JIPB Journal of Integrative Plant Biology

Research Article
https://doi.org/10.1111/jipb.13409

GhAP1-D3 positively regulates flowering time and early maturity with no yield and fiber quality penalties in upland cotton

Caixiang Wang[1†], Juanjuan Liu[1†], Xiaoyu Xie[1], Ji Wang[2], Qi Ma[3], Pengyun Chen[2], Delong Yang[1], Xiongfeng Ma[4*], Fushun Hao[2*] and Junji Su[1,3*]

1. State Key Laboratory of Aridland Crop Science, College of Life Science and Technology, Gansu Agricultural University, Lanzhou 730070, China
2. State Key Laboratory of Cotton Biology, College of Life Science, Henan University, Kaifeng 475004, China
3. Cotton Research Institute, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi 832000, China
4. State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, China

†These authors contributed equally to this study.

*Correspondences: Junji Su (sujj@gsau.edu.cn, Dr. Su is fully responsible for the distributions of all materials associated with this article); Fushun Hao (haofsh@henu.edu.cn); Xiongfeng Ma (maxiongfeng@caas.cn)

Caixiang Wang

Junji Su

ABSTRACT

Flowering time (FTi) is a major factor determining how quickly cotton plants reach maturity. Early maturity greatly affects lint yield and fiber quality and is crucial for mechanical harvesting of cotton in northwestern China. Yet, few quantitative trait loci (QTLs) or genes regulating early maturity have been reported in cotton, and the underlying regulatory mechanisms are largely unknown. In this study, we characterized 152, 68, and 101 loci that were significantly associated with the three key early maturity traits—FTi, flower and boll period (FBP) and whole growth period (WGP), respectively, via four genome-wide association study methods in upland cotton (*Gossypium hirsutum*). We focused on one major early maturity-related genomic region containing three single nucleotide polymorphisms on chromosome D03, and determined that *GhAP1-D3*, a gene homologous to *Arabidopsis thaliana APETALA1 (AP1)*, is the causal locus in this region. Transgenic plants overexpressing *GhAP1-D3* showed significantly early flowering and early maturity without penalties for yield and fiber quality compared to wild-type (WT) plants. By contrast, the mutant lines of *GhAP1-D3* generated by genome editing displayed markedly later flowering than the WT. GhAP1-D3 interacted with GhSOC1 (SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1), a pivotal regulator of FTi, both *in vitro* and *in vivo*. Changes in *GhAP1-D3* transcript levels clearly affected the expression of multiple key flowering regulatory genes. Additionally, DNA hypomethylation and high levels of H3K9ac affected strong expression of *GhAP1-D3* in early-maturing cotton cultivars. We propose that epigenetic modifications modulate *GhAP1-D3* expression to positively regulate FTi in cotton through interaction of the encoded GhAP1 with GhSOC1 and affecting the transcription of multiple flowering-related genes. These findings may also lay a foundation for breeding early-maturing cotton varieties in the future.

Keywords: upland cotton, early maturity, genome-wide association studies, QTL, AP1, flowering time

Wang, C., Liu, J., Xie, X., Wang, J., Ma, Q., Chen, P., Yang, D., Ma, X., Hao, F., and Su, J. (2023). GhAP1-D3 positively regulates flowering time and early maturity with no yield and fiber quality penalties in upland cotton. J. Integr. Plant Biol. **00:** 1–18.

INTRODUCTION

Upland cotton (*Gossypium hirsutum* L.) is the dominant cultivated tetraploid cotton (*Gossypium* spp.) species and an economically essential fiber crop in China (Mao, 2013). In the regions of Xinjiang and Gansu in northwest China, cotton planting area reached 2.52 million hectares and produced 5.16 million tons of lint cotton in 2021, accounting for 83.3% of the total cultivation area and 90.0% of the cotton yield nationally, as reported by the National Bureau of Statistics of China. However, the temperature profile in these regions is not ideal for cotton, with generally cool temperatures that greatly restrict seedling growth and boll development, thus reducing fiber yield and quality (Li et al., 2000; Chen et al., 2014; Feng et al., 2017). Accordingly, planting early-maturing varieties is crucial for cotton production in these regions.

The key indicators reflecting when upland cotton reaches maturity are flowering time (FTi), flower and boll period (FBP) and whole growth period (WGP). These traits are regulated by multiple genes with varying effect sizes (Fan et al., 2006 ; Li et al., 2013a; Su et al., 2016). For instance, numerous quantitative trait loci (QTLs) have been identified for maturity traits in segregating populations and by linkage mapping in upland cotton (Ren et al., 2002; Fan et al., 2006; Zhang et al., 2009; Lacape et al., 2013; Li et al., 2013a; Jia et al., 2016). Many single nucleotide polymorphisms (SNPs) associated with cotton early-maturity indicators have also been detected in natural populations by genome-wide association studies (GWAS) (Su et al., 2016; Li et al., 2018, 2021a; Ma et al., 2018; Shen et al., 2019). However, the molecular mechanisms determining early maturity and early flowering in upland cotton are still largely unknown, with the identity of QTLs and genes associated with FTi remaining to be determined.

Flowering is controlled by at least six major pathways: the vernalization, photoperiod, ambient temperature, aging, autonomous and gibberellin pathways (Andrés and Coupland, 2012; Yamaguchi, 2022). Typically, floral signals derived from each pathway converge to activate several common key floral integrators such as FLOWERING LOCUS T (FT), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and AGAMOUS-LIKE 24 (AGL24). These integrators further activate the expression of floral meristem identity genes such as *APETALA1* (*AP1*), *LEAFY* (*LFY*), *SEPALLATA3* (*SEP3*), and *AGAMOUS* (*AG*), leading to the transition from vegetative to reproductive growth (Khan et al., 2014; Bluemel et al., 2015; Chen et al., 2020; Freytes et al., 2021). *AP1* is a key gene promoting flower bud differentiation and floral organ identity in *Arabidopsis* (Wagner et al., 1999; Pelaz et al., 2001; Winter et al., 2015; Goslin et al., 2017). *AP1* homologs in upland cotton, including *GhMADS22*, *GhMADS42* and *GhAP1.7* (*Gh_D03G0922*), have been identified over the past few years (Zhang et al., 2013, 2016; Su et al., 2018; Cheng et al., 2021b). Overexpression of *GhMADS22* or *GhMADS42* in *Arabidopsis* accelerated flowering (Zhang et al., 2013, 2016). Likewise, *GhAP1.7* overexpression in *Arabidopsis* caused precocious flowering, whereas knocking down *GhAP1.7* transcript levels by virus-induced gene silencing (VIGS) in cotton resulted in delayed flowering and taller plants (Su et al., 2018; Cheng et al., 2021b). Despite these reports, the exact roles of GhAP1 members have not been elucidated in cotton, and their underlying mechanisms remain to be unveiled.

Epigenetic modifications such as histone modifications and DNA methylation play essential roles in plant growth and development, including flowering, by governing gene expression (Gehring, 2019; Yamaguchi, 2022). In *Arabidopsis*, changes in the levels of trimethylation of histone H3 lysine 4 (H3K4me3), H3K27me3, H3K36me3, and/or histone acetylation greatly affect the transcription of *FT*, *LFY*, *FLC* (*FLOWERING LOCUS C*, a crucial repressor of flowering) and *AG*, further controlling floral transition and flower development (Jeong et al., 2015; Zhao et al., 2019; He et al., 2020; Pelayo et al., 2021; Yamaguchi, 2022). Moreover, chromatin factors regulate *AP1* expression during flower development in *Arabidopsis* by modulating the deposition or removal of such histone marks. For example, a loss-of-function mutation in the gene encoding the histone methyltransferase HOMOLOG of TRITHORAX (ATX1, also named SET DOMAIN GROUP27 (SDG27)) that catalyzes H3K4me3 deposition causes the homeotic conversion of floral organs and reduces expression of floral homeotic genes, including *AP1* (Alvarez-Venegas et al., 2003; Guo et al., 2015). Alterations of DNA methylation levels also affect FTi. For instance, loss of DNA METHYLTRANSFERASE 1 (MET1) function diminished global cytosine methylation levels, resulting in hypomethylation at *FWA* (*FLOWERING WAGENINGEN*) and a late flowering phenotype, in *Arabidopsis* (Soppe et al., 2000; Kankel et al., 2003). However, whether histone modifications and DNA methylation modulate the expression of *GhAP1* during cotton flowering is currently unknown.

In this work, we detected a genomic region strongly associated with FTi on chromosome 16 (D03) of upland cotton and identified the candidate gene *GhAP1-D3* regulating early maturity within this region by transcriptome data analysis and reverse transcription quantitative polymerase chain reaction (RT-qPCR) experiments. Moreover, overexpression of *GhAP1-D3* in upland cotton caused early maturity without yield and fiber quality penalties. We determined that GhAP1-D3 interacts with the floral regulator GhSOC1. We also showed that DNA hypomethylation and high levels of H3K9ac in early-maturing cotton cultivars were accompanied by strong *GhAP1-D3* expression. These results will provide a basis for breeding new early-maturing varieties of cotton in the future.

RESULTS

Flowering time, FBP, and WGP show extensive variation in a natural population of upland cotton

We surveyed the extent of phenotypic variation in FTi, FBP, and WGP across a population of 315 upland cotton accessions. Accordingly, we grew all accessions in the field in

three test environments: AY-14 (Anyang, Henan China in 2014), SHZ-14 (Shihezi, Xinjiang China in 2014), and SHZ-15 (Shihezi, Xinjiang China in 2015), as previously reported (Su et al., 2020). We observed extensive variation in FTi, FBP, and WGP, ranging from 66.7 to 83.0 d for FTi, 46.2 to 72.2 d for FBP and 114.1 to 148.4 d for WGP. For all three traits, the phenotypic values conformed to a normal distribution, as supported by a Shapiro–Wilk test (Figure 1A), demonstrating that the three early maturity-related traits likely involve multiple loci. Additionally, broad-sense heritability was high, with values of 73.7%, 71.9%, and 78.7% for FTi, FBP, and WGP, respectively (Table S1). Pearson's correlation analysis detected a significant and positive correlation between traits in the natural population (Table S2). Further, the analysis of variance (ANOVA) values revealed that the three early-maturity traits were significantly affected by genotype, environment and genotype × environment interaction (Table S1). These results suggest that the three early maturity-related traits are controlled by multiple loci, the environment and its interaction with the cotton genome.

Genome-wide association studies identify three SNPs associated with all three early-maturity traits

In our previous study, we obtained 13,391 high-quality SNPs using a specific-locus amplified fragment sequencing (SLAF-seq) method in an association panel, which we grouped into 9,244 SNP linkage disequilibrium blocks (SNPLDBs) with multiple haplotypes/alleles to develop molecular markers. Based on the results from principal component analysis and clustering analysis of the above SNPLDBs, we divided the 315 upland cotton accessions into two groups (Su et al., 2020). We also previously estimated the average distance of linkage disequilibrium (LD) decay to be about 500 kb for the natural population (Su et al., 2020). Here, we implemented three single-locus GWAS methods (general linear model (GLM), mixed linear model (MLM) and fixed and random model circulating probability unification (FarmCPU)) to detect significant associations between SNPs and the phenotypic values obtained in each of the three environments, as well as with their best linear unbiased prediction (BLUP) values. We identified 93, two and three significant SNPs associated with the BLUP values of FTi via the GLM-GWAS, MLM-GWAS, and FarmCPU-GWAS methods, respectively (Figure S1; Table S3). We also detected five, one and five significant SNPs associated with the BLUP values of FBP by GLM-GWAS, MLM-GWAS, and FarmCPU-GWAS, respectively (Figure S2; Table S4). Finally, we identified 34, one, and three significant SNPs associated with the BLUP values of WGP via GLM-GWAS, MLM-GWAS, and FarmCPU-GWAS, respectively (Figure S3; Table S5). Interestingly, we noticed the concentration of 38, five and 23 SNPs on chromosome 16 (D03) in Manhattan plots of the SNPs significantly associated with the BLUP values of FTi, FBP, and WGP obtained by all three GWAS methods (Figure 1B). Notably, we detected three stable-association SNPs (D03_37952328, D03_38393621, and D03_39049001) with high $-\log_{10}$ (*P*-values) using the BLUP values from three environments and the FTi, FBP, and WGP values in at least two environments.

To confirm that these SNPs are significantly associated with FTi, FBP, and WGP, we performed a restricted two-stage multi-locus and multi-allele GWAS (RTM-GWAS) analysis using the 9,244 SNPLDBs (Su et al., 2020). We thus detected 54, 57, and 63 SNPLDBs with significant association with FTi, FBP, and WGP, respectively (Figure 1C; Table S6). Among the SNPLDBs, we determined that the three stable loci LDB_16_37952328, LDB_16_38393621, and LDB_16_39049001 on chromosome 16 (D03) are associated with their BLUP values and the FTi, FBP, and WGP values from three environments, with high $\log_{10}$ (*P*-values). These results suggest that the three SNPs (D03_37952328, D03_38393621, and D03_39049001) define a major genomic region in which the genotype influences the three early maturity-related traits.

Delineating a stable genomic region associated with early maturity-related traits

We selected the three SNPs to identify favorable alleles determining early maturity, given their simultaneous association with all three traits in the above GWAS analyses. The three SNPs presented the allelic variations C/T, A/C, and A/T, respectively (Figure 1D), and were closely linked on chromosome D03. We detected four main haplotypes across these three SNPs (CCA, TAT, TCA, and CAT). In the natural population, accessions with the TAT haplotype had significantly smaller values for FTi, FBP, and WGP than those with the CCA haplotype (Figure 1D), indicating that the TAT haplotype determines early maturity. The three tightly linked SNPs on chromosome D03 mapped within an LD block defining a 2.10-Mb genomic segment (37.45–39.55 Mb) (Figure S4). The results indicate that this 2.10-Mb genome region may harbor a locus that regulates early maturity in cotton.

GH_D03G1119 acts as a candidate gene for regulation of early-maturity traits

Based on the upland cotton reference genome v2.1 (Hu et al., 2019), the 2.10-Mb genomic region contained 76 annotated genes (Table S7; Figure S4B). Of the three associated SNPs, only D03_39049001 was within the non-coding region (intron) of a gene, *GH_D03G1119* (Figure S4B), suggesting that D03_39049001 might not be a functional variation of *GH_D03G1119*.

To examine the functional variants of *GH_D03G1119*, we extracted its coding sequence (CDS) and promoter sequence from two early-maturing (ZMS50 and ZMS74) and two late-maturing accessions (GXM11 and STS458) from whole-genome resequencing data. We noticed one non-synonymous SNP (D03_39050639) located in the fifth exon of the CDS that distinguishes early-maturing cultivars, which carry a T allele, from late-maturing cultivars, which harbor a C allele. We confirmed this allelic variation (T/C) by Sanger sequencing of the four cultivars (Figure S5A). The SNP mapped to the sequence encoding a conserved domain of the K-box in AP1, resulting in an

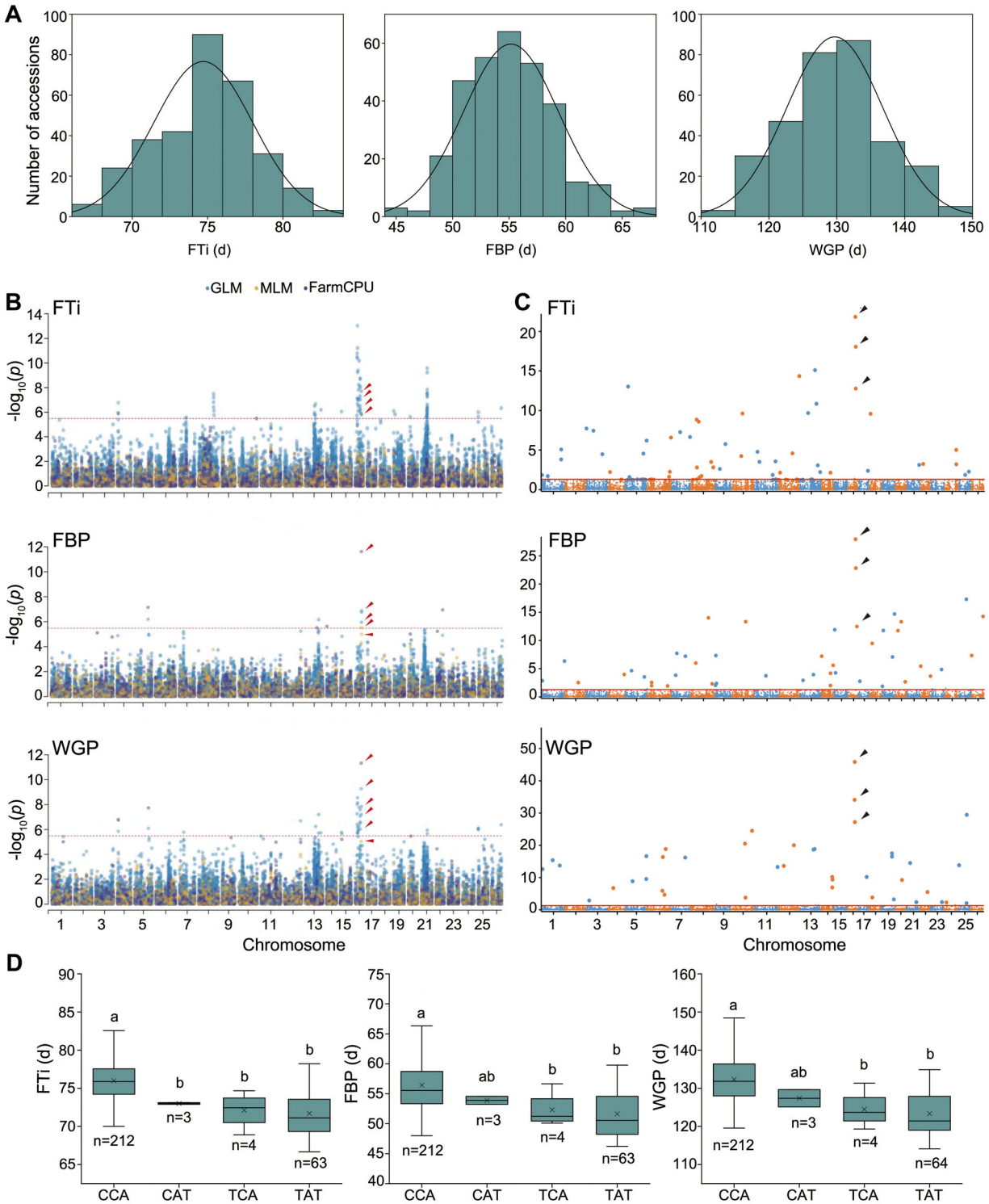

Figure 1. Genome-wide association studies (GWAS) for three early maturity-related traits

(A) Phenotypic variation in three traits related to early maturity (FTi, FBP, and WGP) among 315 upland cotton accessions. FTi, flowering time; FBP, flower and boll period; WGP, whole growth period. **(B)** Manhattan plots for FTi, FBP, and WGP by the general linear model- (GLM), mixed linear model- (MLM) and fixed and random model circulating probability unification (FarmCPU)-genome-wide association study (GWAS) methods under the single-locus model. Red arrowheads show three single nucleotide polymorphisms (SNPs) (D03_37952328, D03_38393621, and D03_39049001) in the three GWAS methods. **(C)** Manhattan plots for FTi, FBP, and WGP at the second stage of restricted two-stage multi-locus and multi-allele GWAS (RTM-GWAS) under the multi-locus model. Black arrowheads indicate three SNPs (D03_37952328, D03_38393621 and D03_39049001). **(D)** FTi, FBP, and WGP among the four haplotype groups for the three SNPs. Different lowercase letters indicate significant differences by one-way analysis of variance and Tukey's Honestly Significant Difference test ($P < 0.05$).

amino acid change from a phenylalanine to a leucine (Figure S5B). Importantly, we detected no sequence variation in the CDS and promoter region of *GH_A03G0856*, a homolog of *GH_D03G1119* on chromosome A03, between early- and late-maturing cultivars. Moreover, the CDS sequence of *GH_A03G0856* was identical to that of *GH_D03G1119* in the late-maturing cultivars (GXM11 and STS458). We hypothesized from these results that *GH_D03G1119* might be the candidate gene affecting early maturity in upland cotton.

GH_D03G1119 is differentially expressed between early-maturity and late-maturity cultivars

To confirm *GH_D03G1119* as the candidate gene associated with early-maturity traits, we analyzed two previously published transcriptome deep sequencing (RNA-seq) datasets (Cheng et al., 2021a) from two upland cotton varieties: the early-maturing cultivar ZMS50 with the TAT haplotype and the late-maturing cultivar GXM11 with the CCA haplotype. We determined that 57 out of the 76 genes were expressed in young leaves and shoot apical meristems at six different growth stages (cotyledon and 1–5 true leaf stages). Of these 57 genes, two (*GH_D03G1119* and *GH_D03G1123*) were expressed four-fold more in ZMS50 than in GXM11. By contrast, the remaining 55 genes were expressed similarly in ZMS50 and GXM11 across all six growth stages tested in this dataset. We noticed that *GH_D03G1119* expression levels were 4.9 times higher in ZMS50 than in GXM11 in young leaves and shoot apical meristems at the third true leaf stage (Figure 2A). These results indicate that *GH_D03G1119* and *GH_D03G1123* are differentially expressed genes (DEGs) between early-maturing and late-maturing cultivars, and may represent good candidates for genes underlying the early-maturity trait. In addition, we exploited another RNA-seq dataset encompassing 22 samples from the upland cotton line TM-1 (http://cotton.zju.edu.cn), and established that *GH_D03G1119* was highly expressed in leaves, stems and pistils, but not expressed in ovules or fiber tissues. Notably, *GH_D03G1123* was highly expressed in all 22 tissues, with the highest relative expression level in the torus (Figure 2B). To further define the candidate genes related to cotton early maturity, we carried out RT-qPCR experiments to assess the messenger RNA (mRNA) abundance

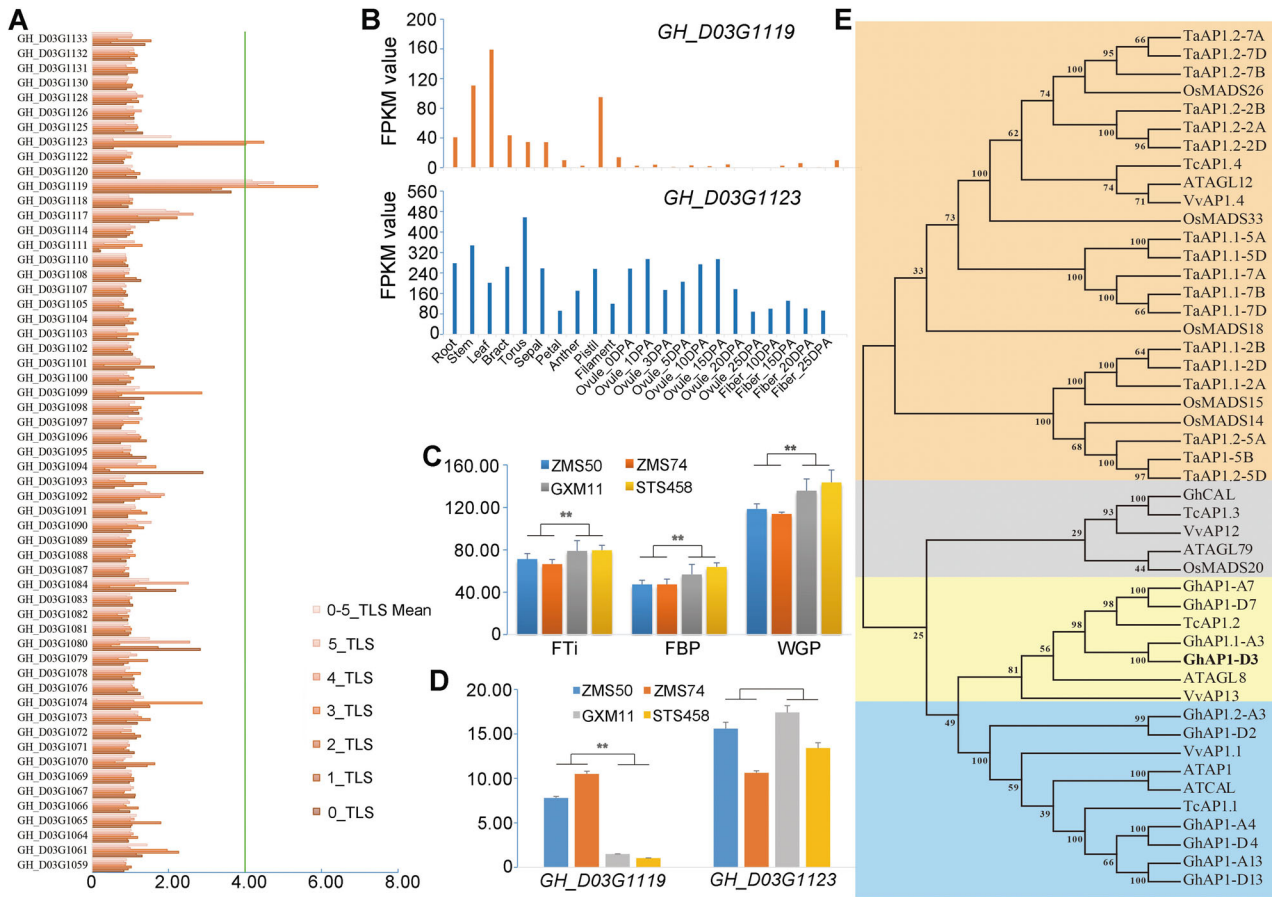

Figure 2. Identification of candidate genes for early maturity traits

(A) Relative expression levels of 57 candidate genes in young leaves and shoot apical meristems during six growth periods. **(B)** Expression patterns of the two candidate genes *GH_D03G1119* and *GH_D03G1123* across 22 upland cotton tissues. **(C)** Flowering time (FTi), flower and boll period (FBP), and whole growth period (WGP) for two early-maturing varieties (ZMS50 and ZMS74) and two late-maturing varieties (GXM11 and STS458). **(D)** Relative expression levels of the two candidate genes *GH_D03G1119* and *GH_D03G1123* in the two early-maturing varieties and the two late-maturing varieties at the three-leaf stage. **(E)** Phylogenetic tree of AP1 proteins from upland cotton (*Gossypium hirsutum*), grapevine (*Vitis vinifera*), *Arabidopsis*, cacao tree (*Theobroma cacao*), rice (*Oryza sativa*), and wheat (*Triticum aestivum*). Bootstrap values are given at each node.

of *GH_D03G1119* and *GH_D03G1123* in the two early-maturing (ZMS50 and ZMS74) and two late-maturing accessions (GXM11 and STS458) previously found to have significant differences in their FTi, FBP and WGP (Figure 2C). We determined that *GH_D03G1119*, but not *GH_D03G1123*, was more highly expressed in the two early-maturing varieties (ZMS50 and ZMS74) with the TAT haplotype than in the two late-maturing varieties (GXM11 and STS458) with the CCA haplotype (Figure 2D). Collectively, our data suggest that GH_D03G1119, which encodes a transcription factor with high similarity to the *Arabidopsis* AP1, may be an important regulator of early maturity.

Phylogenetic analysis of the GhAP1 protein family

The upland cotton genome encodes 11 *GhAP1* members (Zhang et al., 2020). Of the corresponding 11 genes, only *GH_D03G1119* mapped to chromosome D03; we named this gene *GhAP1-D3*. To examine the phylogenetic relationship of GhAP1-D3 to other AP1-like proteins, we constructed a phylogenetic tree using AP1 proteins from *G. hirsutum* (11 genes), grapevine (*Vitis vinifera*; four genes), *Arabidopsis* (five genes), cacao tree (*Theobroma cacao*; four genes), rice (*Oryza sativa*; five genes), and wheat (*Triticum aestivum*; 17 genes) using the neighbor-joining method. GhAP1-D3 showed a close evolutionary relationship with GhAP1.1-A03 (GH_A03G0856), TcAP1-1 (EOY22135), GhAP1-7D (GH_D07G0744), GhAP1-7A (GH_A07G0749), *Arabidopsis* AGAMOUS-LIKE 8 (AGL8, At5g60910), and VvAP1.3 (VIT_17s0000g04990) (Figure 2E).

Overexpressing *GhAP1-D3* results in early flowering

To further investigate the role of *GhAP1-D3* in regulating FTi, we cloned the *GhAP1-D3* complementary DNA (cDNA) from the early-maturing cotton cultivar ZMS50. *GhAP1-D3* has an open reading frame (ORF) of 696 bp and encodes a protein of 231 amino acids (Figure S6). We then overexpressed *GhAP1-D3* in the late-maturing cotton line Jin668 (wild-type (WT), thereafter) using *Agrobacterium tumefaciens*-mediated transformation (Jin et al., 2006). We obtained 13 transgenic lines, of which five (*GhAP1-D3*OE-1, *GhAP1-D3*OE-2, *GhAP1-D3*OE-3, *GhAP1-D3*OE-4, and *GhAP1-D3*OE-5) harbored a single-copy insertion (Figure S7). These five lines displayed early flowering phenotypes compared to WT when grown under greenhouse conditions (Figure 3A). Importantly, RT-qPCR analysis indicated that *GhAP1-D3* transcript levels were much higher in these overexpression (OE) lines than in WT (Figure 3B). The OE lines reached flower budding 7–12 d earlier than WT, and flowered 15–21 d earlier than WT (Figure 3C, D). These results indicate that GhAP1-D3 is a positive modulator of early flowering and plays an important role in early maturity in cotton.

ghap1 mutants are late flowering

To further evaluate the role of *GhAP1-D3* in regulating FTi, we generated single mutants in *GhAP1-D3* or its homolog *GhAP1-A3* (*GH_D03G1119* and *GH_A03G0856*) using clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated nuclease 9 (Cas9)-mediated gene editing. For this, we designed a pair of single guide RNAs (sgRNAs) complementary to a sequence conserved across late-maturity varieties in the coding region of *GhAP1-D3/A3* using the late-maturity line Jin668 as template (Figure 4A). We obtained 30 independent primary (T_0) plants, of which 14 harbored the transgene, based on genotyping PCR with primers specific for *NPTII* and *Cas9* (Figure S8A). Of these, we sequenced the target sites for 12 lines via Sanger sequencing. We established that all 12 lines carried mutations at both target sites, with a range of deletion events and insertion events (Figure S8B). To further investigate the editing profile of T_3 lines derived from each of these 12 mutant plants, we evaluated seven independent mutant lines by PCR and Sanger sequencing (Figure 4B, C). The *ghap1-m3*, *-m6*, *-m13*, *-m19*, and *-m27* mutants each carried a 1-bp insertion of either A or T located 60 bp downstream of the transcription start site (ATG), while the *ghap1-m3*, *-m11*, and *-m27* mutants harbored 1-bp deletions; the *ghap1-m6*, *-m19* and *-m30* mutants contained a 2-bp, 3-bp, or 10-bp deletion, respectively (Figure 4C). The two mutant lines *ghap1-m11* and *ghap1-m27* were transgene-free, as evidenced by genotyping PCR for the *Cas9* and *NPTII* genes (Figure 4B). We thus selected *ghap1-m11* and *ghap1-m27*, as well as *ghap1-m13* (which still carries the transgene), for further study. All three lines displayed a late flowering phenotype compared to WT when grown under greenhouse conditions (Figure 4D). The three mutants reached the flower budding stage 3–9 d later than WT, and flowered 4–6 d later than WT (Figure 4E, F). These results suggest that GhAP1-D3 is important for early flowering of cotton plants.

Overexpression of *GhAP1-D3* results in early maturity without fiber yield and quality penalties

To estimate the value and potential of *GhAP1-D3* in cotton breeding applications, we investigated the agronomic traits of the T_3 transgenic lines overexpressing *GhAP1-D3* in field trials at two test environments (Shihezi (SHZ) and Kuerle (KEL)) in Xinjiang in 2021. Compared to that of the WT, the FTi of the five T_3 OE transgenic lines was earlier by 4.0–7.3 and 2.0–6.3 d in SHZ and KEL, respectively; and boll-opening of OE lines occurred earlier by 2.0–6.3 and 4.0–6.7 d in SHZ and KEL, respectively. The FTi and boll-opening time of the transgenic plants were significantly earlier than those of the WT plants (Figure 5A, B). With the exception of the OE-3 line, all transgenic lines had similar FBP values to WT plants in both SHZ and KEL (Figure 5B). These data suggest that the early boll-opening time of the OE lines can be attributed to early FTi rather than to FBP. In addition, we observed no significant differences among genotypes in yield component traits such as boll weight (BW), boll number per plant (BN), lint percentage (LP) and fruit branch number (FBN) and in fiber-quality indicators such as fiber length (FL), fiber strength (FS), fiber uniformity (FU), fiber micronaire (FM) and fiber elongation (FE) (Figures 5C, D, S9). The above results indicate that overexpression of *GhAP1-D3* in upland cotton results in early maturity without imposing any penalty on yield and fiber quality.

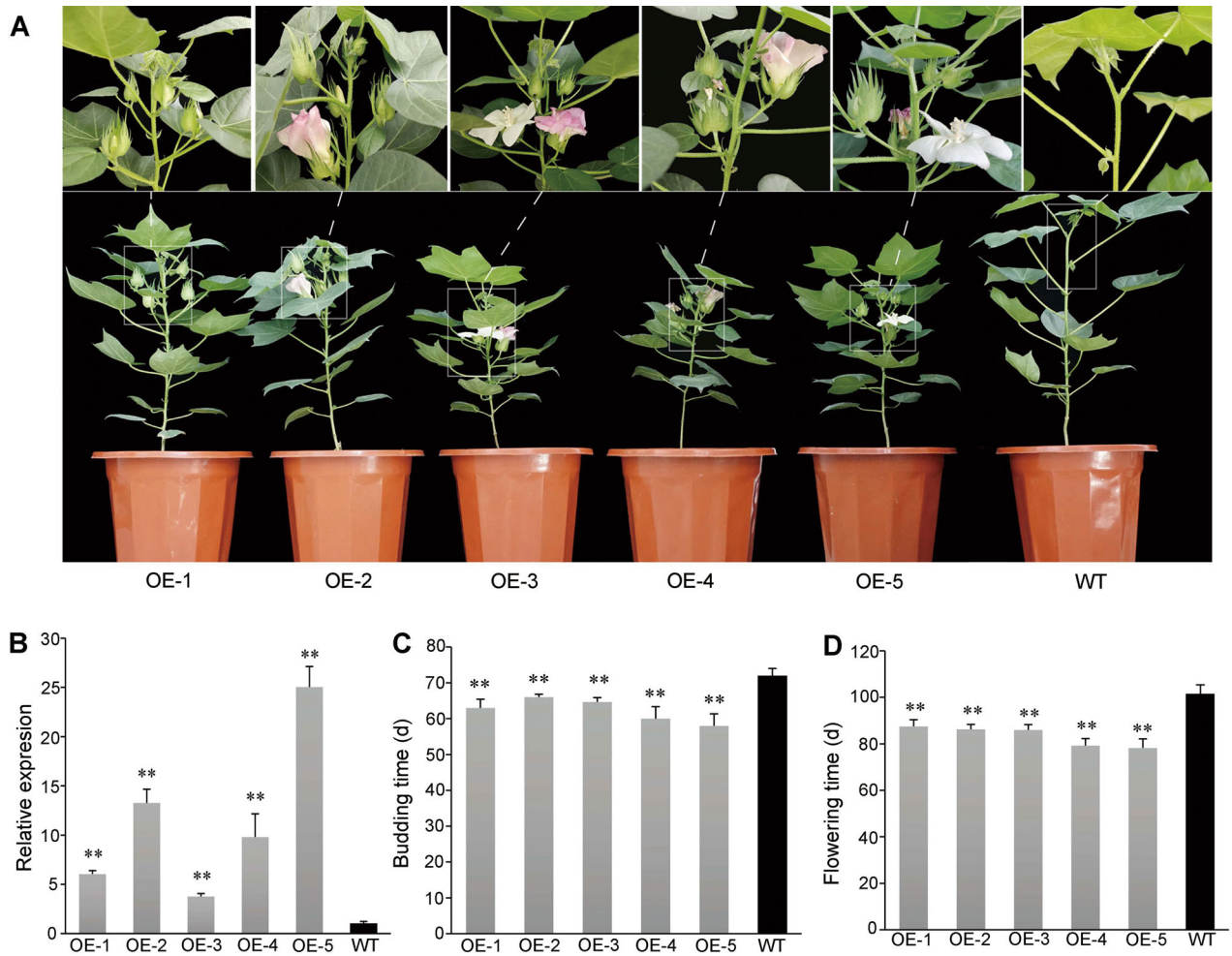

Figure 3. Performances of *GhAP1-D3* overexpressing lines (OEs) in upland cotton

(A) Flowering phenotypes. **(B)** Relative *GhAP1-D3* expression levels in the five T_3 overexpression lines and wild-type (WT) plants, as revealed by reverse transcription quantitative polymerase chain reaction (RT-qPCR). Expression levels in WT were set to 1. **(C)** Budding time. **(D)** Flowering time. Data are means ± *SD* ($n \geq 10$). $^{**}P < 0.01$ levels by Student's *t*-test between WT and OEs.

GhAP1-D3 modulates the expression of multiple key flowering genes in upland cotton

To probe how *GhAP1-D3* affects FTi in cotton, we measured the expression levels of 14 important flowering regulatory genes in the OE lines, the *ghap1* mutants, and WT plants. These genes were *GhFT* (*FLOWERING LOCUS T*), *GhGA1* (*GA REQUIRING 1*), *GhHOS1* (*HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE 1*), *GhSPL3* (*SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE 3*), *GhCAL* (*CAULIFLOWER*), *GhFLK* (*FLOWERING LOCUS KH DOMAIN*), *GhLFY* (*LEAFY*), *GhSVP* (*SHORT VEGETATIVE PHASE*), *GhMYB* (*Myeloblast*), *GhAGL24* (*AGAMOUS-LIKE 24*), *GhFCA* (*FLOWERING CONTROL LOCUS A*), *GhFVE* (*FOWERING LOCUS VE*), *GhFLM* (*FLOWERING LOCUS M*) and *GhSOC1*. Compared to those in WT, the expression levels of *GhGA1*, *GhHOS1* and *GhFLK* were significantly increased in the OE lines and decreased in the *ghap1* mutants (Figure 6A). Similarly, the transcript levels of *GhLFY*, *GhSVP*, *GhMYB*, *GhAGL24*, *GhFT*, *GhFVE*, and *GhFLM* in the OE transgenic lines were markedly higher than those in the WT and the *ghap1* mutant lines. However, the expressions of *GhFCA*, *GhCAL*, and *GhSPL3* were lower in the OE transgenic lines and the *ghap1* mutants than in WT (Figure 6A). The transcript levels of *GhSOC1* were higher in the OE and especially the mutant lines (Figure 6A). These results indicate that *GhAP1-D3* may promote flowering by affecting the expression levels of key flowering genes.

A previous study showed that GhMADS42/GhAP1-D04 (a homolog of GhAP1-D3) interacts with GhSOC1 (Zhang et al., 2016). We established here that the expression levels of *GhSOC1* and *GhFT*, two core genes in the flowering pathway, markedly increased in both OE and mutant lines. We therefore investigated whether GhAP1-D3 might interact with GhSOC1 and GhFT by yeast two-hybrid (Y2H) assay. Indeed, GhAP1-D3 interacted with both GhSOC1 and GhFT in this assay (Figure 6B). To assess the possible formation of a complex between GhAP1-D3, GhSOC1 and GhFT in plant cells, we conducted bimolecular fluorescence complementation (BiFC) assays in

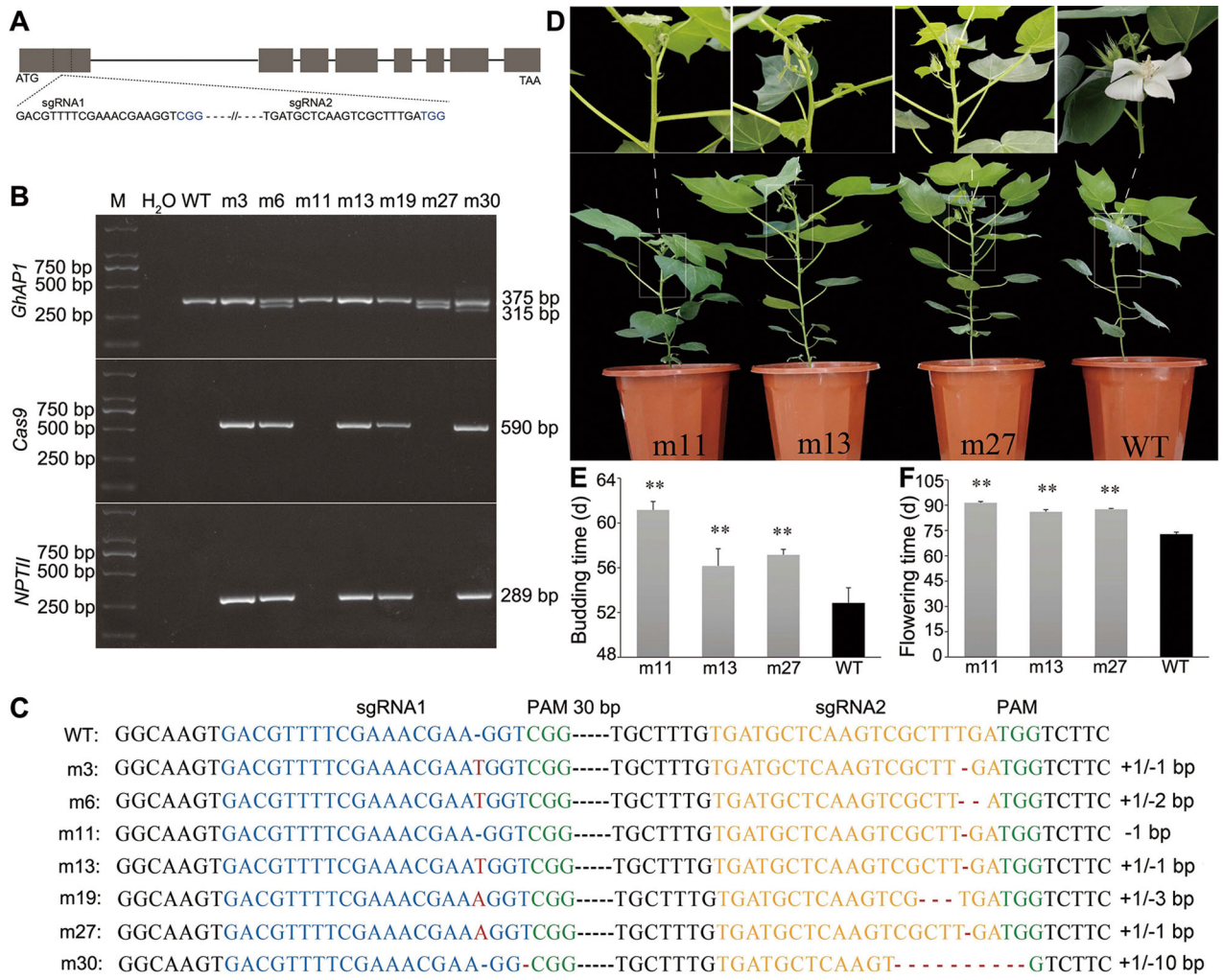

Figure 4. Knock-out of *GhAP1-D3* by clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated nuclease 9 (*Cas9*)-mediated genome editing

(A) Schematic diagram of the *GhAP1-D3* locus and single guide RNA (sgRNA) target sites. **(B)** Polymerase chain reaction (PCR) analysis of the wild-type (WT) (Jin668) and *ghap1* mutants (T_3 generation). **(C)** Editing profiles of mutants at the T_3 generation based on Sanger sequencing results. The blue or red dashes indicate deletions, and red letters show insertions. **(D)** Flowering phenotypes. **(E)** Budding time. **(F)** Flowering time. Data are means ± *SD* ($n \geq 10$). $^{**}P < 0.01$ levels by Student's *t*-test between WT and the mutants.

Nicotiana benthamiana leaves. We determined that GhAP1-D3 interacted with GhSOC1 but not with GhFT *in vivo* (Figure 6C). These results demonstrate that GhAP1-D3 may regulate FTi through a direct interaction with GhSOC1 in upland cotton.

DNA methylation and histone modifications affect *GhAP1-D3* expression

The transcription of many flowering genes has been shown to be modulated by epigenetic modifications (Jeong et al., 2015; He et al., 2020; Yamaguchi, 2022). To explore the potential for regulation of *GhAP1-D3* expression by changes in DNA methylation and histone modifications, we investigated the levels of DNA methylation and histone marks at the *GhAP1-D3* locus in two early-maturing cultivars (ZMS50 and ZMS74) and two late-maturing cultivars (GXM11 and STS458) (Figure 2C). To this end, we divided the *GhAP1-D3* locus into five segments (SetA–SetE) (Figure 7A) and assessed DNA methylation levels by bisulfite sequencing for SetA within the *GhAP1-D3* promoter. We observed DNA methylation in the CG, CHG and CHH contexts in this fragment. The proportions of methylated CG sites were 28.6% and 30.7% in the early-maturing cultivars ZMS50 and ZMS74, respectively, and 33.3% in both late-maturing cultivars, GXM11 and STS458. The rates of CHH methylation at *GhAP1-D3* were 0.8% and 0.6% in ZMS50 and ZMS74, respectively, but 2.2% in GXM11 and 1.8% in STS458. Moreover, the proportions of fully methylated CG and CHH sites at *GhAP1-D3* in ZMS50 and ZMS74 were 3.6% (ZMS50), 3.6% (ZMS74), 5.2% (GXM11), and 4.9% (STS458) (Figure 7B). These results hinted that DNA hypomethylation at the *GhAP1-D3* promoter might be related to the early-maturity trait, with DNA hypermethylation being associated with late maturity. Considering the high expression

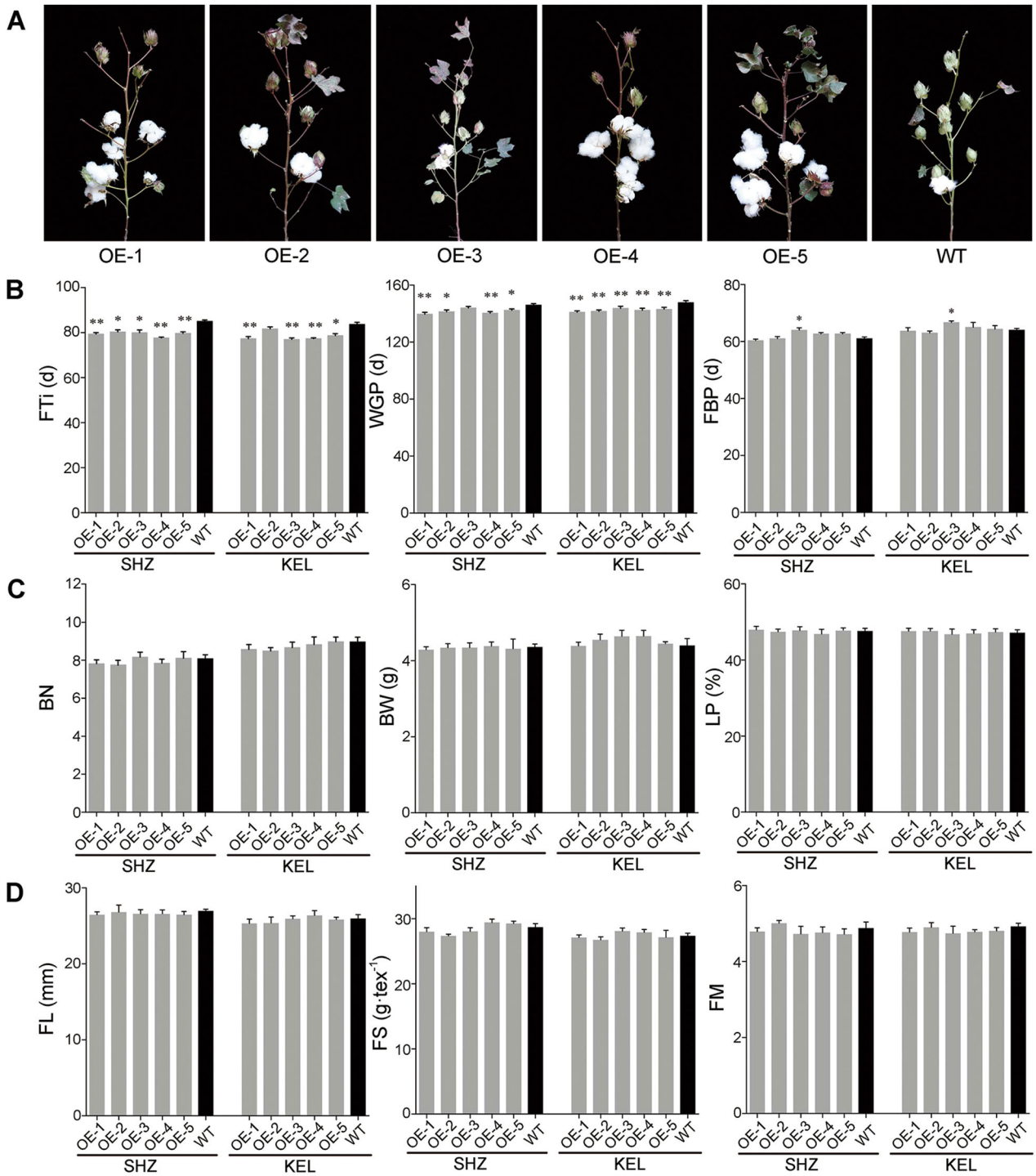

Figure 5. Phenotypic characterization of the *GhAP1-D3* overexpressing lines

(A) Boll-opening phenotypes of the five *GhAP1-D3* overexpression lines and wild-type (WT) plants. **(B)** Mean data for flowering time (FTi), whole growth period (WGP), and flower and boll period (FBP) traits related to early maturity. **(C)** Mean data for the yield component traits boll number per plant (BN), boll weight (BW), and lint percentage (LP). **(D)** Mean data for the three key fiber-quality indicators fiber length (FL), fiber strength (FS), and fiber micronaire (FM). Data are means ± *SD* ($n \geq 10$), $^{*}P < 0.05$; $^{**}P < 0.01$ by Student's *t*-test between overexpression (OE) lines and WT.

levels of *GhAP1-D3* in early-maturity varieties and its low expression levels in late-maturity ones (Figure 2D), we hypothesize that *GhAP1-D3* transcription may be positively modulated by DNA hypomethylation.

Next, we investigated histone modification levels by immunoblotting using specific antibodies against H3K4me2, H3K4me3, H3K9me2, H3K9ac, and H4K5ac, respectively. We established that the levels of H3K4me2, H3K9ac, and

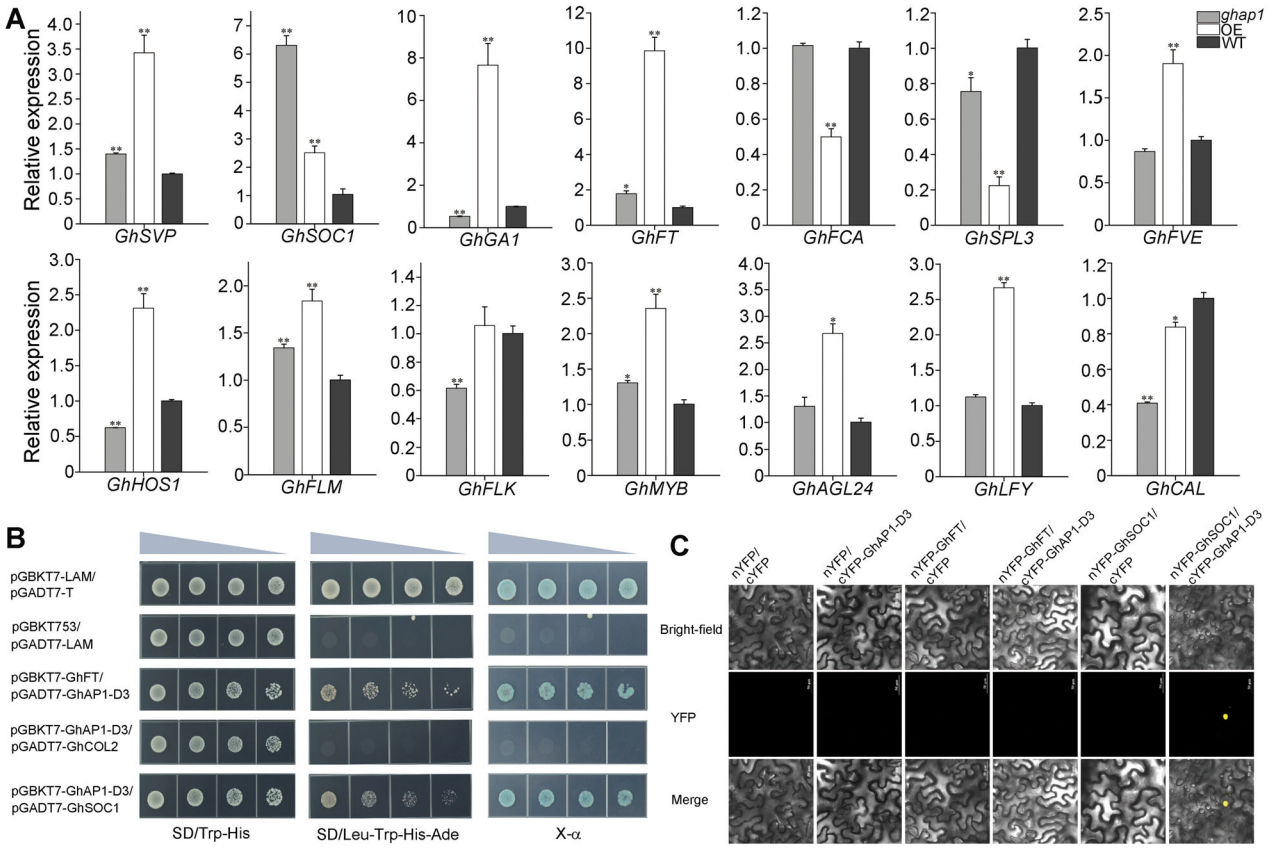

Figure 6. GhAP1-D3 positively regulates cotton flowering time likely through altering the expression of key flowering genes and interacting with GhSOC1

(A) Relative expression levels of 14 key flowering genes in one *GhAP1-D3* overexpression (OE) line, the *ghap1* mutant, and wild-type (WT). Data are means ± *SD* ($n = 3$). $^{*}P < 0.05$; $^{**}P < 0.01$ by Student's *t*-test compared to WT. **(B)** Yeast two-hybrid assay showing the interaction of GhAP1-D3 with GhFT and GhSOC1 in yeast cells. **(C)** Bimolecular fluorescence complementation (BiFC) assay showing the interaction of GhAP1-D3 with GhSOC1.

H4K5ac were high relative to those of to the other two marks in all four cotton cultivars tested (Figure 7C). We then measured the enrichment of H3K4me2, H3K9ac, and H4K5ac marks associated with the *GhAP1-D3* locus by chromatin immunoprecipitation PCR (ChIP-PCR) over the four *GhAP1-D3* regions, SetB–SetE. We observed that H3K9ac was significantly enriched in the SetB–SetE regions in the two early-maturity varieties in comparison to the two late-maturity varieties (Figure 7D), suggesting that the strong expression of *GhAP1-D3* may be positively affected by high levels of H3K9ac in early-maturing cultivars of cotton.

DISCUSSION

Genomic regions associated with cotton early-maturity traits identified by GWAS and linkage mapping

A large proportion of SNP markers have been developed from genome sequencing data in upland cotton, which has allowed the identification of many QTLs for main agronomic traits in recent years (Jia et al., 2016; Ma et al., 2018; Su et al., 2018; Li et al., 2021a). In a complementary approach, many genomic regions associated with early-maturity traits have been explored via linkage mapping and GWAS (Li et al., 2020). For example, 289 QTLs (Table S8) for early maturity-related traits have been detected via linkage mapping based on high-density genetic maps of upland cotton (Jia et al., 2016; Li et al., 2017). Among these, two stable QTLs related to multiple early maturity-related traits were mapped to the chromosomal intervals 95.61–96.91 cM (Jia et al., 2016) and 48–50 cM on chromosome D03 (Li et al., 2017); according to the new reference genome of upland cotton (Hu et al., 2019), these regions lie from 40.24 to 44.20 Mb and from 37.06 to 40.40 Mb, respectively. Moreover, a total of eight, 29, 1,199, 24, 39, and 307 SNPs significantly associated with early maturity-related traits have been identified via GWAS in cotton (Table S9) (Su et al., 2016; Ma et al., 2018; Fu et al., 2019; Shen et al., 2019; Li et al., 2018, 2021a). Based on these QTL mapping and GWAS results, chromosome D03 harbors many QTLs for FTi, FBP, and WGP, suggesting that chromosome D03 harbors many genes modulating early maturity in upland cotton. In agreement with this idea, a major QTL mapping to this chromosome between 38.31 and 41.83 Mb and related to early-maturity traits has been characterized in previous studies (Jia et al., 2016; Su et al., 2016; Ma et al., 2018; Shen et al., 2019; Li et al., 2017, 2021a).

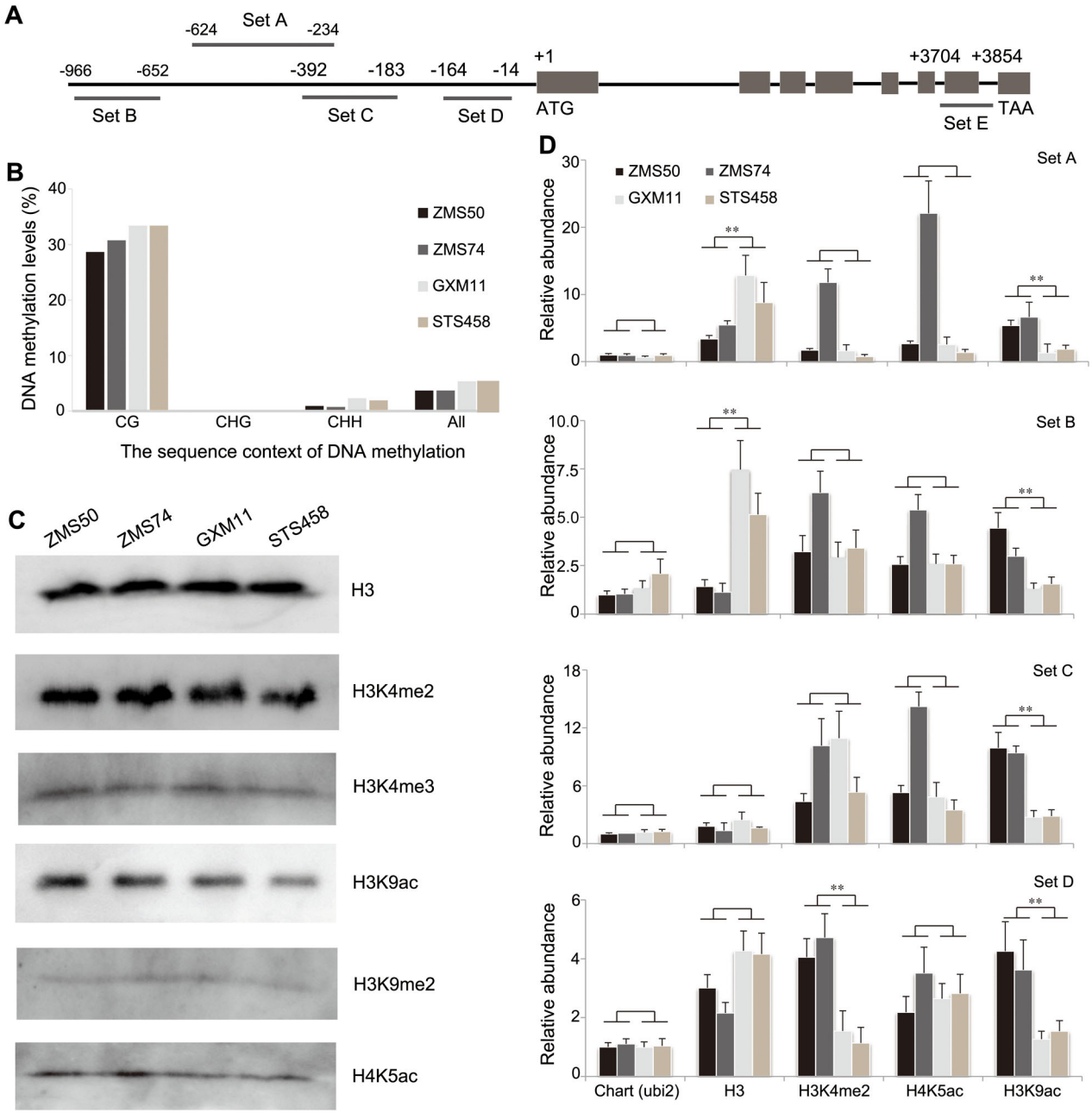

Figure 7. Differential DNA methylation and histone modifications at the *GhAP1-D3* chromatin between early-maturing and late-maturing cotton varieties

(A) Schematic diagram of the *GhAP1-D3* genomic sequence and promoter region showing the five segments (SetA–SetE) used for DNA methylation analysis. **(B)** Mean DNA methylation levels in two early-maturing varieties and two late-maturing varieties over the *GhAP1-D3* promoter. **(C)** Immunoblot analysis of two early-maturing varieties and two late-maturing varieties using specific antibodies against H3K4me2, H3K4me3, H3K9me2, H3K9ac, and H4K5ac. Histone H3 served as loading control. **(D)** Chromatin immunoprecipitation polymerase chain reaction (ChIP-PCR) assays for H3K4me2, H3K9ac, and H4K5ac associated with the *GhAP1-D3* locus in two early-maturing varieties and two late-maturing varieties.

In this study, by taking advantage of the GLM-GWAS, MLM-GWAS, FarmCPU-GWAS and RTM-GWAS methods, we detected a major genomic region containing three SNPs on chromosome D03 that was significantly associated with FTi, FBP, and WGP. A comparison of our GWAS results with those of previous linkage mapping and GWAS data determined that the three significant SNPs were novel and mapped to a region adjacent to that containing the major QTL previously reported on this chromosome.

The 315 upland cotton accessions used here were previously sequenced using the SLAF-seq method, which is a reduced-representation sequencing approach described by Sun et al. (2013). The extent of coverage of SLAF-seq is relatively low, and the number of SNPs developed by SLAF-seq is about 1% of that

obtained from whole-genome resequencing, providing a rationale for our inability to identify the presumptive causal SNP in *GhAP1-D3* from this SNP dataset. Indeed, only the analysis of whole-genome sequences of two early-maturing (ZMS50 and ZMS74) and two late-maturing varieties (GXM11 and STS458) revealed the functional variant represented by the non-synonymous SNP D03_39050639 (Figure S4B).

Identification of candidate genes related to FTi in cotton

Cotton flower development and FTi are closely linked to early maturity (Yu et al., 2017). Previous studies identified several cotton flowering-related genes, such as *GhMADS12*, *GhMADS22*, *GhMADS42* and *GhLFY*, based on their similarity to their *Arabidopsis* homologs (Table S10). The probable functions of these upland cotton genes in the regulation of flowering or flower morphology development were explored by over-expressing these genes in *Arabidopsis* (Wang et al., 2011; Li et al., 2013b; Zhang et al., 2013, 2016). Moreover, some of these candidate genes (Table S10), including *Gh_D03G0922* (Su et al., 2016; Li et al., 2017; Cheng et al., 2021b), *GhEMF2* (*EMBRYONIC FLOWER 2*) (Jia et al., 2016), *Gh_D03G0728* and *Gh_D03G0729* (Ma et al., 2018), *Ghir_D03G011010* (Shen et al., 2019), *Gh_A05G1482* (Fu et al., 2019) and *Ghir_D03G011310* and *Ghir_A05G017290* (Li et al., 2021a), have been identified within the genomic regions associated with early-maturity traits by GWAS and QTL mapping. The functions of these genes were verified by their overexpression in *Arabidopsis*, and/or by VIGS in cotton.

However, whether these candidate genes regulate cotton early maturity has rarely been elucidated in genetically modified cotton plants. For example, transgenic cotton plants expressing an antisense construct for *GhCAL* have been recently shown to exhibit later flower bud differentiation and FTi than WT plants (Cheng et al., 2021a). Overexpression of *GhSOC1* in upland cotton leads to changes in floral organ morphology (Zhang et al., 2016). In this study, we identified *GhAP1-D3* as the key candidate gene mapping to a genomic region identified by GWAS. Importantly, we validated the role of *GhAP1-D3* in accelerating flowering via overexpression lines and genome-edited mutant lines in upland cotton (Figures 3, 4). Field test results showed that transgenic plants overexpressing *GhAP1-D3* flowered significantly earlier than WT plants in terms of both FTi and WGP. Moreover, all overexpression transgenic lines displayed no significant difference in yield and fiber quality component traits (LP, BW, BN, FL and FS, and so on) compared to WT plants in field conditions (Figure 5). These results indicate that *GhAP1-D3* expression can increase early maturity but does not reduce yield and fiber quality. Thus, *GhAP1-D3* may hold promise for cotton breeding to improve early maturity in the northwestern inland cotton-growing region of China in the future.

GhAP1-D3 transcript levels affect the expression of a dozen key flowering genes in upland cotton

The MADS-domain transcription factor AP1 is responsible for floral commitment in *Arabidopsis*, specifying the identity of the floral meristem, sepals, and petals. AP1 also inhibits the expression of the FTi genes *SVP*, *SOC1*, *AGL24*, *FD* and *SPL9* (Mandel et al., 1992; Abe et al., 2005; Liu et al., 2007; Wang et al., 2009; Freytes et al., 2021). *AP1* expression is activated by FT, leading to the upregulation of *AG* and *SEP3*, two MADS-box transcription factor genes and key modulators of flowering. AP1 also interacts with and activates LFY, which reciprocally activates AP1. In addition, AP1 can be activated by SEP3 (SEPALLATA3), and forms a feedback loop with LFY and TERMINAL FLOWER1 (TFL1), thereby controlling floral transition (Liljegren et al., 1999; Gregis et al., 2006; Jung et al., 2016). Moreover, AP1 increases the expression of the gibberellic acid (GA) biosynthetic gene *GA3 oxidase 1* (*GA3ox1*) and the GA catabolism gene *GA2ox1*, thereby adjusting GA levels and cell proliferation (Achard et al., 2009; Kaufmann et al., 2010). Overall, *AP1* is a key flower meristem and floral organ identity gene the protein product of which controls the expression of a series of flowering genes in *Arabidopsis*. *AP1* and *LFY*, *TFL1*, *AGL24* and other genes form a regulatory network to determine FTi in *Arabidopsis* (Wagner et al., 1999; Winter et al., 2015; Goslin et al., 2017). In temperate Gramineae, *VRN1* (*VERNALIZATION1*), an ortholog of *AP1*, regulates FTi mainly through the vernalization pathway (Shrestha et al., 2014; Xu and Chong, 2018). In this study, we showed that *GhAP1-D3* expression affected the transcript levels of some key genes in the GA- and temperature-related flowering regulatory pathways (Figure 6A). Notably, the transcript levels of *GhGA1*, *GhHOS1* and *GhFLK* were significantly increased in the OE lines and decreased in the *ghap1* mutants. This result suggests that *GhGA1*, *GhHOS1*, *GhFLK* and *GhAP1-D3* are co-expressed. Additionally, the transcript abundance of *GhLFY*, *GhAGL24*, and *GhFVE* in the *ghap1* mutants was comparable to that in WT, and their expression levels in the OE transgenic lines markedly increased compared to WT. Therefore, these three genes might act upstream of *GhAP1-D3* in the pathway, and may form a positive feedback regulatory loop with *GhAP1-D3*. Compared to those in WT, the transcript levels of *GhSVP*, *GhMYB*, *GhFT*, *GhFLM*, and *GhSOC1* in both OE and mutant lines were notably elevated, indicating that the proteins encoded by these genes might have inhibitory effects on *GhAP1-D3* expression, while the high levels of *GhAP1-D3* likely enhanced the expression of *GhSVP*, *GhMYB*, *GhFT*, *GhFLM*, and *GhSOC1*.

Previous studies have shown that *GmSOC1* expression increases in both *GmAP1*-overexpression and *gmap1* mutant lines in soybean (*Glycine max*) (Chen et al., 2020); our results are consistent with these findings (Figure 3A). Moreover, *GmSOC1* genes contribute to floral induction in both leaves and the shoot apex through an inter-regulation with *FT* genes (Yue et al., 2021; Kou et al., 2022), suggesting a specific interactive regulatory relationship between AP1s, SOC1s, and FTs in crops. Upland cotton is a tetraploid species with characteristic features of thermophilic plants but lacking photoperiod sensitivity. Exactly how GhAP1 regulates flowering in cotton might differ from how this occurs in *Arabidopsis* and temperate Gramineae. Previous studies have revealed that GhAP1-A04 (also named GhMADS42) can interact with GhSOC1, GhCAL, GhAGL6 and GhSEP4 (Zhang et al., 2016; Cheng et al., 2021a). Our results

showed that GhAP1-D3 interacted with GhSOC1 but not with GhFT (Figure 6B, C), indicating that GhAP1 may be of more general importance in flowering regulation through its interaction with GhSOC1. Considering that the expression levels of several important flowering genes involved in the GA and temperature pathways were clearly altered in the *ghap1* mutants and *GhAP1*-OE lines (Figure 6A), we speculate that GhAP1-D3 might integrate the signals of GAs and environmental temperature to positively regulate flowering through interacting with GhSOC1 in upland cotton.

Association of DNA hypomethylation and high H3K9ac levels with *GhAP1-D3* expression

DNA methylation and histone modifications fulfill critical functions in plant flowering (Gehring, 2019; Yamaguchi, 2022). In *Arabidopsis*, the *METHYLTRANSFERASE 1* mutant, which has reduced DNA methylation, flowers earlier than WT in response to vernalization treatment (Finnegan et al., 1998). Moreover, treatment with DNA methyltransferase inhibitor 5-azacytidine induces flowering in multiple plant species (Kondo et al., 2007; Li et al., 2021b). These results are in agreement with our findings that DNA methylation contents at the *GhAP1-D3* promoter were markedly lower in two early-maturing varieties than in two late-maturing varieties (Figure 7B). These data suggest that DNA hypomethylation is related to early flowering, and may positively modulate the expression of *GhAP1-D3*, thus promoting flowering in cotton. In *Arabidopsis*, histone modifications adjust the transcription of some MADS-box genes like *FLC* and *AGL24*. High levels of H3K27me3 suppress *FLC* expression, resulting in an induction of flowering. H3K36me3 has the role opposite to that of H3K27me3 in influencing *FLC* transcription and floral transition (Yamaguchi, 2022). *LFY* transcription can be activated by high levels of H3K9ac, promoting flowering (Yamaguchi, 2022). Moreover, H3K3me3 changes affect the expression of *AP1* in *Arabidopsis*, also modulating floral transition (Alvarez-Venegas et al., 2003; Guo et al., 2015). In this report, we observed that H3K9ac levels in the promoter region of *GhAP1-D3* were notably higher in two early-maturity varieties relative to two late-maturity varieties (Figure 7D). The possible role of H3K9ac in the regulation of *GhAP1-D3* therefore appears to be similar to that in the modulation of *Arabidopsis AGL24* expression. These results suggest that H3K9ac status may be of general importance in modulating MADS-box gene expression and flowering in plants.

MATERIALS AND METHODS

Plant materials and field trials

The natural population of 290 accessions from upland cotton (*Gossypium hirsutum*) was developed in China; another 25 lines were obtained from outside of China (Table S11). The 315 lines were planted following a completely randomized block design with three replications in three environments: AY-14 (Anyang, Henan, China in 2014), SHZ-14 (Shihezi, Xinjiang, China in 2014) and SHZ-15 (Shihezi, Xinjiang, China in 2015). Each line was planted in single-row plots of 4 m in length, 0.8 m in width, and 0.2 m space between neighbor plants in AY-14, and was grown in double-row plots of 2.0 m in length, 0.76 m in width, and 0.1 m spacing between consecutive plants in SHZ-14 and SHZ-15. The sowing dates were 19 April (AY-14 and SHZ-14), or 25 April (SHZ-15). All management practices, including chemical regulation, insect and weed control, were conducted according to local agronomic practices.

Phenotypic evaluation of three traits related to early maturity and data analysis

The physiological periods of cotton flowering and boll-opening of the natural population were investigated following previously described methods (Jia et al., 2016; Su et al., 2016). The values of FTi, FBP, and WGP were calculated according to the days from sowing to flowering, from flowering to boll-opening, and from sowing to boll-opening, respectively. The BLUP values for all three traits in the 315 lines were calculated using the R package lme4 (Bates et al., 2014). The broad-sense heritability (h^2) of FTi, FBP and WGP were calculated as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_\varepsilon^2/nr)$, where σ_g^2, σ_{ge}^2, σ_ε^2, n, and r represent the genotype variance, genotype × environment variance, error variance, number of environments, and number of replications per environment, respectively (Hanson et al., 1956). The ANOVA and correlation analysis for FTi, FBP, and WGP were carried out using IBM SPSS 26.0 software.

GLM-, MLM-, and FarmCPU-GWAS methods

The genomes of the 315 upland cotton lines were sequenced using the SLAF-seq approach described by Sun et al. (2013). The sequencing and SNP calling methods were based on our previous study (Su et al., 2020). A total of 13,391 high-quality SNPs were defined, including 9,244 SNPLDBs with multiple haplotypes/alleles (Su et al., 2020). To identify the SNPs associated with FTi, FBP, and WGP, three GWAS methods were used by the GLM, MLM (PCs + K) or FarmCPU algorithm in a single-locus linear model, based on the 13,391 identified SNPs and phenotypic values from each of three environments and their BLUP values for the three target traits. When $-\log_{10}(P) > 5.74$, SNPs were considered to be significantly associated with the trait.

RTM-GWAS procedures

To effectively detect and confirm the loci significantly associated with FTi, FBP, or WGP, a restricted two-stage multi-locus and multi-allele GWAS (RTM-GWAS) procedure (He et al., 2017) was followed using the phenotypic traits and 9,244 organized SNPLDB markers. In the first stage, SNPLDBs were preselected using a single-locus linear model ($y = wa + xb + e$) at the 5% significance level. In the second phase, abundant significant genomic regions were identified by a stepwise regression under a multi-locus model ($y = wa + x_1b_1 + x_2b_2 + \cdots + x_kb_k + e$) at the 5% significance level. In the two models, y, x, w, a, and e represent a phenotypic value vector (y), the design matrices of markers under testing (x), the genetic similarity coefficient matrix established using the 9,244 SNPLDBs (w), the population structure effect estimated (a), and residual effects

(*e*). *b* is an allele-effect vector estimated from the single-locus model. *k* is the number of SNPs and b_1,..., b_k are the allele-effect vectors to be calculated in the multi-locus model.

Haplotype analysis and potential candidate gene mining

Based on the LD blocks among major SNPs, the haplotypes were determined and their phenotypic effects were calculated and plotted as box plots in OriginPro 2018C software. ANOVA was conducted to inspect significant differences among phenotypic values of different haplotypes using IBM SPSS 26.0 software. The major SNPs were used to mine possible candidate genes based on the upland cotton reference genome v 2.1 (Hu et al., 2019). The genome regions containing candidate genes were determined by LD based on pairwise R^2 values between the SNPs in the local Manhattan plot region on chromosome D03.

Transcriptome data analysis

A previously published RNA-seq dataset was used, representing six developmental stages (cotyledon and one to five true leaf stages) of the ZMS50 (early-maturing) and GXM11 (late-maturing) varieties (Cheng et al., 2021a). The mean frames per kilobase of exon per million mapped reads (FPKM) values of the gene expression data from the three replicates were calculated as the expression levels of the genes in the mapping interval. The ratio of mean FPKM values between ZMS50 and GXM11 were calculated for all genes for each of the six growth periods. The tissue-specific expression of DEGs was determined using RNA-seq datasets from 22 upland cotton (TM-1) samples (roots, stems, leaves, bracts, torus, sepals, petals, anthers, pistils, filaments, and ovules at 0, 1, 3, 5, 10, 15, 20, and 25 d post-anthesis (DPA)), and fibers at 10, 15, 20, and 25 DPA (http://cotton.zju.edu.cn).

Quantitative RT-PCR analysis

For two early-maturing cultivars (ZMS50 and ZMS74) and two late-maturing cultivars (GXM11 and STS458), shoot apical meristems and the youngest leaves were harvested in the third true leaf period, and total RNA was extracted using a Plant RNA Purification Kit (Tiangen, Beijing, China). First-strand cDNA was reverse transcribed from 1 μg total RNA using a FastKing gDNA Dispelling RT SuperMix Kit (Tiangen, Beijing, China). qPCR experiments were conducted on a Roche lightCycler® 96 Real-Time PCR System (Swiss) with gene-specific primers (Table S12).

Cloning, vector construction and transformation

The full-length coding sequence of *GhAP1-D3* was amplified from the cDNA of early-maturing cotton variety ZMS50 by RT-PCR with the primer pair GhAP1F and GhAP1R (Table S13), and subcloned into the pGM-T vector (Tiangen, Beijing, China). The *GhAP1-D3* coding sequence was cloned into the pBI121 vector with the cauliflower mosaic virus (CaMV) 35S promoter, resulting in the 35S:*GhAP1* construct which was then introduced into *Agrobacterium tumefaciens* strain LBA4404. Construct-positive *Agrobacterium* colonies were then used to transform the cotton receptor line Jin668, following the *Agrobacterium*-mediated hypocotyl transformation method (Jin et al., 2006). Transgenic cotton plants were selected on Murashige and Skoog medium containing 50 mg/L kanamycin and 400 mg/L cefotaxime. The OE lines were genotyped by PCR using a specific primer pair (PBI35S-F/OEGhAP1-D3-R) (Table S14). Relative *GhAP1-D3* expression levels were estimated in all OE lines by RT-qPCR.

CRISPR/Cas9-mediated editing of *GhAP1-D3*

To generate *ghap1* mutants, 20-bp target sequences specific to *GhAP1-D3* and with NGG as protospacer adjacent motif at their 3′ ends were selected using the CRISPR sgRNA design online tool (http://crispr.mit.edu). Two target sequences (5′-GACGTTTTCGAAACGAAGGT-3′ and 5′-TGATGCTCAAGT CGCTTTGA-3′) were selected for cloning into the CRISPR/Cas9 vector. The sgRNA1-tRNA-sgRNA2 fragment was ligated into the pRGEB32-GhU6.9-NPTII expression vector (Wang et al., 2018). The constructs were introduced into *Agrobacterium* strain LBA4404 and transformed into the wild-type cotton line (Jin668) according a published method (Jin et al., 2006). The target site of each sgRNA in the genome was amplified by PCR from the *ghap1* mutants with a specific primer pair (Table S15) with 2× Pro Taq Master Mix (Accurate Biotechnology, AG). About 300 ng PCR product was analyzed by electrophoresis on 2% (w/v) agarose gel for the presence of deletions and Sanger sequencing. Putative *ghap1* mutant plants were also analyzed by PCR using *NPTII*-specific and *Cas9*-specific primers (Table S15) from the pRGEB32-GhU6.9-NPTII vector. The *ghap1* mutants were identified by Sanger sequencing and PCR genotyping at the T_0 or T_3 generation, respectively.

Measurement of comprehensive traits in the transgenic lines

Plants from five transgenic *GhAP1-D3* overexpression lines and WT were grown in double-row plots in SHZ and KEL, Xinjiang, China, during 2021 in a completely randomized block design with three replications in each environment. The phenotypes of FTi, FBP, WGP, FBN and plant height (PH) were scored as follows. In early October in each environment, FBN was investigated and a total of 20 bolls from the central part of each plant were harvested by hand and weighed to calculate BW. Then, a cotton engine was used to peel the entire fiber samples. Fiber samples were weighed to obtain fiber weight (FW) with an electronic balance. The LP values of each line were calculated as LP (%) = FW/BW× 100%. The five fiber-quality evaluation indicators FL, FS, FU, FM, and FE were measured using about 15 g fiber samples with an HVI-MF 100 instrument (User Technologies, Inc., USTER, Switzerland). A Student's one-tailed *t*-test was performed to determine statistical significance in SPSS 22.0 software.

Protein–protein interaction assays

For Y2H experiments, the coding sequence of *GhAP1-D3* was amplified and cloned into the pGBKT7 vector. The

resulting construct was introduced into yeast strain Y2HGold for self-activation activity assays and toxicity tests. The coding sequences of the target genes were individually cloned into the pGADT7 vector. The appropriate pairs of pGBKT7-gene and pGADT7-gene vectors were introduced into yeast strain Y2HGold, with pGBKT7 and pGADT7 used as negative controls. Positive colonies were selected on synthetic defined (SD) –Trp –Leu double dropout medium and protein–protein interactions were tested on SD –Ade –His –Leu –Trp quadruple dropout medium.

For BiFC assays, the coding sequences of *GhAP1-D3*, *GhSOC1* and *GhFT* were amplified from TM-1 cotton cDNA using the specific primers listed in Table S16. The *GhAP1-D3* amplicon was recombined into vector pXY104 harboring the sequence of the C-terminal region of yellow fluorescent protein (*cYFP*), while the *GhSOC1* and *GhFT* amplicons were individually recombined into vector pXY106 harboring the N-terminal sequence of YFP (*nYFP*) (Yu et al., 2008). The resulting constructs were then introduced into *Agrobacterium* strain GV3101 and co-infiltrated into well-expanded *Nicotiana benthamiana* leaves. After 48 h, YFP fluorescence in leaf cells was detected by confocal microscopy (Zeiss LSM510).

DNA methylation assay

The top buds and young leaves of 30-d-old cotton plants were harvested, and genomic DNA was extracted with a MiniBEST Plant Genomic DNA Extraction Kit (TaKaRa, Beijing, China). The DNA was digested, and whole-genome methylation levels of the DNA were measured by high-performance liquid chromatography (Osabe et al., 2014). A 391-bp DNA fragment of the *GhAP1-D3* promoter was selected to assay its methylation level by bisulfite sequencing (Hu et al., 2012). Unmethylated cytosines were converted to uracil after sodium bisulfite treatment using an EZ DNA Methylation-Gold kit (Zymo Research, CA, USA). Bisulfite-treated genomic DNA was amplified with bisulfite-specific primers (Table S17) and sequenced to identify methylated cytosines.

Immunoblotting and ChIP-qPCR analysis of histone modifications

Total proteins were isolated from top buds and young leaves of 30-d-old cotton plants with a Total Protein Miniprep Kit (TIANDZ, Beijing, China). Immunoblotting was conducted following He et al. (2020). Histone H3 was used as a loading control. ChIP-PCR results were obtained based on a previously described method (Zhang et al., 2018). In brief, chromatin was isolated and digested into 200–500 bp fragments using micrococcal nuclease. The chromatin concentration was quantified, and the same amount of chromatin was immunoprecipitated by antibodies respectively against histone H3, H3K4me2, H3K4me3, H3K9me2, H3K9ac, or H4K5ac modifications at 4°C for about 12 h. A rabbit serum was used as the negative control. Immunoprecipitated DNA was quantified by qPCR with specific primers (Table S18). The relative enrichment of each histone mark in the immunoprecipitated DNA was normalized to input levels or to histone H3.

Data availability statement

The sequencing data from the SLAF-seq analysis for the 315 sequenced upland cotton lines are available in the Sequence Read Archive under accession number PRJNA314284 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA314284/).

ACKNOWLEDGMENTS

We especially thank Professor Shuxun Yu and his team (at the Institute of Cotton Research of CAAS), who provided us with good guidance for the experiment. This work was funded by the National Natural Science Foundation of China (31971986 and 32260478), the Education Technology Innovation Project of Gansu Province (2022QB-076), the Gansu Province Science and Technology Program (20JR10RA520), the Biological Breeding program of Gansu Academy of Agricultural Sciences (2022GAAS04) and the Tianshan talent plan of Xinjiang Autonomous Region (2021).

CONFLICTS OF INTEREST

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

J.S., F.H., and X.M. designed the experiments; C.W., J.L., J. W., X.X., and Q.M. performed the experiments; J.S., P.C., and X.X. analyzed the data; J.S. and C.W. wrote the manuscript. D.Y. and F.H. revised the manuscript. C.W. and J.L. contributed equally to this study, and all authors read and approved the final manuscript.

Edited by: Feng Tian, China Agricultural University, China

Received Jul. 5, 2022; **Accepted** Nov. 11, 2022; **Published** Nov. 18, 2022

REFERENCES

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science **309:** 1052–1056.

Achard, P., Gusti, A., Cheminant, S., Alioua, M., Dhondt, S., Coppens, F., Beemster, G.T.S., and Genschik, P. (2009). Gibberellin signaling controls cell proliferation rate in *Arabidopsis*. Curr. Biol. **19:** 1188–1193.

Alvarez-Venegas, R., Pien, S., Sadder, M., Witmer, X., Grossniklaus, U., and Avramova, Z. (2003). ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. Curr. Biol. **13:** 627–637.

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. **13:** 627–639.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). Linear mixed-effects models using eigen and S4. R package Version. Madison, Wisconsin, United States pp. 1–7.

Bluemel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops—What did we learn from *Arabidopsis*? Curr. Opin. Biotechnol. **32:** 121–129.

Chen, G.W., Yu, Y., and Li, H. (2014). On the strategic shift of high-yielding cultivation theory in cotton in Xinjiang-from exploring heat energy potential to light energy. Xinjiang Farm Res. Sci. Technol. **37:** 3–6. (in Chinese).

Chen, L.Y., Nan, H.Y., Kong, L.P., Yue, L., Yang, H., Zhao, Q.S., Fang, C., Li, H.Y., Cheng, Q., Lu, S.J., Kong, F.J., Liu, B.H., and Dong, L.D. (2020). Soybean AP1 homologs control flowering time and plant height. J. Integr. Plant Biol. **62:** 1868–1879.

Cheng, S.S., Chen, P.Y., Su, Z.Z., Ma, L., Hao, P.B., Zhang, J.J., Ma, Q., Liu, G.Y., Liu, J., Wang, H.T., Wei, H.L., and Yu, S.X. (2021a). High-resolution temporal dynamic transcriptome landscape reveals a GhCAL-mediated flowering regulatory pathway in cotton (*Gossypium hirsutum* L.). Plant Biotechnol. J. **19:** 153–166.

Cheng, X.Q., Wang, H.T., Wei, H.L., Gu, L.J., Hao, P.B., Sun, H.R., Wu, A.M., Cheng, S.S., and Yu, S.X. (2021b). The MADS transcription factor GhAP1.7 coordinates the flowering regulatory pathway in upland cotton (*Gossypium hirsutum* L.). Gene **769:** 145235.

Fan, S.L., Yu, S.X., Song, M.Z., and Yuan, R.H. (2006). Construction of molecular linkage map and QTL mapping for earliness in short-season cotton. Cotton Sci. **18:** 135–139. (in Chinese).

Feng, L., Dai, J.L., Tian, L.W., Zhang, H.J., Li, W.J., and Dong, H.Z. (2017). Review of the technology for high-yielding and efficient cotton cultivation in the northwest inland cotton-growing region of China. Field Crops Res. **208:** 18–26.

Finnegan, E.J., Genger, R.K., Kovac, K., Peacock, W.J., and Dennis, E. S. (1998). DNA methylation and the promotion of flowering by vernalization. Proc. Natl. Acad. Sci. U.S.A. **95:** 5824–5829.

Freytes, S.N., Canelo, M., and Cerdán, P.D. (2021). Regulation of flowering time: When and where? Curr. Opin. Plant Biol. **63:** 102049.

Fu, Y.Z., Dong, C.G., Wang, J., Wang, Y.Y., and Li, C.Q. (2019). Genome-wide association study reveals the genetic control underlying node of the first fruiting branch and its height in upland cotton (*Gossypium hirsutum* L.). Euphytica **215:** 35.

Gehring, M. (2019). Epigenetic dynamics during flowering plant reproduction: Evidence for reprogramming? New Phytol. **224:** 91–96.

Goslin, K., Zheng, B.B., Serrano-Mislata, A., Rae, L., Ryan, P.T., Kwaśniewska, K., Thomson, B., Ó'Maoiléidigh, D.S., Madueño, F., Wellmer, F., and Graciet, E. (2017). Transcription factor interplay between LEAFY and APETALA1/CAULIFLOWER during floral initiation. Plant Physiol. **174:** 1097–1109.

Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2006). AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in *Arabidopsis*. Plant Cell **18:** 1373–1382.

Guo, S.Y., Sun, B., Looi, L.S., Xu, Y.F., Gan, E.S., Huang, J., and Ito, T. (2015). Co-ordination of flower development through epigenetic regulation in two model species: Rice and *Arabidopsis*. Plant Cell Physiol. **56:** 830–842.

Hanson, C.H., Robinson, H.F., and Comstock, R.E. (1956). Biometrical studies of yield in segregating populations of Korean Lespedeza. Agron. J. **48:** 267–282.

He, J.B., Meng, S., Zhao, T.J., Xing, G.N., Yang, S.P., Li, Y., Guan, R.Z., Lu, J.J., Wang, Y.F., Xia, Q.J., Yang, B., and Gai, J.Y. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. Theor. Appl. Genet. **130:** 2327–2343.

He, S.B., Hao, Y.F., Zhang, Q., Zhang, P.H., Ji, F.F., Cheng, H., Lv, D., Sun, Y.F., Hao, F.S., and Miao, C. (2020). Histone deacetylase inhibitor SAHA improves high salinity tolerance associated with hyperacetylation-enhancing expression of ion homeostasis-related genes in cotton. Int. J. Mol. Sci. **21:** 7105.

Hu, Y., Chen, J.D., Fang, L., Zhang, Z.Y., Ma, W., Niu, Y.C., Ju, L.Z., Deng, J.Q., Zhao, T., Lian, J.M., Baruch, K., Fang, D., Liu, X., Ruan, Y.L., Rahman, M.U., Han, J.L., Wang, K., Wang, Q., Wu, H.T., Mei, G. F., Zang, Y.H., Han, Z.G., Xu, C.Y., Shen, W.J., Yang, D.F., Si, Z.F., Dai, F., Zou, L.F., Huang, F., Bai, Y.L., Zhang, Y.G., Brodt, A., Ben-Hamo, H., Zhu, X.F., Zhou, B.L., Guan, X.Y., Zhu, S.J., Chen, X.Y., and Zhang, T.Z. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat. Genet. **51:** 739–748.

Hu, Y., Zhang, L., He, S.B., Huang, M., Tan, J.J., Zhao, L., Yan, S.H., Li, H., Zhou, K., Liang, Y.N., and Li, L.J. (2012). Cold stress selectively unsilences tandem repeats in heterochromatin associated with accumulation of H3K9ac. Plant Cell Environ. **35:** 2130–2142.

Jeong, H.J., Yang, J., Yi, J., and An, G. (2015). Controlling flowering time by histone methylation and acetylation in *Arabidopsis* and rice. J. Plant Biol. **58:** 203–210.

Jia, X.Y., Pang, C.Y., Wei, H.L., Wang, H.T., Ma, Q.F., Yang, J.L., Cheng, S.S., Su, J.J., Fan, S.L., Song, M.Z., Wusiman, N., and Yu, S.X. (2016). High-density linkage map construction and QTL analysis for earliness-related traits in *Gossypium hirsutum* L. BMC Genomics **17:** 909.

Jin, S., Zhang, X., Nie, Y., Guo, X., Liang, S., and Zhu, H. (2006). Identification of a novel elite genotype for *in vitro* culture and genetic transformation of cotton. Biol. Plant. **50:** 519–524.

Jung, J.H., Lee, H.J., Ryu, J.Y., and Park, C.M. (2016). SPL3/4/5 integrate developmental aging and photoperiodic signals into the FT-FD module in *Arabidopsis* flowering. Mol. Plant Pathol. **9:** 1647–1659.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). *Arabidopsis* MET1 cytosine methyltransferase mutants. Genetics **163:** 1109–1122.

Kaufmann, K., Wellmer, F., Muiño, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueño, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., and Riechmann, J.L. (2010). Orchestration of floral initiation by APETALA1. Science **328:** 85–89.

Khan, M.R.G., Ai, X.Y., and Zhang, J.Z. (2014). Genetic regulation of flowering time in annual and perennial plants. Wiley Interdiscip. Rev. RNA **5:** 347–359.

Kondo, H., Miura, T., Wada, K.C., and Takeno, K. (2007). Induction of flowering by 5-azacytidine in some plant species: Relationship between the stability of photoperiodically induced flowering and flower-inducing effect of DNA demethylation. Physiol. Plant. **131:** 462–469.

Kou, K., Yang, H., Li, H.Y., Fang, C., Chen, L.Y., Yue, L., Nan, H.Y., Kong, L.P., Li, X.M., Wang, F., Wang, J.H., Du, H.P., Yang, Z.Y., Bi, Y.D., Lai, Y.C., Dong, L.D., Cheng, Q., Su, T., Wang, L.S., Li, S.C., Hou, Z.H., Lu, S.J., Zhang, Y.H., Che, Z.J., Yu, D.Y., Zhao, X.H., Liu, B.H., and Kong, F.J. (2022). A functionally divergent SOC1 homolog improves soybean yield and latitudinal adaptation. Curr. Biol. **32:** 1728–1742.

Lacape, J.M., Gérard, G., Tuong-Vi, C., Christopher, V., Danny, L., Shiming, L., Jacobs, J., Becker, D., Barroso, P.A.V., Assunção, J. H., Palaï, O., Georges, S., Jean, J., and Giband, M. (2013). Mapping QTLs for traits related to phenology, morphology and yield components in an inter-specific *Gossypium hirsutum*×*G. barbadense* cotton RIL population. Field Crops Res. **144:** 256–267.

Li, C.Q., Fu, Y.Z., Liu, Q., Du, L., and Trotsenko, V. (2020). A review of genetic mechanisms of early maturity in cotton (*Gossypium hirsutum* L.). Euphytica **216:** 4140–4148.

Li, C.Q., Wang, X.Y., Dong, N., Zhao, H.H., Xia, Z., Wang, R., Converse, R.L., and Wang, Q.L. (2013a). QTL analysis for early-maturing traits in cotton using two upland cotton (*Gossypium hirsutum* L.) crosses. Breed. Sci. **63:** 154–163.

Li, C.Q., Wang, Y.Y., Ai, N.J., Li, Y., and Song, J.F. (2018). A genome-wide association study of early-maturation traits in upland cotton based on the CottonSNP80K array. J. Integr. Plant Biol. **60:** 970–985.

Li, J., Fan, S.L., Song, M.Z., Pang, C.Y., Wei, H.L., Li, W., Ma, J.H., Wei, J.H., Jing, J.G., and Yu, S.X. (2013b). Cloning and characterization of a FLO/LFY ortholog in *Gossypium hirsutum* L. Plant Cell Rep. **32:** 1675–1686.

Li, L.B., Zhang, C., Huang, J.Q., Liu, Q.B., Wei, H.L., Wang, H.T., Liu, G. Y., Gu, L.J., and Yu, S.X. (2021a). Genomic analyses reveal the genetic basis of early maturity and identification of loci and candidate genes in upland cotton (*Gossypium hirsutum* L.). Plant Biotechnol. J. **19:** 109–123.

Li, L.B., Zhao, S.Q., Su, J.J., Fan, S.L., Pang, C.Y., Wei, H.L., Wang, H. T., Gu, L.J., Zhang, C., Liu, G.Y., Yu, D.W., Liu, Q.B., Zhang, X.L., and Yu, S.X. (2017). High-density genetic linkage map construction by F2 populations and QTL analysis of early-maturity traits in upland cotton (*Gossypium hirsutum* L.). PLoS ONE **12:** e0182918.

Li, X.J., He, Q., and Yuan, Y.J. (2000). Character analysis of amount of heat in serious reduction years of cotton climatic output in Xinjiang. J. Xinjiang Agr. Univ. **23:** 20–26. (in Chinese).

Li, Z.Q., Tang, M.Q., Luo, D.J., Kashif, M.H., Cao, S., Zhang, W.X., Hu, Y.L., Huang, Z., Yue, J., Li, R., and Chen, P. (2021b). Integrated methylome and transcriptome analyses reveal the molecular mechanism by which DNA methylation regulates kenaf flowering. Front. Plant Sci. **12:** 709030.

Liljegren, S.J., Gustafson-Brown, C., Pinyopich, A., Ditta, G.S., and Yanofsky, M.F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. Plant Cell **11:** 1007–1018.

Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., and Yu, H. (2007). Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. Development **134:** 1901–1910.

Mandel, M.A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. Nature **360:** 273–277.

Ma, Z.Y., He, S.P., Wang, X.F., Sun, J.L., Zhang, Y., Zhang, G.Y., Wu, L. Q., Li, Z.K., Liu, Z.H., Sun, G.F., Yan, Y.Y., Jia, Y.H., Yang, J., Pan, Z., Gu, Q.S., Li, X.Y., Sun, Z.W., Dai, P.H., Liu, Z.W., Gong, W.F., Wu, J.H., Wang, M., Liu, H.W., Feng, K.Y., Ke, H.F., Wang, J.D., Lan, H. Y., Wang, G.N., Peng, J., Wang, N., Wang, L.R., Pang, B.Y., Peng, Z., Li, R.Q., Tian, S.L., and Du, X.M. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat. Genet. **50:** 803–813.

Mao, S.C. (2013). Cotton Farming in China. 66-91 Shanghai: Shanghai Scientific and Technical Press. pp. 884–898. (in Chinese).

Osabe, K., Clement, J.D., Bedon, F., Pettolino, F.A., Ziolkowski, L., Llewellyn, D.J., Finnegan, E.J., and Wilson, I.W. (2014). Genetic and DNA methylation changes in cotton (*Gossypium*) genotypes and tissues. PLoS ONE **9:** e86049.

Pelayo, M.A., Yamaguchi, N., and Ito, T. (2021). One factor, many systems: The floral homeotic protein AGAMOUS and its epigenetic regulatory mechanisms. Curr. Opin. Plant Biol. **61:** 102009.

Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.L., and Yanofsky, M.F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. Plant J. **26:** 385–394.

Ren, L.H., Guo, W.Z., and Zhang, T.Z. (2002). Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line. Acta Bot. Sin. **44:** 815–820. (in Chinese).

Shen, C., Wang, N., Huang, C., Wang, M.J., Zhang, X.L., and Lin, Z.X. (2019). Population genomics reveals a fine-scale recombination landscape for genetic improvement of cotton. Plant J. **99:** 494–505.

Shrestha, R., Gomez-Ariza, J., Brambilla, V., and Fornara, F. (2014). Molecular control of seasonal flowering in rice, *Arabidopsis* and temperate cereals. Ann. Bot. **114:** 1445–1458.

Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol. Cell **6:** 791–802.

Su, J.J., Li, L.B., Zhang, C., Wang, C.X., Gu, L.J., Wang, H.T., Wei, H.L., Liu, Q.B., Huang, L., and Yu, S.X. (2018). Genome-wide association study identified genetic variations and candidate genes for plant architecture component traits in Chinese upland cotton. Theor. Appl. Genet. **131:** 1299–1314.

Su, J.J., Pang, C.Y., Wei, H.L., Li, L.B., Liang, B., Wang, C.X., Song, M. Z., Wang, H.T., Zhao, S.Q., Jia, X.Y., Mao, G.Z., Huang, L., Geng, D. D., Wang, C.S., Fan, S.L., and Yu, S.X. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. BMC Genomics **17:** 687.

Su, J.J., Wang, C.X., Yang, D.L., Shi, C.H., Zhang, A., Ma, Q., Liu, J.J., Zhang, X.L., Huang, L., and Ma, X.F. (2020). Decryption of favourable haplotypes and potential candidate genes for five fibre quality properties using a relatively novel genome-wide association study procedure in upland cotton. Ind. Crops Prod. **158:** 113004.

Sun, X.W., Liu, D.Y., Zhang, X.F., Li, W.B., Liu, H., Hong, W.G., Jiang, C. B., Guan, N., Ma, C.X., Zeng, H.P., Xu, C.H., Song, J., Huang, L., Wang, C.M., Shi, J.J., Wang, R., Zheng, X.H., Lu, C.Y., Wang, X.W., and Zheng, H.K. (2013). SLAF-seq, an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS ONE **8:** e58700.

Wagner, D., Sablowski, R.W., and Meyerowitz, E.M. (1999). Transcriptional activation of APETALAI by LEAFY. Science **285:** 582–584.

Wang, J.W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. Cell **138:** 738–749.

Wang, L.N., Fan, S.L., Song, M.Z., Pang, C.Y., and Yu, S.X. (2011). Cloning of *GhMADS12* gene in cotton and construction of plant expression vector. Mol. Plant Breed. **9:** 34–40. (in Chinese).

Wang, P.C., Zhang, J., Sun, L., Ma, Y.Z., Xu, J., Liang, S.J., Deng, J.W., Tan, J.F., Zhang, Q.H., Tu, L.L., Daniell, H., Jin, S.X., and Zhang, X. L. (2018). High efficient multi-sites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. Plant Biotechnol. J. **16:** 137–150.

Winter, C.M., Yamaguchi, N., Wu, M.F., and Wagner, D. (2015). Transcriptional programs regulated by both LEAFY and APETALA1 at the time of flower formation. Physiol. Plant. **155:** 55–73.

Xu, S.J., and Chong, K. (2018). Remembering winter through vernalisation. Nat. Plants **4:** 997–1009.

Yamaguchi, N. (2022). The epigenetic mechanisms regulating floral hub genes and their potential for manipulation. J. Exp. Bot. **73:** 1277–1287.

Yue, L., Li, X.M., Fang, C., Chen, L.Y., Yang, H., Yang, J., Chen, Z. H., Nan, H.Y., Chen, L.N., Zhang, Y.H., Li, H.Y., Hou, X.L., Dong, Z.C., Weller, J.L., Abe, J., Liu, B.H., and Kong, F.J. (2021). FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. J. Integr. Plant Biol. **63:** 1004–1020.

Yu, S.X., Wang, H.T., Wei, H.L., and Su, J.J. (2017). Research progress and application of early maturity in upland cotton. Cotton Sci. **29:** 1–10. (in Chinese).

Yu, X., Li, L., Li, L., Guo, M., Chory, J., and Yin, Y. (2008). Modulation of brassinosteroid-regulated gene expression by Jumonji domain-containing proteins ELF6 and REF6 in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. **105:** 7618–7623.

Zhang, A., Wang, C.X., Su, J.J., Zhang, X.L., Shi, C.H., Liu, J.J., Peng, Y.L., and Ma, X.F. (2020). Identification of MADS-box family and analysis of tissue specific expression in *Gossypium hirsutum* L. Cotton Sci. **32:** 404–417. (in Chinese).

Zhang, H., Yue, M.X., Zheng, X.K., Gautam, M., He, S.B., and Li, L.J. (2018). The role of promoter-associated histone acetylation of *Haem Oxygenase-1*

(*HO-1*) and *Giberellic Acid-Stimulated Like-1* (*GSL-1*) genes in heat-induced lateral root primordium inhibition in maize. Front. Plant Sci. **9:** 1520.

Zhang, W.X., Fan, S.L., Pang, C.Y., Wei, H.L., Ma, J.H., Song, M.Z., and Yu, S.Y. (2013). Molecular cloning and function analysis of two SQUAMOSA-like MADS-box genes from *Gossypium hirsutum* L. J. Integr. Plant Biol. **55:** 597–607.

Zhang, X.H., Wei, J.H., Fan, S.L., Song, M.Z., and Pang, C.Y. (2016). Functional characterization of *GhSOC1* and *GhMADS42* homologs from upland cotton (*Gossypium hirsutum* L.). Plant Sci. **242:** 178–186.

Zhang, X.L., Gao, J.S., Song, G.L., Liu, F., Li, S.H., Liu, K.F., Chu, Z.Y., Shao, Y.Y., and Wang, K.B. (2009). Additive and epistatic effects QTL analysis on upland cotton CRI-G6. Mol. Plant Breed. **7:** 312–320. (in Chinese).

Zhao, T., Zhan, Z., and Jiang, D. (2019). Histone modifications and their regulatory roles in plant development and environmental memory. J. Genet. Genomics **46:** 467–476.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.13409/suppinfo

Figure S1. Manhattan plot for the best linear unbiased prediction (BLUP) value for the genome-wide association studies (GWAS) of flowering time (FTi) in upland cotton

Figure S2. Manhattan plot for the best linear unbiased prediction (BLUP) value for the genome-wide association studies (GWAS) of flower and boll period (FBP) in upland cotton

Figure S3. Manhattan plot for the best linear unbiased prediction (BLUP) value for the genome-wide association studies (GWAS) of whole growth period (WGP) in upland cotton

Figure S4. Schematic model of the associated single nucleotide polymorphisms (SNPs) and the annotated genes in the major genomic regions associated with early maturity traits

Figure S5. Variation in the coding or protein sequence for GhAP1-D3 between early- and late-maturing cultivars

Figure S6. Complementary DNA and full-length open reading frame (ORF) of *GhAP1-D3*

Figure S7. Southern blotting analysis of 13 *GhAP1-D3* overexpression (OE) lines. Lines 5, 6, 8, 10, and 11 represent five stable single-copy T_3 lines (*GhAP1-D3* OE-1, *GhAP1-D3* OE-2, *GhAP1-D3* OE-3, *GhAP1-D3* OE-4, and *GhAP1-D3* OE-5)

Figure S8. Isolation of *ghap1* mutants by clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated nuclease 9 (Cas9)

Figure S9. Phenotypes in *GhAP1-D3* overexpression lines. Plant height (PH), boll number per plant (BN), fiber uniformity (FU) and fiber elongation (FE) are shown. Data are means $\pm SD$ ($n \geq 10$); * and **, $P < 0.05$ and $P < 0.01$ levels by Student's *t*-test between wild-type (WT) and over-expressing lines, respectively

Table S1. Variance for the three early maturity-related traits flowering time (FTi), flower and boll period (FBP), and whole growth period (WGP) in upland cotton

Table S2. Correlation between values for flowering time (FTi), flower and boll period (FBP), and whole growth period (WGP)

Table S3. Significantly associated single nucleotide polymorphisms (SNPs) with flowering time (FTi) in upland cotton

Table S4. Significantly associated single nucleotide polymorphisms (SNPs) with flower and boll period (FBP) in upland cotton

Table S5. Significantly associated single nucleotide polymorphisms (SNPs) with the whole growth period (WGP) in upland cotton

Table S6. Significantly associated single nucleotide polymorphism linkage disequilibrium blocks (SNPLDBs) with the three early maturity-related traits flowering time (FTi), flower and boll period (FBP), and whole growth period (WGP) in multiple-environment model

Table S7. List of 76 genes annotated within the 2.10-Mb genomic region of chromosome D03

Table S8. Genomic regions detected by linkage mapping based on high-density single nucleotide polymorphisms (SNP) maps for the early maturity-related traits in upland cotton

Table S9. Significantly associated single nucleotide polymorphisms (SNPs) with the early maturity-related traits flowering time (FTi), flower and boll period (FBP) and whole growth period (WGP) in upland cotton

Table S10. Flowering-related genes identified in previous studies in upland cotton

Table S11. Information on the 315 upland cotton accessions used in this study

Table S12. Primers used for reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Table S13. Primers used to clone *GhAP1-D3*

Table S14. Primers used for the confirmation of overexpression plants

Table S15. Primers used to genotype *ghap1* mutant plants and *NPTII*- and *Cas9*-specific genes

Table S16. Primers used for the construction of bimolecular fluorescence complementation (BiFC) constructs

Table S17. Primers used for bisulfite sequencing

Table S18. Primers used for the chromatin immunoprecipitation - quantitative polymerase chain reaction (ChIP-qPCR) assay