | 34.72         | 0.55      | IPR000297               |
| NAU1102−241      | $a$    | −0.28   | 36.71         | 0.6       | IPR003329               |
| $ae_1y_1$        | −0.38  | 8.35    | 0.63          |           |                         |
| $ae_1y_2$        | −0.46  | 11.99   |               |           |                         |
| $ae_2y_1$        | −0.52  | 14.8    |               |           |                         |
| GH111−245×NAU874−215 | $ae_1y_1$ | 0.37 | 8.25         | 0.37     | $\times$IPR001128     |
|                  | $ae_2y_1$ | 0.39 | 8.93         |           |                         |
| GH111−245×NAU5433−330 | $ae_1y_1$ | 0.55 | 16.75        | 0.96     | $\times$IPR001865      |
|                  | $ae_2y_1$ | 0.84 | 37.43        |           |                         |
| HAU1029−197×NAU1125−243 | $ae_1y_2$ | 0.47 | 12.16        | 1.59     | IPR001012×IPR016196    |
|                  | $ae_2y_2$ | 0.55 | 16.85        |           |                         |
|                  | $ae_3y_1$ | 0.93 | 46.13        |           |                         |
|                  | $ae_3y_2$ | 0.63 | 21.32        |           |                         |
| HAU1969−375×JESPR42−128 | $ae_1y_1$ | 1.08 | 60.86        | 2.43     | IPR000181×Gorai.009G322400 |
|                  | $ae_2y_1$ | 1.27 | 83.72        |           |                         |
| HAU1639−320×NAU5099−233 | $ae_1y_1$ | 0.38 | 8.55         | 0.53     | IPR000217×IPR001214    |
|                  | $ae_2y_1$ | 0.5 | 14.18        |           |                         |
| NAU2715−184×JESPR42−128 | $ae_3y_1$ | −0.63 | 22.12 | 0.65 | IPR004022×Gorai.009G322400 |
|                  | $ae_3y_2$ | −0.38 | 8.43         |           |                         |
| NAU2873−352×NAU5433−330 | $ae_1y_1$ | −0.50 | 13.93 | 0.77 | IPR003311×IPR001865    |
|                  | $ae_2y_1$ | −0.77 | 32.56        |           |                         |
| NAU4042−175×GH111−245 | $ae_1y_1$ | 0.47 | 12.53 | 0.64 | Cotton_A_04013×          |
|                  | $ae_2y_1$ | 0.68 | 25.62        |           |                         |
| STV61−131×NAU3305−155 | $ae_2y_1$ | −0.91 | 43.35 | 1.49 | DUF1685×IPR002913    |
|                  | $ae_3y_1$ | −0.65 | 22.48        |           |                         |
| TMB1181−220×NAU3563−149 | $ae_1y_1$ | 0.6 | 19.26 | 1.25 | Gorai.002G088500×IPR003311 |
|                  | $ae_2y_1$ | 0.42 | 9.91         |           |                         |

**Note:** $a$ = additive effect; $ae_1y_1$, $ae_1y_2$, $ae_2y_1$, $ae_2y_2$, $ae_3y_1$ as defined in Table 2; $ae_1y_1$ = additive by environment interaction effect at Anyang in 2008, $ae_1y_1$ = epistasis by environment interaction effect at Nanjing in 2008, $ae_1y_1$ = epistasis by environment interaction effect at Anyang in 2008, $ae_1y_1$ = epistasis by environment interaction effect at Anyang in 2009; $-\log_{10}P$ = minus $log_{10}$($P$-value), $h^2$ (%) = heritability (%).

doi:10.1371/journal.pone.0095882.003
cytochrome P450 have negative effects for lint yield. But the decompose genes of glycoside, lipid and protein maybe organic substances will have positive effects on cotton lint yield, (Table 2). It seems that the genes catalyzed the oxidation of (IPR002913), and peptidase aspartic (IPR021109) at Nanjing areas (Table 2). It seems that the genes catalyzed the oxidation of organic substances will have positive effects on cotton lint yield, but the decompose genes of glycoside, lipid and protein maybe have negative effects for lint yield.

For boll number, there were 3 additive SSR loci (HAU1385-150, NAU1102-241, MGHE41-333) related to genes of pep-tidyl-prolyl cis-trans isomerases (PPLase, IPR002927), plastocyanin-like (IPR003245), and acylneuraminate cytidylyltransf erase (IPR003292). 7 epistasis loci with positive effects for boll number (QQQ×QQQ genotype) might be related to gene interactions of cytochrome P450 (IPR001128), UBX domain-containing protein (IPR001012), tubulin (IPR000217), ribosomal protein (IPR001845), major facilitator superfamily (IPR016196), formyl-methionine deformylase (IPR000181), and SET domain of SUVH4 (IPR001214) etc. There were 3 epistasis loci with negative effects (qq×qq genotype) related to gene interactions of DDT domain superfamily (IPR004022), AUX/LLI protein (IPR003311), ribosomal protein (IPR001865), protein of unknown function (DUF1685), lipid-binding START (IPR002913) etc. It seems that the gene interactions for the metabolism and transport of cysteine, prolyl and formylmethionine could increase the boll number. But the regulation of the gene interactions of the lipid-binding DNA binding and auxin-binding protein could also affect the boll number.

Table 4. Predicted genetic effects with significance and heritability of boll weight for cotton regional trials in three locations and three years.

| Locus          | Effect | Predict | −Log10P | h² (%) | Candidate Gene |
|----------------|--------|---------|---------|--------|----------------|
| NAU3744-200    | a      | −0.07   | 66.02   | 1.29   | IPR020828      |
| TMB312-212     | a      | −0.06   | 54.75   | 1.07   | IPR011335      |
| TMB1296-230    | a      | 0.07    | 60.15   | 1.18   | IPR002591      |
| BNL1313-180    | a      | 0.08    | 95.32   | 1.89   | –              |
| NAU3588-315    | a      | −0.15   | 299.91  | 6.14   | –              |
| GH132-180      | a      | −0.06   | 47.17   | 0.92   | IPR001611      |
| JESP274-129    | a      | −0.06   | 43.33   | 0.84   | –              |
| JESP101-122    | a      | 0.05    | 42.82   | 0.83   | IPR004813      |
| NAU3305-155    | a      | −0.05   | 42.63   | 0.83   | IPR002913      |
| NAU5099-280    | a      | 0.05    | 42.79   | 0.83   | IPR001214      |
| TMB1963-243    | a      | −0.05   | 37.99   | 0.73   | –              |
| HAU1385-155    | a      | 0.05    | 36.04   | 0.7    | IPR003245      |
| BNL3033-175    | a      | −0.05   | 33.68   | 0.65   | IPR008540      |
| NAU2931-250    | a      | −0.04   | 27.91   | 0.53   | IPR001841      |
| NAU1163-160    | a      | −0.04   | 26.63   | 0.51   | –              |
| TMB312-212×GH132-180 | aae, y1 | −0.08 | 11.95 | 0.35 | IPR011335×IPR001611 |
| TMB312-212×GH132-180 | aae, y2 | −0.07 | 8.79 | 0.25 | –              |
| NAU3031-245×NAU5163-216 | aae, y1 | −0.09 | 12.53 | 0.66 | Cotton_A_34510×IPR002917 |
| NAU3031-245×NAU5163-216 | aae, y2 | −0.08 | 10.01 | 0.10 | –              |

Note: a = additive effect, a = additive by additive epistasis effect; aae = epistasis by environment interaction effect at Kuche in 2007; aae, y = epistasis by environment interaction effect at Kuche in 2008; aae, y = epistasis by environment interaction effect at Nanjing in 2007. aae, y = epistasis by environment interaction effect at Nanjing in 2009. −Log10P = minus log10 (P-value), h² (%) = heritability (%). doi:10.1371/journal.pone.0095882.t004

Table 5. Predicted genetic effects with significance and heritability of lint percentage for cotton regional trials in three locations and three years.

| Loci            | Effect | Predict | −Log10P | h² (%) | Candidate Gene |
|-----------------|--------|---------|---------|--------|----------------|
| NAU3325-238     | a      | −0.44   | 164.11  | 2.47   | IPR000778      |
| NAU3519-200     | a      | −0.23   | 46      | 0.67   | IPR000719      |
| NAU3519-220×TMB1268-157 | aa | 0.29 | 71.31 | 1.04 | IPR000719×IPR002048 |
| HAU1185-174×TMB1791-217 | aa | −0.21 | 38.63 | 0.55 | IPR000194×IPR000209 |
| TMB1638-189×DPLS13-320 | aa | 0.21 | 37.87 | 0.54 | IPR002109×– |
| NAU3110-292×NAU2862-248 | aa | −0.20 | 36.04 | 0.52 | IPR003329×IPR001841 |
| JESP101-122×BN1231-228 | aa | −0.20 | 35.61 | 0.51 | IPR004813×– |

Note: Genetic effect: a = additive effect, aa = additive by additive epistasis effect; −Log10P = minus log10 (P-value), h² (%) = heritability (%). doi:10.1371/journal.pone.0095882.t005
For boll weight, there were 15 additive SSR loci with positive effects related to following genes: Type 1 phosphodiesterase (IPR000591), SET domain (IPR001214) of SUTH4, plastocyanin-like (IPR000243), oligopeptide transporter (OPT) superfamily (IPR004813). While the genes could have negative effects, such as glyceraldehyde 3-phosphate dehydrogenase (IPR020828), restriction endonuclease (IPR003135), leucine-rich repeat (IPR001841), STAR-related lipid-transfer (START) domain (IPR004577), brassinosteroid signalling positive regulator (BZR1) family protein (IPR000340), and RING-type zinc finger (IPR001841), respectively. Except the main additive effects, there are also 2 epistasis loci with negative effects (qq×qq genotype) which maybe resulted from the interactions of leucine-rich repeat (IPR001611), GTP-binding protein (IPR002917) etc. It was inferred that boll weight is mainly controlled by the genes with the phosphorus, lipid and signalling transport, which may be further regulated by the zinc finger, leucine-rich-repeat, and the methylation protein.

For lint percentage, 2 SSR loci with negative additive effects could be related to protein kinase (IPR000719) and cytochrome b245 (IPR000778), respectively. There were 2 epistasis loci with positive effects (QQ×QQ genotype) related to the interactions of the genes of protein kinase (IPR000719), calcium-binding EF-hand (IPR004813), glutaredoxin (IPR002109). Three epistasis loci with negative effects (qq×qq genotype) relate to the interactions of the genes of peptidase S8/S3 (IPR000209), zinc finger with RING-type (IPR001841), and oligopeptide transporter OPT superfamily (IPR004813).

Supporting Information

Table S1  Variety germplasm of Upland cotton used in association mapping for fiber traits. Note: New lines were breeding lines and genetic stocks bred by Institute of Cotton Research of Chinese Academy of Agricultural Sciences (ICR, CAAS) in recent years.

Table S2  651 primer alleles that were polymorphic in the tested population.

(XLS)

Author Contributions

Conceived and designed the experiments: XMD. Performed the experiments: YHJ XWS JLS ZEP XWW SHX WJS ZLZ BYP LRW JGL JM. Analyzed the data: JZ. Contributed reagents/materials/analysis tools: XMD JZ. Wrote the paper: XWS JZ XMD.

References

1. He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, et al. (2003) Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. Euphytica 144: 141–149.
2. Shen XL, Zhang TZ, Guo WZ, Zhu XF, Zhang XY (2006) Mapping Fiber and Yield QTLs with Main, Epistatic, and QTL × Environment Interaction Effects in Recombinant Inbred Lines of Upland Cotton. Crop Science 46: 61.
3. Murray CE, Lewinger JP, Sauderman WJ (2008) Gene-Environment Interaction in Genome-Wide Association Studies. American Journal of Epidemiology 169: 219–226.
4. Chen H, Qian N, Guo WZ, Song QP, Li BC, et al. (2010) Using three selected overlapped RILs to fine-map the yield component QTL on Chr. D8 in Upland cotton. Euphytica 176: 321–329.
5. Liu RZ, Wang BH, Guo WZ, Qin YS, Wang LG, et al. (2011a) Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in Gossypium hirsutum L. Mol Breeding 29: 297–311.
6. Yadava SK, Arumugan N, Mukhopadhyay A, Sodhi YS, Gupta V, et al. (2012) QTL mapping of yield-associated traits in Brassica juncea: meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. Theor Appl Genet 125: 1533–1546.
7. Yu JW, Zhang K, Li SY, Yu SX, Zhai HH, et al. (2012) Mapping quantitative trait loci for lint yield and fiber quality across environments in a Gossypium hirsutum × Gossypium barbadense backcross inbred line population. Theor Appl Genet 126: 273–287.
8. Lacapé JM, Gouzy JF, Cao TV, Vico G, Llewellyn DJ, et al. (2013) Mapping QTLs for traits related to phenotype, morphology and yield components in an inter-specific Gossypium hirsutum × G. barbadense cotton RIL population. Field Crop Research 144: 256–267.
9. Huang XH, Wei XH, Sang T, Zhao Q, Peng Q, et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genetics 42: 961–967.
10. Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci USA 108: 6893–6898.
11. Kump KI, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, et al. (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nature Genetics 43: 163–168.
12. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, et al. (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics 43: 159–162.
13. Li H, Peng ZY, Yang XH, Wang WD, Fu JJ, et al. (2012) Genome-wide association study dissect the genetic architecture of oil biosynthesis in maize kernels. Nature Genetics: 1–10.
14. Zhang TZ, Qian N, Zhu XF, Chen H, Wang S, et al. (2013) Variations and Transmission of QTL Alleles for Yield and Fiber Qualities in Upland Cotton Cultivars Developed in China. PLoS ONE 8: e57220.
15. Ziegler A, Ko¨nig IR, Thompson JR (2008) Biostatistical Aspects of Genome-Wide Association Studies. Biom J 50: 8–28.
16. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5: 618–625.
17. Phillips PC (2008) Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. Nature Publishing Group 9: 835–867.
18. van Os J, Rutten BPF (2009) Gene-environment-interaction studies in psychiatry. Am J Psychiatry 166: 964–966.
19. Culverhouse R, Suarez BK, Lin J, Reich T (2002) A perspective on epistasis: limits of models displaying no main effect. Am J Hum Genet 70: 461–471.
20. Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12: 232.
21. Zuck O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109: 1193–1198.
22. Godoy S, Palomo G (1999) Genetic analysis of earliness in upland cotton. II. Yield and fiber properties. Euphytica 105: 161–166.
23. Campbell BT, Jones MA (2005) Assessment of genotype × environment interactions for yield and fiber quality in cotton performance trials. Euphytica 168: 69–76.
24. Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363: 557–572.