Research Article

A genome-wide association study of early-maturation traits in upland cotton based on the CottonSNP80K array

Chengqi Li1†*, Yuanyuan Wang1†, Nijiang Ai2, Yue Li1 and Jiafeng Song1

1. Collaborative Innovation Center of Modern Biological Breeding, Henan Province/Cotton Research Institute, Henan Institute of Science and Technology, Xinxiang, Henan Province, 453003, China
2. Shihezi Agricultural Science Research Institute, Shihezi, Xinjiang Province, 832000, China

†These authors contributed equally to this work.
*Correspondence: Chengqi Li (lichq2010@126.com)

Edited by: Jin-Sheng Lai, China Agricultural University, China
Running title: Association mapping of cotton early-maturation traits

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/jipb.12673]

This article is protected by copyright. All rights reserved
Received: February 02, 2018 / Accepted: May 31, 2018
Abstract Genome-wide association studies (GWAS) efficiently identify genetic loci controlling traits at a relatively high resolution. In this study, variations in major early-maturation traits, including seedling period (SP), bud period (BP), flower and boll period (FBP), and growth period (GP), of 169 upland cotton accessions were investigated, and a GWAS of early maturation was performed based on a CottonSNP80K array. A total of 49,650 high-quality single-nucleotide polymorphisms (SNPs) were screened, and 29 significant SNPs located on chromosomes A6, A7, A8, D1, D2, and D9, were repeatedly identified as associated with early-maturation traits, in at least two environments or two algorithms. Of these 29 significant SNPs, 1, 12, 11, and 5 were related to SP, BP, FBP, and GP, respectively. Six peak SNPs, TM47967, TM13732, TM20937, TM28428, TM50283, and TM72552, exhibited phenotypic contributions of ~10%, which could allow them to be used for marker-assisted selection. One of these, TM72552, as well as four other SNPs, TM72554, TM72555, TM72558, and TM72559, corresponded to the quantitative trait loci previously reported. In total, 274 candidate genes were identified from the genome sequences of upland cotton and were categorized based on their functional annotations. Finally, our studies identified Gh_D01G0340 and Gh_D01G0341 as potential candidate genes for improving cotton early maturity.

Keywords: Candidate gene; Early maturity; Genome-wide association study (GWAS); Single-nucleotide polymorphism (SNP); Upland cotton
INTRODUCTION

Cotton is an important commercial crop that provides most of the natural fiber in the world and is also an important source of edible oil (Campbell et al. 2011). Upland cotton (*Gossypium hirsutum* L.) ($2n = 4x = 52$), one of 50 *Gossypium* species and the leading natural fiber crop worldwide, contributes more than 95% of total cotton production due to its high yield and wide adaptability (Chen et al. 2007). Early-maturity cotton, a specific type of upland cotton, generally exhibits a dwarf, compact phenotype with fewer leaves, shorter internodes and fruiting branches, and a shorter growth period than middle-late-maturing cotton (Yu and Huang 1990). The breeding and popularization of early-maturity cotton varieties is important to relieve environmental pressure and optimize the structure of the cotton industry.

Cotton early maturity is a comprehensive trait that includes components such as seedling period (SP), bud period (BP), flower and boll period (FBP), growth period (GP), node of the first fruiting branch and its height, and yield percentage before frost (Yu and Huang 1990; Li et al. 2013). These components are all complex quantitative traits (White 1966), which are controlled by quantitative trait loci (QTLs) and environmental factors, and may exhibit various genetic patterns, in different combinations and environments (Song et al. 2005; Dong et al. 2010). However, there is a significant negative correlation between early maturity and yield (Song et al. 2005; Fan et al. 2006b).

Obtaining a satisfactory yield, in a relatively short growing season, compounds the tradeoff between early maturity and yield. Thus, it is difficult to use traditional breeding techniques to simultaneously improve both of these important cotton traits. Using molecular marker techniques to identify QTLs for early-maturation traits and applying the molecular markers that are closely linked to, or significantly associated with, target QTLs for marker-assisted selection (MAS) could effectively disrupt the negative correlation between early maturity and other traits, rapidly producing early-maturity cotton varieties with high yield and high quality.

Linkage mapping and association mapping are the two main methods
used to analyze the genetic basis of plant quantitative traits. Linkage mapping is a traditional method used in gene/QTL analysis. Here, the QTLs for target traits are located in particular linkage segments, and the genetic effects of QTLs are analyzed, based on high-density genetic linkage maps constructed from limited parental hybridization.

In association mapping, after both phenotypic and genotypic data are obtained, for a natural population or germplasm resource population, a statistical method is used to detect the association between genetic polymorphisms and a trait’s heritable variation, based on linkage disequilibrium (LD). Compared with linkage mapping, association mapping provides dramatically better QTL detection efficiency, as it uses the germplasm population, which has a greater level of genetic variation (Gupta et al. 2005).

Before the cotton genome was sequenced, researchers mainly used simple sequence repeat markers (SSRs) to detect QTLs and/or molecular markers related to target traits. For cotton early maturity, several genetic linkage maps have been constructed using F$_2$, F$_{2:3}$, and recombinant inbred line populations, and more than 80 QTLs for early-maturation traits have been identified, using linkage mapping (Fan et al. 2006a; Zhang et al. 2008; Guo et al. 2008, 2009; Li et al. 2012; Li et al. 2013). Subsequently, a number of SSR markers that were significantly associated with early-maturation traits were identified, using association mapping (Li et al. 2016a). The above results laid an important foundation for analyzing the genetic basis of cotton early maturity. However, due to the limited number of molecular markers used in these studies, the genetic bases of quantitative traits could not be fully dissected, at the genome-wide level.

In recent years, high-throughput sequencing technology has been widely used. For the upland cotton genome, Zhang et al. (2015) sequenced the allotetraploid *Gossypium hirsutum* L. acc. TM-1 genome by integrating whole-genome shotgun reads, bacterial artificial chromosome (BAC)-end sequences, and genotype-by-sequencing genetic maps. Additionally, Li et al. (2015) also produced a draft genome of TM-1 using 181-fold paired-end sequences, assisted by fivefold BAC-to-BAC sequences and a high-resolution genetic map.

A new cotton single-nucleotide polymorphism (SNP) chip, the
CottonSNP80K array, covering the genome of upland cotton, was successfully developed based on the sequencing of TM-1 (Zhang et al. 2015) and the re-sequencing of 100 different cultivars in upland cotton with 5× coverage, on average (Fang et al. 2017). The array contains the following information: (1) 77,774 SNP loci and a genotype accuracy of ≥ 99.12%; (2) filtered SNPs in repeat regions; (3) no other SNPs or insertions or deletions (InDels) in the 50 base pairs (bp) flanking the SNP sites; (4) heterozygosity rates of ≤ 15%; (5) SNPs in gene regions have Illumina design scores of ≥ 0.7; and (6) SNPs in intergenic regions have Illumina design scores of ≥ 0.9 (Cai et al. 2017). The CottonSNP80K array provides important marker resources for genome-wide association studies (GWASs) in upland cotton.

Early maturity is an important and complex agronomic trait in upland cotton. To better understand its genetic architecture, we phenotypically characterized a population of 169 upland cotton accessions for the major early-maturation traits, in multiple environments, and genotyped it with the CottonSNP80K array. We conducted a GWAS using three algorithms: the general linear model (GLM), the mixed linear model (MLM), and the factored spectrally transformed linear mixed model (FaST-LMM). Based on our GWAS, we identified marker-associated candidate genes within the LD decay distance of upland cotton. Our findings serve as a foundation for the analysis of genetic mechanisms underlying early maturity, at the genome-wide level, and for MAS for early-maturation traits, in cotton.
RESULTS

Performances of early-maturation traits
We generated descriptive statistics and an ANOVA for the four early-maturation traits, seedling period (SP), bud period (BP), flower and boll period (FBP), and growth period (GP), of the 169 upland cotton accessions (Table 1). The coefficient of variation (CV) ranges for SP, BP, FBP, and GP in the four environments, E1 (2012Xinxiang), E2 (2013Xinxiang), E3 (2012Shihezi) and E4 (2013Shihezi), were, respectively, 4.59–8.98%, 5.40–8.83%, 3.88–9.81%, and 2.92–6.92%, and their average CVs were 7.17%, 7.31%, 6.18%, and 4.74%, respectively, which reflected the degree of variation in these traits.

For all four traits, the variance in genotype, environment, and the interaction between genotype and environment (G × E) were all highly significant, at the α = 0.01 level and/or at the α = 0.001 level, indicating that they are all complex quantitative traits controlled by both genotype and environment. The histogram of phenotypic traits investigated in the four environments (Figure 1) showed that early-maturation traits exhibited the genetic characteristics of quantitative traits, with continuous distributions across different environments. Furthermore, some of the traits exhibited multimodal or partial distributions, suggesting that genes/QTLs of major effect related to early maturity exist in the cotton genome. The broad-sense heritability values for these four early-maturation traits were 64.6%, 61.2%, 67.3%, and 78.7%, respectively, suggesting that they are under relatively high levels of genetic control.

Distribution of polymorphic SNPs across the genome of *G. hirsutum*
We next examined the genotypes of the 169 accessions, using Illumina GenomeStudio software. Only the SNPs with minor allele frequencies of ≥ 0.05 and integrities of ≥ 50% in the population were used for screening polymorphic loci; a total of 49,650 high-quality SNPs (of 77,774 total) were retained after applying these criteria (Figure 2, Table 2). These SNPs were unevenly distributed across the *G. hirsutum* genome, with 25,820 and 23,830 SNPs in the A and D subgenomes, respectively.
The average marker density was approximately one SNP per 41.179 kb. In the A subgenome, chromosome A6 had the most markers (2,746), with a marker density of one SNP per 37.571 kb, and A4 had the fewest markers (966), with a marker density of one SNP per 65.128 kb. For the D subgenome, chromosome D6 had the most markers (2,919), with a marker density of one SNP per 22.026 kb, and D4 had the fewest markers (972), with a marker density of one SNP per 52.936 kb. The polymorphism information content (PIC) values ranged from 0.246 to 0.309 among chromosomes, and the mean PIC values of the A and D subgenomes were 0.282 and 0.281, respectively.

Population structure and LD analysis
To estimate the number of subpopulations in the 169 upland cotton accessions, we performed a population structure analysis using all 49,650 SNPs. The minimum number of cross-validation errors was $K = 7$, which was thus determined to be the optimum $K$, and the testing accessions could be separated into seven subpopulations (Figure 3A, B). Subpopulations 1–7 included 41, 41, 24, 23, 22, 7, and 11 accessions, respectively (Table S1). Most of the upland cotton accessions from each subpopulation had mixed ancestry, and the obvious geographic subpopulation structure was not observed, indicating that these lines might have experienced introgression, or gene flow, during cotton breeding in China.

Although the 169 accessions were derived from multiple pedigrees and different ecological cotton-growing regions, the varietal population in this study was considered not highly structured, meaning that the population could be used for further marker–trait association mapping. The LD analysis estimated the average LD decay distance of our population for the AD genome to be ~400 kb, with $R^2 = 0.38$ at half of the maximum value, which was longer than in the A subgenome and shorter than in the D subgenome (Figure 3C).

Genome-wide association mapping
We next used three GWAS algorithms, GLM, MLM and FaST-LMM, to perform a GWAS. The SNPs with $-\log_{10}P$ values higher than 4.69 (1/49, 650) were selected as significant trait-associated SNPs. Here, 134 SNPs associated with early-maturation traits were detected using the values of individual
environments, including best linear unbiased predictors (BLUPs), and the three algorithms (Table S2). To obtain reliable results, we used only the significant SNPs associated with early-maturation traits repeatedly detected in at least two environments or two algorithms. As a result, 29 significant SNPs for the four early-maturation traits were selected (Table 3, Figures S1–S4).

For SP, we detected only one SNP, TM47967, located at bp 3,646,257 of chromosome D1. It was significantly associated with the E2 and BLUP values by Fast-LMM, and explained 0.18–0.20 of the phenotypic variation.

For BP, 12 SNPs located in chromosomes A6 and A7 were detected. Three SNPs—TM13730 and TM13732, located at bp 3,495,697 and 3,509,639, respectively, of chromosome A6, and TM20868 located at bp 67,275,473 of chromosome A7—were significantly associated with the E3 and/or BLUP values by GLM and MLM, and explained 0.14–0.16, 0.12–0.18, and 0.10–0.11, respectively, of the phenotypic variation. Seven SNPs, TM20788, TM20831, TM20860, TM20888, TM20918, TM20940, and TM20946, located in the bp 66,292,784–68,002,395 region of A7, were significantly associated with the BLUP value by GLM and Fast-LMM, and explained 0.11–0.21 of the phenotypic variation. Two SNPs, TM20903 and TM20937, located at bp 67,533,709 and 67,860,203 of chromosome A7, respectively, were significantly associated with the BLUP value by GLM, MLM, and Fast-LMM, and explained 0.13–0.20 of the phenotypic variation.

For FBP, 11 SNPs located on chromosomes A8, D2, and D9 were detected. Six SNPs—TM28405 and TM28428, located at bp 70,830,363 and 70,964,312, respectively, of chromosome A8; TM50280, TM50282, and TM50283, located at bp 900,486, 920,715, and 924,620, respectively, of chromosome D2; and TM72569, located at bp 43,617,803 of chromosome D9—were significantly associated with the E1, E3, and/or BLUP values by GLM and MLM, contributing 0.07–0.12 of the phenotypic variation. Five SNPs, TM72552, TM72554, TM72555, TM72558, and TM72559, located in the bp 43,446,639–43,499,161 region of chromosome D9, were significantly associated with the E1 and/or BLUP values by GLM, MLM, and Fast-LMM, and explained 0.08–0.25 of the phenotypic variation.

For GP, five SNPs located in chromosome D9 were detected. Three SNPs, TM72554, TM72555, and TM72559, located at bp 43,478,716, 43,483,274,
and 43,499,161, respectively, of chromosome D9, were significantly associated with the E1 and/or BLUP values by GLM and Fast-LMM, contributing 0.07–0.19 of the phenotypic variation. Two SNPs, TM72552 and TM72558, located at bp 43,446,639 and 43,495,825, respectively, of chromosome D9, were significantly associated with the E1 and/or BLUP values by GLM, MLM, and Fast-LMM, contributing 0.08–0.20 of the phenotypic variation.

Among all 29 SNPs associated with early-maturation traits (Table 3), five SNPs located on chromosome D9, TM72552, TM72554, TM72555, TM72558, and TM72559, were simultaneously associated with FBP and GP, which could result from the pleiotropy of a single causal gene, or the tight linkage of multiple causal genes.

Identification of candidate genes

Because of the LD decay distance of the upland cotton population used in this study, we designed our analysis to identify candidate genes that were located within 400 kb upstream and downstream of peak SNPs (the most significant SNPs, with maximum $-\log_{10} P$ values) based on genes annotated in the G. hirsutum acc. TM-1 genome (Zhang et al. 2015). Among 29 significant SNPs, six—TM47967 (D1) associated with SP, TM13732 (A6) and TM20937 (A7) associated with BP, TM28428 (A8) and TM50283 (D2) associated with FBP, and TM72552 (D9) associated simultaneously with FBP and GP—were identified as peak SNPs, which had maximum $-\log_{10} P$ values within a SNP peak region (Figure 4). An LD analysis showed that a high LD level existed among the SNPs within 400 kb upstream and downstream of the six peak SNPs on chromosomes D1 (Figure 4A), A6 (Figure 4B), A7 (Figure 4C), A8 (Figure 4D), D2 (Figure 4E), and D9 (Figure 4F). Therefore, we used the regions within 400 kb upstream and downstream of the six SNPs for further identification of candidate genes.

With this approach, we identified a total of 274 candidate genes, of which 47 were related to SP, 64 to BP, 79 to FBP, and 84 to both FBP and GP (Table S3). Because these four traits are directly related to cotton early maturity and are significantly positively correlated with each other, we merged these candidate genes into one group for a systematic summary (Figure 5). Among
the 274 candidate genes, 23 were correlated with transport and metabolism, 28 were correlated with transcription regulation and posttranslational modification, 4 were correlated with cell wall, cell membrane, or cellular components, 2 were correlated with defense mechanisms, 2 were correlated with energy production and conversion, 2 were correlated with senescence, 3 were correlated with RNA processing and modification, 23 were correlated with signal transduction mechanisms, 16 were correlated with translation or ribosomal structure and biogenesis, 10 were correlated with DNA replication, recombination, and repair, 24 were associated with general functional predictions only, 43 encoded putative and uncharacterized proteins, 75 encoded proteins with other functions (M), and 19 encoded proteins of unknown functions.

DISCUSSION

Early maturity is an important agronomic trait in upland cotton. The breeding and popularization of early-maturity cotton varieties are important in deciding to occupy farmlands with either cotton or cereal, while optimizing cropping systems (Yu and Huang 1990; Li et al. 2013). We therefore performed a GWAS of early maturity in upland cotton to dissect the genetic architecture of this agronomically-important trait.

**Abundant genetic variation among materials is the basis for performing a GWAS**

A suitable association mapping panel should embrace as much genetic variation, at the phenotypic and molecular levels, as can be reliably measured in common environments (Flint-Garcia et al. 2005). It is especially critical to select, as much as possible, samples that encompass genetic diversity. In this study, 169 upland cotton backbone cultivars and breeding lines, originating from different ecological cotton-growing areas in China and abroad, were selected for the association mapping panel. Although the sample was restricted to 169 accessions, they were mainly derived from multiple pedigrees that included Deltapine, Stoneville, Foster, and Uganda (the pedigrees of some materials are unknown). Thus, the range of pedigrees was very rich.
The phenotypic evaluation of early-maturation traits showed abundant phenotypic diversity and relatively high heritability values for these traits. The marker polymorphism data showed that from 77,774 SNPs, 49,650 high-quality SNPs were obtained across all of the accessions, accounting for 63.8% of the SNPs, which reflected a high genetic variation, at a molecular level. The PIC values of markers ranged from 0.246 to 0.309 among chromosomes, and the mean PICs of the A and D subgenomes were 0.282 and 0.281, respectively. Our results were similar to those reported by Sun et al. (2017), in which the PIC values ranged from 0.208 to 0.312, and the mean PICs of the A and D subgenomes were, respectively, 0.287 and 0.283. Sun et al. conducted a successful GWAS for fiber-quality-related traits using 719 diverse accessions of upland cotton and 10,511 polymorphic SNPs. Because of the greater genetic variation at the phenotypic and molecular levels among our accessions, this association panel was applicable for performing a GWAS for the target traits.

**CottonSNP80K array is a meaningful chip for a GWAS in upland cotton**

Among various molecular markers, SNPs represent the most abundant form of genetic variation. The availability of SNP data has increased the ability to analyze diversity in germplasm collections (van Treuren and van Hintum 2014; Mason et al. 2015; Li and Erpelding 2016) and to make advances in plant breeding programs (Varshney et al. 2009; Rafalski 2010; Fang et al. 2013; Thomson 2014; Gao et al. 2015; Wang et al., 2015). The first SNP chip for cotton, the CottonSNP63K array (Illumina, USA), was developed in 2015, allowing assays for 45,104 putative intra-specific SNP markers for use within the cultivated cotton species *G. hirsutum* L. and 17,954 putative inter-specific SNP markers for use with crosses of other cotton species with *G. hirsutum* (Hulse-Kemp et al. 2015). To date, the chip has been used in a diversity analysis of cotton germplasm (Hinze et al. 2017), high-density linkage mapping and QTL analysis (Hulse-Kemp et al. 2015; Li et al. 2016b), and GWASs of yield components and fiber quality traits (Gapare et al. 2017; Sun et al. 2017). Compared with CottonSNP63K, the second cotton SNP chip, CottonSNP80K, is particularly useful for comprehensively dissecting the genetic architecture of agronomically important traits in upland cotton because the SNP loci in the
CottonSNP80K array benefited from the whole-genome sequencing of *G. hirsutum* acc. TM-1 (Zhang et al. 2015) and 1,372,195 intraspecific non-unique SNPs identified by the re-sequencing of *G. hirsutum* accessions (Fang et al. 2017). Therefore, the selected SNPs in CottonSNP80K could be distributed along the entire genome of upland cotton. In addition, in contrast with the CottonSNP63K array, in the CottonSNP80K array each SNP marker is addressable, which avoids complications caused by homeologous/paralogous genes. Cai et al. (2017) recently used this chip to assess the phylogenetic relationships among, and performed a GWAS analysis of salt stress traits in upland cotton. They suggested that CottonSNP80K played important roles in variety verification, genetic relationship identification, and molecular breeding in cotton. In the present study, we performed the first GWAS study of major early-maturation traits in upland cotton via the CottonSNP80K array. The results form an important foundation for comprehensive analysis of the genetic mechanisms underlying cotton early maturation at the genome-wide level.

**Stable SNPs for early-maturation traits detected in this study**

The current GWAS method is a single-locus analysis approach conducted with polygenic background and population structure controls. The number of tests involved is the number of markers, requiring a Bonferroni correction for multiple tests. To control the experimental error, at a genome-wide level of 0.05, the significance level for each test should be adjusted to 0.05/n (where n is the total number of SNPs). The $-\log_{10}P$ value threshold is 5.99 if the 49,650 SNPs in this study are to be scanned.

The use of stringent probability thresholds reduces the risk of false positives, but does not reduce the risk of rejecting true positives caused by setting the thresholds too high. Therefore, referring to the method used by Cai et al. (2017) and Sun et al. (2017), we adopted an adjusted $P$ value of 1/n after the Bonferroni correction for estimating the significant levels of association. Additionally, to obtain reliable results, we used only significant SNPs associated with target traits repeatedly detected in at least two environments, or two algorithms. Ultimately, we detected 29 significant SNPs associated with early-maturation traits. In total, we identified six SNPs—TM47967 (D1) associated with SP, TM13732 (A6) and TM20937 (A7) associated with BP,
TM28428 (A8) and TM50283 (D2) associated with FBP, and TM72552 (D9) associated simultaneously with GP—as peak SNPs. Other SNPs concentrated near these peak SNPs in a tight genome region, forming a peak region, which further improved the stability of marker-trait associations. For example, 10 SNPs, TM20788, TM20831, TM20860, TM20868, TM20888, TM20903, TM20918, TM20937, TM20940, and TM20946, located at bp 66,292,784–68,002,395 of chromosome A7, were all detected as being associated with BP. This was usually the case for trait-associated loci, because the markers in this region were in strong LD with each other. In addition, the six peak SNPs all exhibited high phenotypic contributions of ~10%. Therefore, they could be used for MAS in future breeding programs.

**Comparison of the GWAS with QTLs identified in previous studies**

QTL/marker loci that are detected across multiple generations, populations, and/or environments are highly stable and can enhance the efficiency and accuracy of the MAS (Li et al. 2013). In previous studies, QTLs for early-maturation traits, and in particular for the four major early-maturation traits SP, BP, FBP, and GP, were identified throughout all 26 chromosomes of upland cotton, based on linkage and association mapping. QTLs for SP were identified on chromosomes A6, A7, A9, A13, D2, D3, D6, D8, D11, and D13 (Li et al. 2013; Li et al. 2016a); QTLs for BP were identified on chromosomes A1, A2, A9, A10, A12, D1, D3, D5, D9, and D13 (Li et al. 2013; Li et al. 2016a); QTLs for FBP were identified on 23 chromosomes of upland cotton, except A1, D1, and D11 (Li et al. 2013; Li et al. 2016a; Jia et al. 2016; Su et al. 2016); and QTLs for GP were identified on 21 chromosomes of upland cotton, except A1, A3, A5, A7, and D1 (Li et al. 2013; Li et al. 2016a; Jia et al. 2016, Su et al. 2016).

We compared the 134 SNP loci detected in the current study (Table S2) with SNP and SSR loci for early-maturation traits identified in previous studies, by electronic PCR (e-PCR), according to the physical locations of these loci on the genome sequence (Zhang et al. 2015). Among the 134 SNPs associated with the four early-maturation traits, 24 SNPs corresponded to previously reported QTLs, detected based on linkage and association mapping. Three SNPs for BP, TM55080, TM55081, and TM55084, located on chromosomes
D3, corresponded to QTLs reported by Li et al. (2013); 10 SNPs for FBP, TM57317, TM59105, TM59132, TM72552, TM72554, TM72555, TM72558, TM72559, TM78943, and TM78963, located on chromosomes D5, D6, D9, and D12, corresponded to QTLs reported by Li et al. (2016a) and/or Jia et al. (2016); and 11 SNPs for GP, TM59105, TM59145, TM72552, TM72554, TM72555, TM72558, TM72559, TM78939, TM78940, TM78945, and TM78953, located on chromosomes D6, D9, and D12, corresponded to QTLs reported by Li et al. (2016a) and/or Jia et al. (2016).

The markers related to these stably inherited QTLs, which were repeatedly detected in different populations with different genetic backgrounds, can potentially be used in the MAS of target traits. The five SNPs for FBP and GP, TM72552, TM72554, TM72555, TM72558, and TM72559, located on chromosomes D9, not only were detected across multiple environments and/or algorithms but also corresponded to previously reported QTLs. Thus, they should also be considered in the MAS.

**Potential candidate early-maturity genes**

The identification of stable markers/QTLs could provide useful information for MAS. Candidate gene analyses are necessary for further gene cloning and functional verification. Several studies have identified candidate genes related to cotton early maturity. Wang et al. (2014) identified a candidate gene, \( \text{GhFPF1} \), related to flowering time control in upland cotton whose transcript level was elevated in the floral apices of early-maturity cotton. Further genetic transformation data demonstrated that the overexpression of \( \text{GhFPF1} \) in \( \text{Arabidopsis} \) promoted flowering under inductive photoperiods.

Su et al. (2016) identified 32 candidate genes in the region near the SNP locus rs13562854 on chromosome D3 in upland cotton, which was positioned between DPL0200 and CIR347 in our previous study (Li et al. 2013). Of these 32 candidate genes, \( \text{CotAD\textunderscore 01947} \) showed significantly higher expression in the early-maturity cotton than in the late-maturing varieties, as assessed by qRT-PCR. After a further BLAST-algorithm-based alignment, they postulated that \( \text{CotAD\textunderscore 01947} \) was a potential candidate gene for improving early maturity through the regulation and control of early flowering time in upland cotton. A potential candidate gene on chromosome D3, named \( \text{EMF2} \), was identified by
Jia et al. (2016). The EMF2 expression trend was consistent with that of flower bud differentiation in cotton. They therefore concluded that EMF2 may also have a significant influence on flowering.

In this study, we identified a total of 274 candidate genes. These genes were categorized based on their functional annotations from several major databases. We cannot accurately determine which genes are directly related to early maturity, based on the data from this study; however, these results provide important information for future work. Among the candidate genes we identified, two, Gh_D01G0340 and Gh_D01G0341, are both highly homologous with AT1G17020 (senescence-related gene 1, SRG1) in Arabidopsis thaliana, which shows a significant correlation with flowering time (Sasaki et al. 2017). Further, SRG1 exhibited increased histone acetylation in the Arabidopsis hda6 mutant, axe1-5 that displayed a late-flowering phenotype, whereas histone acetylation played an important role in epigenetic controls of gene expression (Yu et al. 2011). Thus, it is reasonable to postulate that Gh_D01G0340 and Gh_D01G0341 may be potential candidate genes for improving cotton early maturation. However, early maturation is a complicated physiological and biochemical process that might involve a large number of structural, regulatory, and biochemical-pathway-related genes. Thus, a candidate gene analysis, including candidate gene association mapping and gene expression analysis, may further identify strong candidate genes.

MATERIALS AND METHODS

Experimental materials
A total of 169 upland cotton backbone cultivars and breeding lines cultivated in, or introduced to, China were selected to construct an association mapping panel, of which 62 were from ecological cotton-growing areas of the Yellow River, 25 were from the Yangtze River, 50 were from Northwestern China, 22 were from Northern China, and 10 were from other countries (Table S1). All accessions showed stable inheritance after many generations of self-pollination.
Experimental design and trait phenotyping
All materials were planted in two different ecological cotton-growing areas of China, the Yellow River (Xinxiang City, Henan Province) and Northwestern China (Shihezi City, Xinjiang Province) regions, in 2012 and 2013. The experiment followed a randomized complete block design with single row plot and two replications. In Xinxiang, 14–16 plants were arranged in each row with a row length of 5 m and a row interval of 1.0 m. In Shihezi, 38–40 plants were arranged in each row with a row length of 5 m and a row interval of 0.45 m. Standard local crop management practices were followed for all activities. For description purposes, the four environments, 2012Xinxiang, 2013Xinxiang, 2012Shihezi, and 2013Shihezi, were designated as E1, E2, E3, and E4, respectively. The four traits defining early maturation, SP (the period from seedling emergence to flower bud), BP (the period from flower bud emergence to flowering), FBP (the period from flowering to boll opening), and GP (the period from seedling emergence to boll opening), were investigated. Ten consecutive plants in the middle of each row were tagged for trait phenotyping. To reduce environmental error, BLUPs for the four early-maturation traits per genotype were estimated using the lme4 package in the R program (Bates et al. 2011). The values of BLUP and single environment were used for all association mapping. A statistical analysis was carried out using SAS v9.4 software. The analysis of the phenotypic changes in early trait trends was shown using histograms drawn using R. Broad-sense heritability, $H^2$, was calculated for each trait using the lme4 package in R (Bates et al. 2011).

DNA extraction and SNP genotyping
The DNAsecure Plant Kit (TIANGEN) was used according to the manufacturer’s instructions to isolate genomic DNA from fresh leaves for each accession. All DNA samples were quantified using a NanoDrop 2000 (Thermo Scientific) and diluted to a 50 ng/μL concentration. The genomic DNAs of all samples were whole-genome amplified for 20–24 h at 37°C, fragmented, precipitated, and resuspended in an appropriate hybridization buffer. The samples were hybridized on the prepared CottonSNP80K Genotyping BeadChips for 16–24 h at 48°C. Following the hybridization, nonspecifically
hybridized samples were removed by washing, while the remaining specifically hybridized loci were processed for the single-base extension reaction, stained, and imaged on an Illumina iScan Reader. Image files were saved and analyzed using the GenomeStudio Genotyping Module (v1.9.4, Illumina). Genotype calls for each SNP were performed based on the cluster file generated specifically for the CottonSNP80K array. Because the SNP calling module was developed for diploids, there were three possible genotypes, AA, AB, and BB, for each locus. For the purpose of the GWAS, the genotypes were coded as numeric values according to the additive mode of inheritance, which assumes that alleles have a dose effect: that is, if one allele variant has no effect (e.g., A) on the trait and the other has an effect (e.g., B), a homozygous individual with two copies of the allele (BB) will express twice the value of the effect as a heterozygous individual (AB). Thus, SNPs were genotyped as 0 (AA), 1 (AB), and 2 (BB).

Population structure and LD analyses
Only SNPs with minor allele frequencies of $\geq 0.05$ and integrities of $\geq 50\%$ in the population were used for population structure and LD analyses. Population structure was assessed using ADMIXTURE software (Alexander et al. 2009). The number of genetic clusters ($K$) was predefined as 1–10 to explore the population structure of the tested accessions. This analysis provided maximum likelihood estimates of the proportion of each sample derived from each of the $K$ populations. To determine the mapping resolution for GWAS, an LD analysis was performed for upland cotton accessions using the whole set of polymorphic SNPs. Pair-wise LD between markers was calculated as the squared correlation coefficient ($R^2$) of alleles using GAPIT software (Lipka et al. 2012).

Genome-wide association mapping
The GWAS was performed using three algorithms: GLM and MLM in the TASSEL version 3.0 software package (Bradbury et al. 2007) and FaST-LMM (Lippert et al. 2011) (http://www.nature.com/naturemethods/). Manhattan plots were drawn using the R package qqman (Turner et al. 2014). By referring to the method used by Cai et al. (2017) and Sun et al. (2017), significance levels
of association were estimated using an adjusted $P$ value of $1/n$ after the Bonferroni correction in this study, where $n$ was the total number of SNPs used in the association mapping. The LD heatmap surrounding the peak SNPs of GWAS for early-maturation traits were generated using the R software package LDheatmap.

**Candidate gene annotation and prediction**

To identify potential candidate genes for early maturity, the method of Sun et al. (2017) was followed. Genes that were located within the LD decay distance upstream and downstream of peak SNPs (the most significant SNPs with a maximum of $-\log_{10} P$ values) were identified. The gene annotations were obtained from several databases for non-redundant protein sequences (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/) (Altschul et al. 1997), a gene ontology analysis (http://www.geneontology.org/) (Ashburner et al. 2000), the Cluster of Orthologous Groups of proteins (http://www.ncbi.nlm.nih.gov/COG/) (Tatusov et al. 2000), and the Kyoto Encyclopedia of Genes and Genomes (ftp://ftp.genome.jp/pub/kegg/) (Kanehisa et al. 2004).

**ACKNOWLEDGEMENTS**

This research was supported by the National Natural Science Foundation of China (31671743), the Innovative Talent Support Program of Science and Technology of the Henan Institute of Higher Learning (16HASTIT014), and the Agricultural Scientific and Technological Project of Eighth Division (2013NY01). The authors would like to thank Dr. Xiongming Du of the Cotton Research Institute, Chinese Academy of Agriculture Sciences, for providing parts of upland cotton germplasm resources used in the present study.

**AUTHOR CONTRIBUTIONS**

C.L. designed the experiments. Y.W. performed most of the experiments and analyzed the data. The other authors assisted with the experiments. C.L. and Y.W. wrote the manuscript.
REFERENCES


Li C, Dong Y, Zhao T, Zhao T, Li L, Li C, Yu E, Mei L, Daud MK, He Q, Chen J, Zhu S (2016b) 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** 7: 218


maturity via GWAS in upland cotton. **BMC Genom** 17: 687


Turner SD (2014) qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots, Biorxiv


SUPPORTING INFORMATION

Figure S1. Manhattan plots of a genome-wide association study (GWAS) for SP
The significant SNPs were detected repeatedly in at least two environments or two algorithms.

Figure S2. Manhattan plots of a genome-wide association study (GWAS) for BP
The significant SNPs were detected repeatedly in at least two environments or two algorithms.

Figure S3. Manhattan plots of a genome-wide association study (GWAS) for FBP
The significant SNPs were detected repeatedly in at least two environments or two algorithms.

Figure S4. Manhattan plots of a genome-wide association study (GWAS) for GP
The significant SNPs were detected repeatedly in at least two environments or two algorithms.

Table S1. Names and ecological sources of the 169 upland cotton accessions
Table S2. Significant SNPs associated with early-maturation traits detected in four environments and BLUPs and three algorithms

Table S3. Proposed candidate genes for cotton early maturity

Figure legends

Figure 1. Histogram of early-maturation traits investigated in four environments
E1, E2, E3, and E4 indicate the four environments 2012Xinxiang, 2013Xinxiang, 2012Shihezi, and 2013Shihezi, respectively. SP, seeding period; BP, bud period; FBP, flower and boll period; GP, growth period.

Figure 2. Single-nucleotide polymorphism (SNP) distributions on the 26 chromosomes of upland cotton
A1–A13 and D1–D13 on the vertical axis represent the serial numbers of the 26 chromosomes; the horizontal axis shows chromosome length; the window size is 1 Mb.

Figure 3. Population structure and linkage disequilibrium (LD) decay of upland cotton accessions
(A) Total panel was grouped into 7 subpopulations (K = 7), designated S1–S7, respectively; each cotton line is represented by a single vertical line, and each color represents one subpopulation. (B) Estimated cross-validation (CV) errors in the data calculated for K, ranging from 1 to 10. (C) Genome-wide average LD decay estimated in the AD genome (blue), A subgenome (red), and D subgenome (green).

Figure 4. Manhattan plot and LD heatmap surrounding peak SNPs identified by GWAS for early-maturation traits in cotton
(A) Seedling period (SP) on chromosome D1. (B, C) Bud period (BP) on chromosomes A6 and A7. (D, E) Flower and boll period (FBP) on
chromosomes A8, D2, and D9. (F) Growth period (GP) on chromosome D9. Significant SNPs ($P = 1/49,650$; i.e., $-\log_{10} P > 4.69$) are marked with blue dashed lines; the pair-wise LD between SNPs is indicated as the squared correlation coefficient ($R^2$) value, which is indicated by the color bar; arrows indicate positions of the peak SNPs.

**Figure 5. Functional classification of the candidate genes within 400 kb upstream and downstream of the peak SNPs**
Table 1. Descriptive statistics and ANOVA for early-maturation traits across four environments

<table>
<thead>
<tr>
<th>Trait</th>
<th>Env</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>CV (%)</th>
<th>F value</th>
<th>Genotype</th>
<th>Env</th>
<th>G×E</th>
<th>H² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP (d)</td>
<td>E1</td>
<td>27.00</td>
<td>40.50</td>
<td>34.84</td>
<td>2.89</td>
<td>8.31</td>
<td>11.24***</td>
<td>599.76***</td>
<td>1.71***</td>
<td>64.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>33.00</td>
<td>43.00</td>
<td>35.44</td>
<td>2.40</td>
<td>6.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>23.50</td>
<td>35.00</td>
<td>29.67</td>
<td>2.66</td>
<td>8.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>31.50</td>
<td>40.00</td>
<td>35.40</td>
<td>1.62</td>
<td>4.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP (d)</td>
<td>E1</td>
<td>18.00</td>
<td>29.00</td>
<td>22.96</td>
<td>1.78</td>
<td>7.75</td>
<td>3.62***</td>
<td>591.23***</td>
<td>1.33**</td>
<td>61.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>21.50</td>
<td>32.50</td>
<td>25.05</td>
<td>2.21</td>
<td>8.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>23.00</td>
<td>33.00</td>
<td>27.00</td>
<td>1.96</td>
<td>7.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>27.00</td>
<td>36.00</td>
<td>30.96</td>
<td>1.67</td>
<td>5.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP (d)</td>
<td>E1</td>
<td>39.50</td>
<td>59.50</td>
<td>47.65</td>
<td>4.67</td>
<td>9.81</td>
<td>5.65***</td>
<td>651.26***</td>
<td>2.93***</td>
<td>67.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>39.00</td>
<td>51.00</td>
<td>44.35</td>
<td>2.24</td>
<td>5.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>60.00</td>
<td>77.50</td>
<td>66.89</td>
<td>4.01</td>
<td>5.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>57.00</td>
<td>71.50</td>
<td>64.23</td>
<td>2.49</td>
<td>3.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP (d)</td>
<td>E1</td>
<td>93.00</td>
<td>121.00</td>
<td>105.43</td>
<td>7.30</td>
<td>6.92</td>
<td>16.98***</td>
<td>571.02***</td>
<td>1.89**</td>
<td>78.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>95.00</td>
<td>119.00</td>
<td>104.85</td>
<td>4.15</td>
<td>3.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>110.50</td>
<td>137.00</td>
<td>123.48</td>
<td>6.36</td>
<td>5.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>123.00</td>
<td>139.50</td>
<td>130.58</td>
<td>3.81</td>
<td>2.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E1, E2, E3, and E4 indicate the four environments 2012Xinxiang, 2013Xinxiang, 2012Shihezi, and 2013Shihezi, respectively. SP, seeding period; BP, bud period; FBP, flower and boll period; GP, growth period (all values in days). SD, standard deviation; CV, coefficient of variation; G×E, genotype–environment interaction; H², broad-sense heritability; **, significant at α = 0.01 level; ***, significant at α = 0.001 level.
Table 2. Summary of SNPs in the 26 chromosomes of *Gossypium hirsutum*

<table>
<thead>
<tr>
<th>Chr</th>
<th>Size (kb)</th>
<th>No. of SNPs</th>
<th>SNP density (kb/SNP)</th>
<th>Polymorphism information value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>99884.7</td>
<td>2175</td>
<td>45.924</td>
<td>0.305</td>
</tr>
<tr>
<td>A2</td>
<td>83447.9</td>
<td>1282</td>
<td>65.091</td>
<td>0.279</td>
</tr>
<tr>
<td>A3</td>
<td>100263</td>
<td>1627</td>
<td>61.624</td>
<td>0.273</td>
</tr>
<tr>
<td>A4</td>
<td>62913.7</td>
<td>966</td>
<td>65.128</td>
<td>0.282</td>
</tr>
<tr>
<td>A5</td>
<td>92047.0</td>
<td>2433</td>
<td>37.832</td>
<td>0.300</td>
</tr>
<tr>
<td>A6</td>
<td>103170.</td>
<td>2746</td>
<td>37.571</td>
<td>0.297</td>
</tr>
<tr>
<td>A7</td>
<td>78251.0</td>
<td>1976</td>
<td>39.600</td>
<td>0.290</td>
</tr>
<tr>
<td>A8</td>
<td>103626.</td>
<td>2217</td>
<td>46.741</td>
<td>0.250</td>
</tr>
<tr>
<td>A9</td>
<td>74999.9</td>
<td>2237</td>
<td>33.527</td>
<td>0.279</td>
</tr>
<tr>
<td>A10</td>
<td>100866.</td>
<td>1856</td>
<td>54.346</td>
<td>0.270</td>
</tr>
<tr>
<td>A11</td>
<td>93316.1</td>
<td>1756</td>
<td>53.141</td>
<td>0.273</td>
</tr>
<tr>
<td>A12</td>
<td>87484.8</td>
<td>1917</td>
<td>45.636</td>
<td>0.280</td>
</tr>
<tr>
<td>A13</td>
<td>83159.5</td>
<td>2632</td>
<td>31.595</td>
<td>0.282</td>
</tr>
<tr>
<td>D1</td>
<td>61456.0</td>
<td>1758</td>
<td>34.957</td>
<td>0.283</td>
</tr>
<tr>
<td>D2</td>
<td>67284.5</td>
<td>2221</td>
<td>30.294</td>
<td>0.305</td>
</tr>
<tr>
<td>D3</td>
<td>46690.6</td>
<td>1323</td>
<td>35.291</td>
<td>0.277</td>
</tr>
<tr>
<td>D4</td>
<td>51454.1</td>
<td>972</td>
<td>52.936</td>
<td>0.280</td>
</tr>
<tr>
<td>D5</td>
<td>61933.0</td>
<td>1493</td>
<td>41.482</td>
<td>0.279</td>
</tr>
<tr>
<td>D6</td>
<td>64294.6</td>
<td>2919</td>
<td>22.026</td>
<td>0.270</td>
</tr>
<tr>
<td>D7</td>
<td>55312.6</td>
<td>2550</td>
<td>21.691</td>
<td>0.302</td>
</tr>
<tr>
<td>D8</td>
<td>65894.1</td>
<td>2146</td>
<td>30.705</td>
<td>0.309</td>
</tr>
<tr>
<td>D9</td>
<td>50995.4</td>
<td>2045</td>
<td>24.936</td>
<td>0.246</td>
</tr>
<tr>
<td>D1</td>
<td>63374.6</td>
<td>1638</td>
<td>38.690</td>
<td>0.287</td>
</tr>
<tr>
<td>D1</td>
<td>66087.7</td>
<td>1324</td>
<td>49.915</td>
<td>0.270</td>
</tr>
<tr>
<td>D1</td>
<td>59109.8</td>
<td>1849</td>
<td>31.968</td>
<td>0.267</td>
</tr>
<tr>
<td>D1</td>
<td>60534.3</td>
<td>1592</td>
<td>38.024</td>
<td>0.277</td>
</tr>
</tbody>
</table>
Table 3. Significant SNPs associated with early-maturation traits repeatedly detected in at least two environments or two algorithms

<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Chr.</th>
<th>Position (bp)</th>
<th>Alleles</th>
<th>GLM $-\log_{10} P$</th>
<th>GLM $R^2$</th>
<th>MLM $-\log_{10} P$</th>
<th>MLM $R^2$</th>
<th>Fast-LMM $-\log_{10} P$</th>
<th>Fast-LMM $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP</td>
<td>TM47967</td>
<td>D1</td>
<td>3646257</td>
<td>[T/C]</td>
<td>4.84–5.00</td>
<td>0.18–0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>TM13730</td>
<td>A6</td>
<td>3495097</td>
<td>[T/G]</td>
<td>6.34</td>
<td>0.14</td>
<td>5.43</td>
<td>0.16</td>
<td>4.84</td>
<td>0.18–0.20 E2,</td>
</tr>
<tr>
<td></td>
<td>TM13732</td>
<td>A6</td>
<td>3509639</td>
<td>[T/A]</td>
<td>5.30–7.11</td>
<td>0.12–0.16</td>
<td>5.85</td>
<td>0.18</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20788</td>
<td>A7</td>
<td>66292784</td>
<td>[T/G]</td>
<td>4.88</td>
<td>0.11</td>
<td>4.78</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20831</td>
<td>A7</td>
<td>66784327</td>
<td>[T/C]</td>
<td>5.00</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20860</td>
<td>A7</td>
<td>67183689</td>
<td>[T/C]</td>
<td>5.15</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20868</td>
<td>A7</td>
<td>67275473</td>
<td>[T/C]</td>
<td>5.29</td>
<td>0.10</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20888</td>
<td>A7</td>
<td>67418643</td>
<td>[A/C]</td>
<td>4.87</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20903</td>
<td>A7</td>
<td>67533709</td>
<td>[T/G]</td>
<td>5.48</td>
<td>0.13</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20918</td>
<td>A7</td>
<td>67687356</td>
<td>[G/C]</td>
<td>5.42</td>
<td>0.12</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20937</td>
<td>A7</td>
<td>67860203</td>
<td>[A/G]</td>
<td>5.72</td>
<td>0.13</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20940</td>
<td>A7</td>
<td>67875314</td>
<td>[A/G]</td>
<td>4.80</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td></td>
<td>TM20946</td>
<td>A7</td>
<td>68002395</td>
<td>[T/C]</td>
<td>4.81</td>
<td>0.11</td>
<td>4.85</td>
<td>0.20</td>
<td>4.85</td>
<td>0.20 BLUP</td>
</tr>
<tr>
<td>FBP</td>
<td>TM28405</td>
<td>A8</td>
<td>70830363</td>
<td>[A/C]</td>
<td>4.78</td>
<td>0.10</td>
<td>4.80</td>
<td>0.12</td>
<td>4.80</td>
<td>0.12 E3</td>
</tr>
<tr>
<td></td>
<td>TM28428</td>
<td>A8</td>
<td>70964312</td>
<td>[T/C]</td>
<td>4.96</td>
<td>0.07</td>
<td>5.44</td>
<td>0.11</td>
<td>5.44</td>
<td>0.11 E3</td>
</tr>
<tr>
<td></td>
<td>TM50280</td>
<td>D2</td>
<td>900486</td>
<td>[C/G]</td>
<td>4.97</td>
<td>0.08</td>
<td>4.80</td>
<td>0.07</td>
<td>4.80</td>
<td>0.07 E1</td>
</tr>
<tr>
<td></td>
<td>TM50282</td>
<td>D2</td>
<td>920715</td>
<td>[A/T]</td>
<td>5.10</td>
<td>0.08</td>
<td>4.96</td>
<td>0.07</td>
<td>4.96</td>
<td>0.07 E1</td>
</tr>
<tr>
<td></td>
<td>TM50283</td>
<td>D2</td>
<td>924620</td>
<td>[T/C]</td>
<td>5.95</td>
<td>0.11</td>
<td>5.31</td>
<td>0.09</td>
<td>5.31</td>
<td>0.09 E1</td>
</tr>
<tr>
<td></td>
<td>TM72552</td>
<td>D9</td>
<td>43446639</td>
<td>[A/G]</td>
<td>5.10–9.28</td>
<td>0.08–0.13</td>
<td>6.32</td>
<td>0.16</td>
<td>6.32</td>
<td>0.16 E1</td>
</tr>
<tr>
<td></td>
<td>TM72554</td>
<td>D9</td>
<td>43478716</td>
<td>[T/C]</td>
<td>7.87</td>
<td>0.12</td>
<td>5.23</td>
<td>0.14</td>
<td>5.23</td>
<td>0.14 E1</td>
</tr>
<tr>
<td></td>
<td>TM72555</td>
<td>D9</td>
<td>43483274</td>
<td>[T/C]</td>
<td>4.88–7.92</td>
<td>0.09–0.12</td>
<td>5.43</td>
<td>0.15</td>
<td>5.43</td>
<td>0.15 E1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM72558 D9 43495825</td>
<td>A/T</td>
<td>8.15</td>
<td>0.13</td>
<td>E1</td>
<td>5.78</td>
<td>0.17</td>
<td>E1</td>
<td>6.83</td>
<td>0.25</td>
<td>E1</td>
</tr>
<tr>
<td>TM72559 D9 43499161</td>
<td>A/G</td>
<td>7.75</td>
<td>0.12</td>
<td>E1</td>
<td>5.02</td>
<td>0.14</td>
<td>E1</td>
<td>6.3</td>
<td>0.23</td>
<td>E1</td>
</tr>
<tr>
<td>TM72569 D9 43617803</td>
<td>A/C</td>
<td>6.78</td>
<td>0.10</td>
<td>E1</td>
<td>4.73</td>
<td>0.11</td>
<td>E1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM72552 D9 43446639</td>
<td>A/G</td>
<td>6.04–6.46</td>
<td>0.08–0.09</td>
<td>BLUP</td>
<td>4.76–4.9</td>
<td>0.10–0.11</td>
<td>E1</td>
<td>4.84</td>
<td>0.16</td>
<td>E1</td>
</tr>
<tr>
<td>TM72554 D9 43478716</td>
<td>T/C</td>
<td>5.19–5.77</td>
<td>0.08–0.09</td>
<td>E1, BLUP</td>
<td>5.98, 5.93</td>
<td>0.18–0.19</td>
<td>E1, BLUP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM72555 D9 43483274</td>
<td>T/C</td>
<td>5.48–5.80</td>
<td>0.08–0.09</td>
<td>E1, BLUP</td>
<td></td>
<td></td>
<td>E1, BLUP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM72558 D9 43495825</td>
<td>A/T</td>
<td>5.72</td>
<td>0.09</td>
<td>E1</td>
<td>4.80</td>
<td>0.13</td>
<td>E1</td>
<td>6.08, 6.00</td>
<td>0.19–0.20</td>
<td>E1, BLUP</td>
</tr>
<tr>
<td>TM72559 D9 43499161</td>
<td>A/G</td>
<td>4.99–5.69</td>
<td>0.07–0.09</td>
<td>E1, BLUP</td>
<td>5.62, 5.61</td>
<td>0.16–0.18</td>
<td>E1, BLUP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

E1, E2, E3, E4 and BLUP indicate 2012Xinxiang, 2013Xinxiang, 2012Shihezi, 2013Shihezi, and best linear unbiased predictor, respectively. SP, seeding period; BP, bud period; FBP, flower and boll period; GP, growth period; GLM, general linear model; MLM, mixed linear model; FaST-LMM, factored spectrally transformed linear mixed model; $R^2$, phenotypic variation explained by SNP marker.
Figure. 1

Figure. 2
Figure. 3
Figure. 4
Figure 5

A: Transport and metabolism
B: Transcription regulation and posttranslational modification
C: Cell wall/membrane/cellular component
D: Defense mechanisms
E: Energy production and conversion
F: Senescence
G: RNA processing and modification
H: Signal transduction mechanisms
I: Translation, ribosomal structure and biogenesis
J: Replication, recombination and repair
K: General function prediction only
L: Putative and uncharacterized protein
M: Other function
N: Unknown function