A glutathione S-transferase *GhTT19* determines flower petal pigmentation via regulating anthocyanin accumulation in cotton

Qichao Chai[1,†] (ID), Xiuli Wang[1,†], Mingwei Gao[1], Xuecheng Zhao[2], Ying Chen[1], Chao Zhang[1], Hui Jiang[1], Jiabao Wang[1], Yongcui Wang[1], Meina Zheng[3], Ahmedov Miraziz Baltaevich[1], Jian Zhao[2,*] (ID) and Junsheng Zhao[1,3,*]

[1]*Key Laboratory of Cotton Breeding and Cultivation in Huang-Huai-Hai Plain, Institute of Industrial Crops, Shandong Academy of Agricultural Sciences, Jinan, China*
[2]*Key Laboratory of Tea Science of Ministry of Education, College of Horticulture, Hunan Agricultural University, Changsha, China*
[3]*College of Life Sciences, Shandong Normal University, Jinan, China*

Summary

Anthocyanin accumulations in the flowers can improve seed production of hybrid lines, and produce higher commodity value in cotton fibre. However, the genetic mechanism underlying the anthocyanin pigmentation in cotton petals is poorly understood. Here, we showed that the red petal phenotype was introgressed from *Gossypium bickii* through recombination with the segment containing the R_3^{bic} region in the A07 chromosome of *Gossypium hirsutum* variety LR compared with the near-isogenic line of LW with white flower petals. The cyanidin-3-O-glucoside (Cy3G) was the major anthocyanin in red petals of cotton. A *GhTT19* encoding a TT19-like GST was mapped to the R_3^{bic} site associated with red petals via map-based cloning, but *GhTT19* homologue gene from the D genome was not expressed in *G. hirsutum*. Intriguingly, allelic variations in the promoters between *GhTT19^{LW}* and *GhTT19^{LR}*, rather than genic regions, were found as genetic causal of petal colour variations. GhTT19-GFP was found localized in both the endoplasmic reticulum and tonoplast for facilitating anthocyanin transport. An additional MYB binding element found only in the promoter of *GhTT19^{LR}*, but not in that of *GhTT19^{LW}*, enhanced its transactivation by the MYB activator GhPAP1. The transgenic analysis confirmed the function of *GhTT19* in regulating the red flower phenotype in cotton. The essential light signalling component GhHY5 bonded to and activated the promoter of *GhPAP1*, and the GhHY5-GhPAP1 module together regulated *GhTT19* expression to mediate the light-activation of petal anthocyanin pigmentation in cotton. This study provides new insights into the molecular mechanisms for anthocyanin accumulation and may lay a foundation for faster genetic improvement of cotton.

Introduction

Cotton has obvious heterosis, its coloured flowers not only provide convenient selection markers for effectively distinguishing the hybrids but also as a visual signal to attract pollinators. There has been a report on the purple spot increasing the visit of honeybees and improving the seed production in cotton hybrid lines (Abid *et al.*, 2022). In addition, colourful cotton fibres are desirable agronomic traits that are targeted in cotton breeding (Hinchliffe *et al.*, 2016; Ke *et al.*, 2022; Yan *et al.*, 2018). The pigments of most coloured petals and fibres of cotton are derived from anthocyanins, a major form of flavonoids (Tan *et al.*, 2013). Thus, revealing the mechanism of anthocyanin biosynthesis and regulation is very important for cotton breeding.

It is reported that anthocyanins are synthesized on the cytosolic side of the endoplasmic reticulum (ER) by a multi-enzyme complex (Grotewold, 2006; Winkel-Shirley, 2001). After synthesis, they are transported into the vacuole for storage: the vesicle trafficking-, membrane transporter- and glutathione S-transferase (GST)-mediated models are three distinct, but interrelated mechanisms for vacuolar sequestration of anthocyanins (Zhao, 2015). GST-mediated transports and storages of anthocyanins, which is

suggested to be coupled with vesicle trafficking- and ATP binding cassette (ABC) or multidrug and toxic extrusion (MATE) transporters for vacuolar sequestration of anthocyanins in *Arabidopsis* (Francisco *et al.*, 2013; Gomez *et al.*, 2009, 2011). Anthocyanin-producing plants were found to have anthocyanin-specific GST proteins, such as maize BZ2, *Arabidopsis* TT19 and grapevine VvGST1 (Marrs *et al.*, 1995; Perez-Diaz *et al.*, 2016; Sun *et al.*, 2012). The TT19-like-GSTs act as a carrier to transport anthocyanins from the endoplasmic reticulum to the vacuole. For instance, BZ2 is capable of conjugating cyanidin-3-O-glucoside (Cy3G) with glutathione (Marrs *et al.*, 1995), *Arabidopsis* TT19 and *Grapevine* VvGST1 bond to anthocyanins (Conn *et al.*, 2008; Sun *et al.*, 2012). Recent reverse genetic and molecular studies indicated that cotton *GST* was an essential gene for anthocyanin accumulation in the red leaf of upland cotton (Li *et al.*, 2020; Shao *et al.*, 2021). Also, *Gar07G08900*, a GST homologue was responsible for the red-spot-petals in *Gossypium arboreum* (Zhang *et al.*, 2022). However, no forward genetic and genomic study was reported for more details about anthocyanin pigmentation in cotton petals. Light is the most important environmental factor that regulates anthocyanin synthesis and accumulation, by modifying the core MYB-bHLH-WD40 (MBW) transcriptional complex (Xu *et al.*, 2014). The

anthocyanin-specific MYB transcription factors such as *Arabidopsis PAP1* and its orthologs from other plants, such as tomato and apple, play a key role in the regulation of anthocyanin biosynthesis and accumulation (Ballester *et al.*, 2010; Espley *et al.*, 2007; Xu *et al.*, 2014). Moreover, HY5, the light signalling component, could bind to the promoter of *PAP1* and enhance its activity, mediating the light-regulated anthocyanin production (An *et al.*, 2017; Gangappa and Botto, 2016; Shin *et al.*, 2013), indicating that anthocyanin accumulation in plants is widely regulated at transcriptional levels (Shao *et al.*, 2022; Xu *et al.*, 2015).

Five anthocyanin-related loci *R1*, *R2*, *R3*, *Re* and *Lc1* have been reported in cotton. *R1* (Li *et al.*, 2019) and *Re* (Wang *et al.*, 2022) are related to red foliar organs and stem, R2 to red petal spots (Abid *et al.*, 2022) and *R3* to red petal and *Lc1* to brown fibre (Hinchliffe *et al.*, 2016). Most loci, except for *R3*, have been identified by map-based cloning, and encode R2R3-MYB transcription factors. In this work, using the genetic population between LR with the red flower and LW with the white flower, we mapped the *R3* locus. By comparison of genome sequences, we confirmed it originated from the *Gossypium bickii* R_3^{bic} for red flower and encoded a GST protein. Biochemical, genetic, molecular and physiological studies confirmed the function of *GhTT19* in anthocyanin transport and the allelic variations in promoters of *GhTT19*^{LR} and *GhTT19*^{LW} caused their differential transactivation by anthocyanin activator GhPAP1. Furthermore, the GhHY5-GhPAP1 module regulation of *GhTT19* expression could explain the redness variations of cotton flower petals under different light conditions in the field.

Results

Cy3G is the major anthocyanin in red petals of LR

Different flower colours existed in the three types of cotton (Figure S1). The red petals with a large basal spot in *G. hirsutum* LR variety were originally transferred from *G. bickii* (Liang *et al.*, 1997). LR is the near-isogenic line of *G. hirsutum* L. acc. LW with a pale white flower (Figure 1a,b). Liquid chromatography–mass spectrometry (LC–MS)/MS analysis of the water-soluble anthocyanins extracted from LR's red flowers showed that one major peak was abundantly detected in LR extracts, but not detectable in extracts of LW flowers (Figure 1c). The fragment ions at m/z 287 and m/z 449 of the mother allowed the identification of Cy3G as the corresponding anthocyanin (Figure 1d). Quantification of anthocyanins confirmed the significantly higher Cy3G in petals of LR than LW (Figure 1e).

Map-based cloning of the red flower R_3^{bic} locus

To determine the genetic locus of anthocyanin pigments in cotton petals, we constructed a mapping population by crossing LR with LW plants. All F_1 plants of 76 offsprings exhibited red flower phenotypes of LR, but their F_2 generations were segregated into 252 red flower individuals of LR and 100 white flower individuals of LW, nearly fitting a 3 : 1 segregation ratio ($\chi^2 = 2.202 < \chi^2_{0.05,1} = 3.84$). This result suggested that the phenotype of LR's red flower was controlled by a single nuclear dominant gene. Linkage analysis with mapping population including 352 F_2 plants revealed that the R_3^{bic} region, a chromosomal segment of a 33.2-kb flanked between markers of A07-0594 and A07-0592 on Chr.A07, was responsible for the red petal phenotype (Figure 2a). To define which of four candidate genes determine the red petal phenotype, we cloned and analysed all four genes and analysed their expression patterns (Table S3,

Figure 2b,c), and finally excluded the other three as unrelated ones to flower petal colour. Quantification of the transcript levels of candidate genes in petals with qRT-PCR showed that the expression of *ORF1* could not be detected in LW, but a considerable expressing level was detected in LR (Figure 2b,c,d). Therefore, ORF1 was considered the candidate gene for R_3^{bic}.

Based on a high-quality reference genome sequence of *G. hirsutum* L. acc. TM-1 (http://cotton.zju.edu.cn/), this region contains four open reading frames (ORFs) that all have corresponding transcripts in LR (Tables S2 and S3). To further track the origin of the R_3^{bic} site, we compared the genome sequences of *G. hirsutum* (both LR and LW) and *G. bickii* (Sheng *et al.*, 2022) and found that this R_3^{bic} site located in a large chromosomal segment of about 7 Mb in *G. hirsutum*-LR (8–15 Mb segment of A07) that was introgressed from *G. bickii* (9.94–19.03 Mb segment on Chr07) by recombinant replacement of A07 regions (Figure S4). This region contains 274 genes (Table S4, Figure S5), including four genes in the R_3^{bic} site. Among these genes, 83 genes showed more than twice the significant differential expression between petals of LR and LW: 10 up-regulated genes in LR and 73 genes in LW. *ORF1*(*GH_A07G0814*) has the highest FPKM in LR and the highest fold-change between the petals of LR and LW (Figure S5a). Except for *ORF1*, no other genes were reported to be related to anthocyanin metabolism (Figure S5b).

The ORF1 (*GH_A07G0814*) and their homologues in the D subgenome were amplified from LR and LW (Figure S2a). It encodes a putative GST protein of 214 amino acids highly homology to *Arabidopsis TT19* and thus was named *GhTT19*. There were eight single nucleotide polymorphism (SNPs) in *GhTT19-At* coding regions in LR and LW varieties (here named as *GhTT19*^{LR}-*At* and *GhTT19*^{LW}-*At*, respectively), resulting in three amino acid changes (Figure S2a). There were no differences between the sequences of *GhTT19*^{LR}-*Dt*. However, *GhTT19*^{LR}-*At* and *GhTT19*^{bickii} have identical ORF sequences (Figure S2a) as well as promoter sequences (Figure S3). Two SNPs in *GhTT19* coding regions could distinguish *GhTT19*^{LR}-*At*, *GhTT19*^{LW}-*At* and *GhTT19*-*Dt* using pyrosequencing (Figure 2e). A-A combination represents *GhTT19*^{LR}-*At*, A-G represents *GhTT19*^{LW}-*At* and G-G represents *GhTT19*-*Dt*. We cloned the sequence containing two loci from genomic DNA. Thirty positive clones from LR and LW were sequenced respectively. Only A-A and G-G combination exist in LR, A-G and G-G combination exist in LW (Figure 2f). Pyrosequencing of LR cDNA showed that although *GhTT19*^{LR}-*Dt* exists in LR/LW, it is not expressed (Figure 2g).

We developed a pair of SNP detection primers, based on point mutation at the 288th position of the *GhTT19* coding region that could specifically detect the 288th-A (LR) but not the 288th-G (LW) genotype (Figure S2b). An enlarged F_2 population of 1191 individuals, including 868 red-flower individuals and 323 white-flower individuals, was used for screening with the pair of SNP primers. All 868 red-flower individuals were successfully amplified with the expected bands, while the 323 white-flower individuals failed (Figure S2c). Thus, *GhTT19* was the candidate gene attributed to the red flower of LR.

### *GhTT19* was required for anthocyanin accumulation

Among 12 *GST* genes in cotton plants belonging to the *GSTF* subgroup (Phi), *GhTT19* from LR and LW shared the highest homology to *AtTT19* in protein sequence, exhibiting 59% identity (Figure S6a). The *GhTT19*^{LR} over-expressing lines of *tt19* mutant, which had green hypocotyls similar to the wild-type, showed purple hypocotyls (Figure S6b). When *GhTT19* was suppressed in
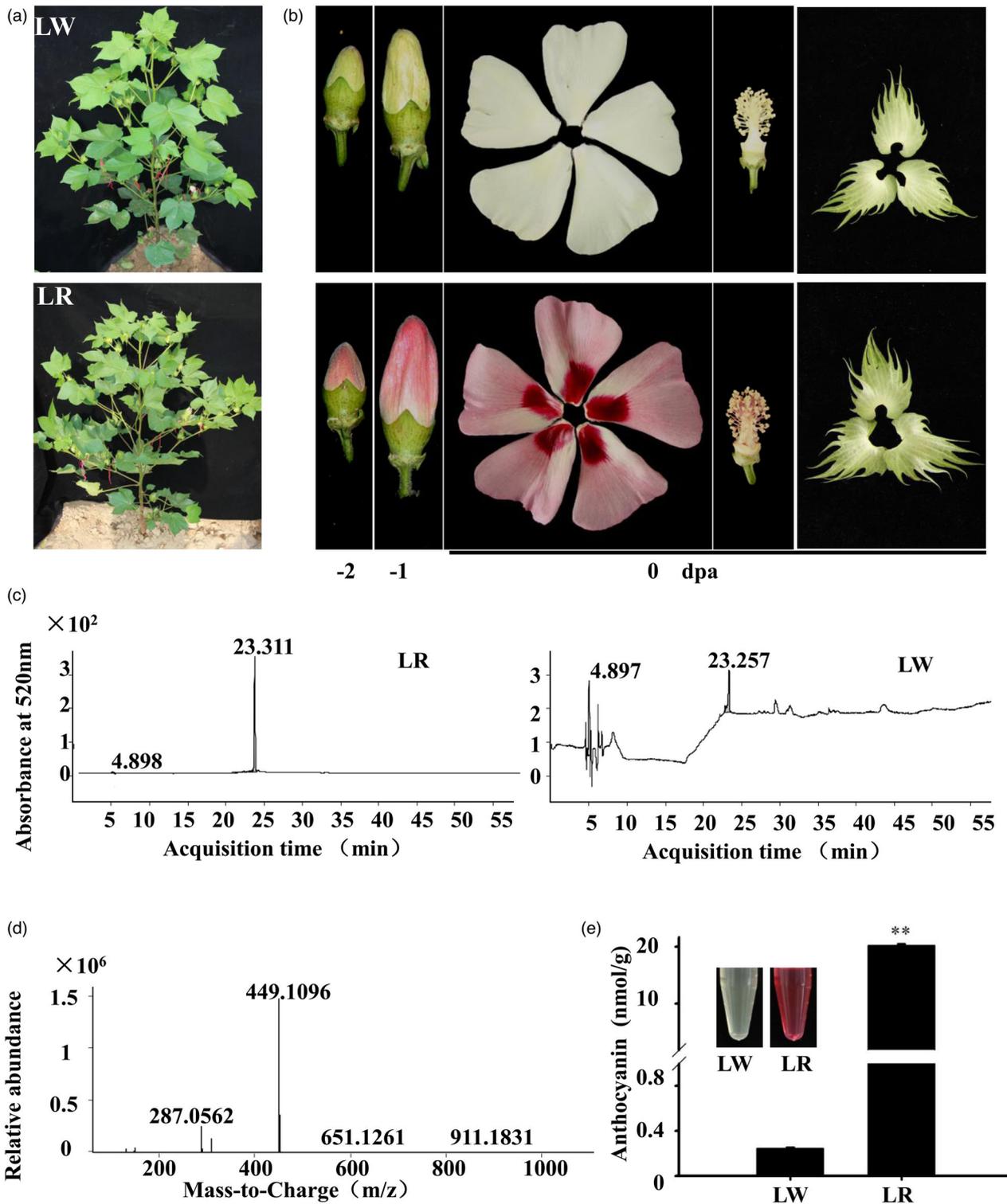
Figure 1 Phenotypic differences of LR and LW. (a) Three-month-old LR and LW plants are grown in the field. (b) The flower phenotype at 2 days, 1 day before or at 0 day post-anthesis (DPA). (c) HPLC chromatograms of anthocyanin in LR and LW lines. (d) Identification of anthocyanin molecule in red petals of LR flowers. An anthocyanin peak, corresponding to cyanidin-3-O-glucoside (Cy3G) was identified from the red petals of LR using ESI–LC–MS/MS. (e) Quantification of total anthocyanins in LR and LW flower petals. Total anthocyanins were extracted with an acidic methanol solution and measured at 530 and 657 nm. Data are presented as means ± SE from at least three biological replicates. $**P < 0.01$, $*P < 0.05$ at Student's t-test.

LR using the pCLCrV-based virus-induced gene silencing (VIGS) assay, the down-regulation of *GhTT19* in LR changed the petal colour from red into light pink (Figure 3a–e). We introduced

GhTT19LR into white cotton cultivar W0 via *Agrobacterium*-mediated transformation. Overexpression of *GhTT19LR* in the T$_0$ plant causes the new leaf, stem and flower to turn red (Figure 3f–
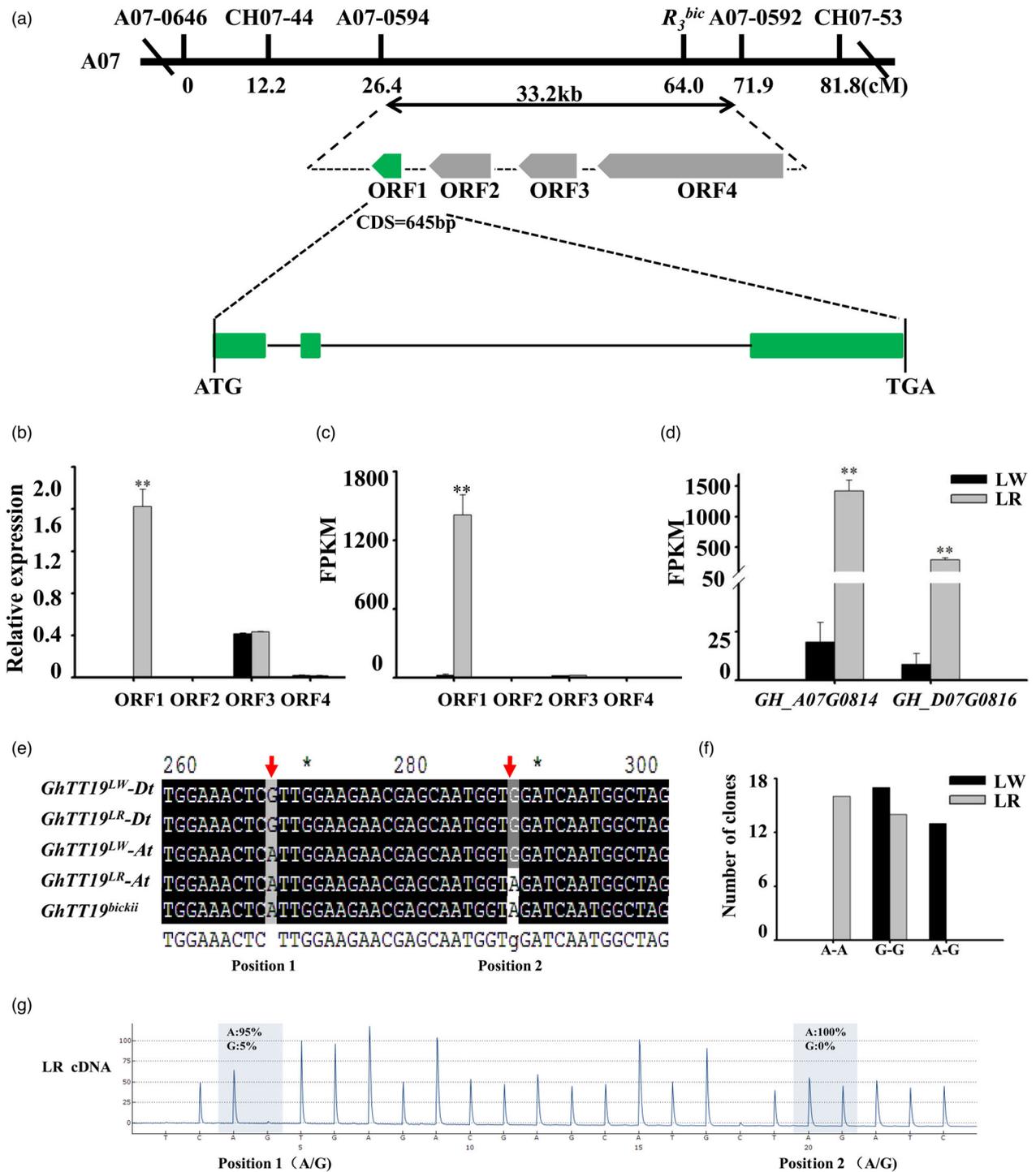
Figure 2 Map-based cloning of the red petal gene R_3^{bic}. (a) Physical map of the R_3^{bic} locus. Four genes (ORF1-4) in the region are indicated by green and grey thick arrows. (b) qRT-PCR analysis of 4 candidate genes in the petals of LR and LW. ORF1 was strongly expressed in the red petals of LR but was almost not expressed in the white petals of LW. (c) FPKM expression values of candidate genes in the petals of LR and LW. (d) FPKM expression values of ORF1 (*GH_A07G0814*) and homologous gene of D subgenome (*GH_D07G0816*) in the petals of LR and LW. (e) Alignment of the ORF sequences of *GhTT19* from LW and LR. A base combination of two positions could distinguish *GhTT19^LR-At*, *GhTT19^LW-At* and *GhTT19-Dt* using pyrosequencing. (f) The number of base combinations of two positions in 30 positive clones. The genomic DNA of LR and LW were used for cloning. (g) Pyrophosphate sequencing using cDNA of LR petal. Data are presented as means ± SE from at least three biological replicates. **$P < 0.01$, *$P < 0.05$ at Student's t-test.

i, Figure S9). The *GhTT19*-dependent anthocyanin accumulation was further supported by qRT-PCR analyses of *GhTT19* expression in the F$_2$ segregated population developed from a cross between

LW as the female parent and LR as the male parent. The F$_2$ population consisted of 39 white flower individuals (F$_2$, 1–39) and 97 red flower individuals (F$_2$, 40–136), of them, all the red flower

individuals had highly expressed *GhTT19*, but all the white flower individuals had no detectable expressions of *GhTT19* (Figure S10). In sum, *GhTT19* was the important gene for anthocyanin accumulation in cotton.

GhTT19 bond to Cy3G and localized on ER and vacuole

To explain why this GST protein is critically required for anthocyanin accumulation in cotton, we examined whether GhTT19 can function as an anthocyanin-binding protein carrier in the anthocyanin transport process. The bacterially expressed and purified GhTT19[LR] recombinant proteins were examined by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) and immunoblotting (Figure 4a,b). The affinities of Cy3G to respective GhTT19[LR] and GhTT19[LW] were measured by Biacore 8000, showing that both proteins could bind to Cy3G in vitro (Figure 4c). The values of affinities of the proteins GhTT19[LR] and GhTT19[LW] to Cy3G were 7.01×10^{-5} and 3.45×10^{-5}, respectively, based on the steady-state affinity model. Both proteins could bind to Cy3G in vitro and their affinities to Cy3G were comparable. The subcellular localization assay showed that GhTT19-GFP was mostly localized on the membranes of ER and the vacuole, in line with the *Arabidopsis* TT19 (Sun *et al.*, 2012), implying the GhTT19 functioned in anthocyanin transport. The results showed that both GhTT19[LR] and GhTT19[LW] could bind to Cy3G and localized on both ER and vacuole.

An MYB binding element enhanced *GhTT19^LR* transcriptional activity via interaction with GhPAP1

To detect the cause of the differential expression levels of *GhTT19* in varieties of LR and LW, we detected the tissue-specific expression variations of *GhTT19* in both plants. The transcriptional level of *GhTT19* in petals and filaments was 240- and 12-fold higher in LR than in LW, respectively. *GhTT19* is expressed in roots, stems and leaves at comparatively lower levels both in LR and LW (Figure 5a). We further checked the promoter activities of *GhTT19^LR* and *GhTT19^LW*. The 1694-bp and 1652-bp promoter regions upstream of the *GhTT19*-At gene were isolated using LR and LW genomic DNAs, respectively (Figure S11). Two promoters were subcloned into constructs to drive a *GUS* expression in tobacco leaves. The *GUS* expression was significantly higher in pro*GhTT19*[LR]-transformant than in pro*GhTT19*[LW]-transformant (Figure 5b,c). Hence, it is clear that pro*GhTT19^LR* had stronger activity than pro*GhTT19^LW*. To probe the cause of the high promoter activity of LR, we compared the *cis*-acting elements in promoter sequences of both LR and LW through NewPLACE (www.dna.affrc.go.jp/PLACE/?action=newplace). The GhTT19 promoters of both LR and LW had MYB binding element CAACTG. However, two AC elements (ACC[A/T]A[A/C][T/C] and ACC[A/T][A/C/T][A/C/T]), and one EEC element (GAAATTC) of MYB existed in the promoter of GhTT19[LR] but absented in that of the GhTT19[LW] (Figure S11). The *MYB* transcription factor *GhPAP1* is a homologous gene of *Arabidopsis* PAP1 (Figure S12), which has been reported as the major controlling gene of the red plant phenotype in cotton (Li *et al.*, 2019). It was shown that *GhPAP1* had the same gene expression patterns as *GhTT19*. *GhPAP1* was almost not expressed in roots, stems and leaves, but had higher expressing levels in the petals and filament both in LR and LW (Figure 5d). Y1H assays showed that both promoters of *GhTT19^LR* and *GhTT19^LW* could bind to GhPAP1, but the former had a stronger combination than the latter (Figure 5e). Besides, GhHY5 could bind to pro*GhTT19^LR*, but not to pro*GhTT19^LW* (Figure 5e). An electrophoretic mobility shift (EMSA) assay was

employed to evaluate the effects of sequence mutations on *GhTT19* promoter activity. The results showed that *GhPAP1* could bind to the elements of CAACTG and EEC (GAAATTC), but not to the two AC elements (Figure 5f). We analysed promoter sequences of *GhTT19* of different cotton species with diverse genetic backgrounds that had different flower colours. All of them have the CAACTG element, but only the cotton species with red flowers have the EEC (GAAATTC) element (Figure S13). Transactivation assays via pro*GhTT19^LR* and pro*GhTT19^LW* driven LUC reporter system further showed that GhPAP1 could activate pro*GhTT19^LR* but not pro*GhTT19^LW* (Figure 5g). The overall conclusion was that GhPAP1 could strongly interact with the promoter *GhTT19^LR* through GAAATTC and enhanced its transcript activity and anthocyanidin accumulation.

## *GhHY5-GhPAP1* modulated anthocyanin accumulation relying on light intensity

Light intensity critically regulates flower coloration in cotton plants. When LR plants grew in the field under a strong light intensity of 80 000 *Lux*, the flowers displayed more anthocyanin accumulation. However, when LR grew in an artificial climate growth chamber where light intensity was 3200 *Lux*, the flowers were much less pigmented (Figure 6a). Correspondingly, the transcriptional levels of *GhPAP1* and *GhTT19* of LR cultivated under the stronger light intensity were drastically higher than those under weaker light intensity (Figure 6c). We then examined the cotton homologue genes of *HY5*, which is a positive regulator of light signalling and mediates the light-regulated anthocyanin production (Lee *et al.*, 2007; Xu, 2020). Among 163 bZIP transcription factors identified from the *G. hirsutum* genome, *GH_D08G2693* was the closest homologue of *Arabidopsis AtHY5* and was named *GhHY5* (Figure S14a). Gene expression profiling revealed that *GhHY5* was the only *bZIP* gene in cotton, appearing with similar expression patterns to *GhTT19* and *GhPAP1*, displaying the highest expression levels in the petal and filaments (Figure S14b). qRT-PCR analyses also validated the results (Figure 6b). Accordingly, *GhTT19*, *GhPAP1* and *GhHY5* of LR exhibited higher or lower expression levels, respectively, under stronger and weaker light intensity (Figure 6c). Y1H assays showed that GhHY5 protein bonds to the promoter of *GhPAP1* (Figure 6d). Transactivation assays with the *proGhPAP1::LUC* reporter system showed that GhHY5 could activate *proGhPAP1* (Figure 6e). EMSA assay, using bacterially expressed and purified GhHY5 recombinant protein, showed that GhHY5 could bind to the 332 bp sequence at the promoter of *GhPAP1* (Figure 6f). Thus, *GhHY5* was closely associated with *GhPAP1-GhTT19*-mediated anthocyanin accumulation in petals under different light conditions.

Discussion

Colourful flowers, fruits, vegetables and seeds packed with natural plant pigments are able to attractive to pollinators, enhance stress tolerance and represent a visible selection marker for crop breeders (Su *et al.*, 2019; Zhang *et al.*, 2019; Zhu *et al.*, 2017). We created four hybrid cotton varieties with red flowers, such as the national registered cotton variety Lu 05H9, Lu HB biaoza-1, these varieties were planted and welcomed by cotton farmers. Anthocyanins, including pelargonidin, cyanidin, delphinidin and their derivatives from glycosylation, acylation and methylation, are widely found in higher plants (Grotewold, 2006). We showed here that the red pigments in the petals of LR flowers

were mainly Cy3G, one of the most common anthocyanins present in the flower of cotton, which was absent in the LW variety possessing white flowers (Figure 1). There are four *R2R3*

MYB transcription factors, *GhTT2* (Hinchliffe *et al.*, 2016; Yan *et al.*, 2018), *GhPAP1* (Li *et al.*, 2019), *GhBM* (Abid *et al.*, 2022) and *Re* gene (Wang *et al.*, 2022), which control flavonoid
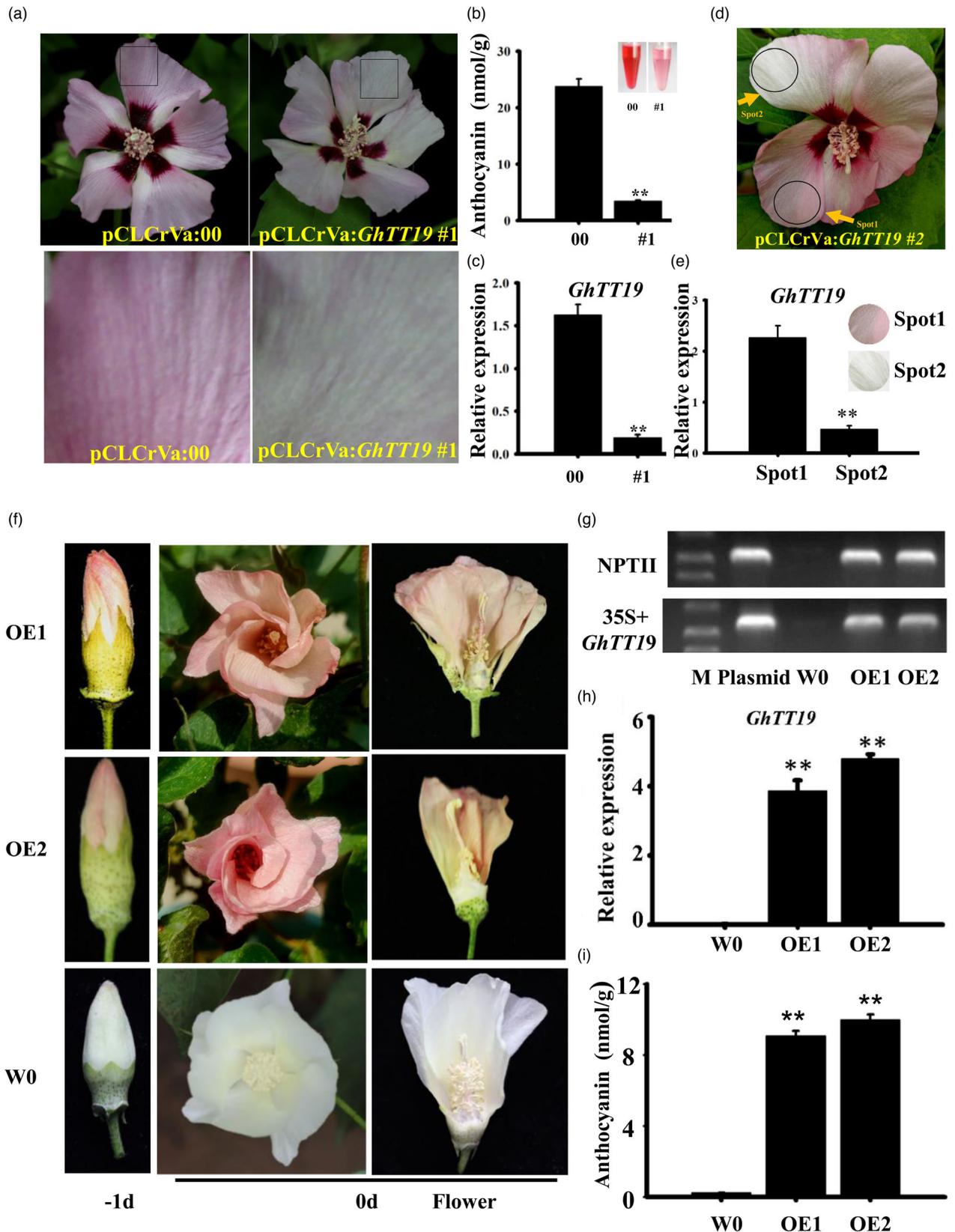
Figure 3 *GhTT19* gene is responsible for anthocyanin accumulation in cotton petals. (a) The phenotypes of LR petal after *GhTT19* were silenced by VIGS. The image below is an enlargement of the contents of the box. (b) Quantification of total anthocyanins in pCLCrVa:00 (CK) and *GhTT19* silenced LR flower petals. Total anthocyanins were extracted with an acidic methanol solution and measured at 530 and 657 nm. (c) qRT-PCR analysis of *GhTT19* in pCLCrVa:00 and *GhTT19* silenced LR flower petals. (d) The phenotype of LR after *GhTT19* was silenced by VIGS. Part of the petal turns white. (e) qRT-PCR analysis of *GhTT19* in the red and white spot of the petal. The *GhTT19* expression was strongly decreased in the white spot than in the red spot. (f) Overexpression of *GhTT19*^LR^ in W0 (white petal cotton) causes the flower (−1 and 0 day) to turn red. (g) PCR detection of *GhTT19*^LR^ transgenic cotton. (h) Expression of *GhTT19* in OE1, OE2 and the wild-type W0. (i) Quantification of total anthocyanins in OE1, OE2 and W0 flower petals. Total anthocyanins were extracted with an acidic methanol solution and measured at 530 and 657 nm.

production in flowers and fibres of cotton. It has been reported that the genetic positions of red petals (R_3^{bic}) and large basal spots (R_2^{bic}) are different but closely linked (Liang *et al.*, 1997). Our study confirmed that overexpression of *GhTT19*^LR^ in W0 (a white petal cultivar) could caused the petal but without the large basal spot (Figures 3f and S9b). Our genetic analysis revealed that the red petals of LR were controlled by a single dominant gene *GhTT19*^LR^. A recent study reported that *Gar07G08900*, a GST homologue gene, is responsible for the red-spot petals in *G. arboreum* (Zhang *et al.*, 2022). An R2R3-MYB113 could control base spot formation in *Gossypium barbadense* (Abid *et al.*, 2022). These seem to indicate that polyploid cotton has multiple MYB and GST genes in regulating petal and base spot colour in different cotton varieties.

We mapped the R_3^{bic} on chromosome A07 in a 33.2-kb region containing four ORFs (Figure 2a) and found that *ORF3* was a glutamine amidotransferase-like gene. It has been reported that *ORF4* is a methyltransferase MT-A70 gene encoding an mRNA m6A writer protein. Gene functions of *ORF1* and *ORF2* are related to the transport or synthesis of plant pigments. *ORF1* is a homologue of *TT19* involved in the transport of anthocyanins from the ER to the vacuole (Sun *et al.*, 2012). *ORF2* is a phytoene synthase (*PSY*), a key enzyme catalysing two C20 geranylgeranyl diphosphate to form phytoene in the carotenoid pathway (Tanaka *et al.*, 2008). However, *ORF2* did not express in the petals of LW and LR (Figure 2b,c). In our results, Cy3G was the major anthocyanin molecule in the red petal of LR (Figure 1c,d), ORF1-GST GhTT19 bond to Cy3G, and *GhTT19*'s expression was positively associated with Cy3G accumulation in LR petals (Figures 2b–d and 5a). Thus, we suggested that *GhTT19* of R_3^{bic} might be the locus for the red pigmentation in cotton flowers. The SNP primers that could distinguish *GhTT19*^LR^ from *GhTT19*^LW^ clearly separated petal phenotypes between red and white in the genetic population (Figure S2). In addition, *GhTT19* could rescue the anthocyanin accumulation in *tt19*, a mutant of *Arabidopsis* (Figure S6b), and its down-regulation in LR could convert the petal colour from red into light pink (Figures 3a–e and S8). Overexpression of *GhTT19*^LR^ in W0 resulted in petals from white to pink (Figure 3f–i). In the F$_2$ independent segregated population, all red flower individuals had high expression levels of *GhTT19*, while all white flower individuals showed no detected *GhTT19* expression (Figure S10). Moreover, gene and promoter cloning, whole genome re-sequencing showed that the R_3^{bic} originated from *G. bickii* (Figures S2–S4). In LR, the *GhTT19*^LR^-*At* gene and promoter fragment replace the *GhTT19*^LW^-*At*. Pyrosequencing of LR cDNA showed that although *GhTT19*^LR^-*Dt* exists in LR, it is not expressed (Figure 2e–g). These results demonstrated that *GhTT19*-*At* located in R_3^{bic} was responsible for the red pigment accumulation in LR flowers.

The biosynthesis, storage and function of anthocyanins usually take place at different subcellular locations (Zhao, 2015; Zhao and Dixon, 2010). TT19-like GSTs are indispensable for anthocyanin pigment accumulation in many plants, likely by binding to anthocyanins and facilitating their translocation from the ER to the vacuole for storage (Sun *et al.*, 2012; Zhao, 2015). We proved that *GhTT19*, which is mainly localized on membranes of both ER and vacuolar, could bind to Cy3G (Figure 4) and might facilitate the Cy3G stability and transport into the vacuole. This is consistent with TT19-like GSTs from other plants. There are eight SNPs including three amino acid changes in *GhTT19* from LR and LW (Figure S2), GhTT19^LR^ and GhTT19^LW^ proteins did not show differences in Cy3G binding capacities and subcellular localizations (Figure 4). Overexpressing either *GhTT19*^LR^ or *GhTT19*^LW^ in *tt19* mutant appeared purple hypocotyl (Figure S6b). These results indicate that the different petal pigmentations between LR and LW flowers might be caused by their different expressing levels.

It was revealed that *GhTT19* transcription in pigmented petals and filaments of LR was higher than that of LW and other tissues (Figure 5a). Indeed, the analysis showed that proGhTT19^LR^ had much stronger activity than proGhTT19^LW^ (Figure 5b,c). *GhPAP1* is a homologous gene of *Arabidopsis AtPAP1* (Figure S12) and is also involved in the regulation of anthocyanins in cotton (Li *et al.*, 2019). Although the CAACTG element, a binding site of MYB, presented in both promoters of *GhTT19*^LW^ and *GhTT19*^LR^, another *MYB* binding element EEC (GAAATTC), was absent in the *GhTT19*^LW^ promoter (Figure S11). EMSA experiments showed that GhPAP1 could bind to both CAACTG and EEC (GAAATTC) elements (Figure 5f). Respective Y1H and EMSA experiments showed that GhPAP1 could more efficiently bind to proGhTT19^LR^ than proGhTT19^LW^ (Figure 5e,f). CAACTG element was verified to exist in all cotton GST promoters with different petal colours, while the EEC (GAAATTC) element only existed in the GST promoters of red petal cotton species (Figure S13). More importantly, *GhPAP1* could activate proGhTT19^LR^ but not proGhTT19^LW^ (Figure 5g). The conclusion was that *GhPAP1* could enhance the proGhTT19^LR^ activity by interaction with the EEC (GAAATTC) element.

It is often observed that more anthocyanin accumulated in LR petals when cultivated at stronger than weaker light intensities (Figure 6a). The major light signalling component, HY5, has been reported to bind to the promoters of flavonol synthase and dihydroflavonol 4-reductase and anthocyanin regulator *PAP1* (An *et al.*, 2017; Lee *et al.*, 2007; Shin *et al.*, 2013). GhHY5 had higher expressing levels in the petal and filament than in other organs and tissues (Figure 6b). The transcriptions of *GhTT19*, *GhPAP1* and *GhHY5* were all more up-regulated under stronger light intensities (Figure 6c). The Y1H, EMSA and transactivation assays clearly proved that GhHY5 could bind to the promoters of *GhPAP1* and activate *GhPAP1* expression (Figure 6d–f). Moreover,

an ACGT-box, ACTCAT elements and core sequence of GT box (Figure S11) were found in the promoter of *GhTT19^{LR}* but not in that of *GhTT19^{LW}*, as it has been known that these cis-elements are typical binding sites by bZIP TFs (Chen *et al*., 2016; Lee *et al*., 2007; Satoh *et al*., 2004). Y1H assay clearly showed that the GhHY5 only binds to *proGhTT19^{LR}* but not to *proGhTT19^{LW}*
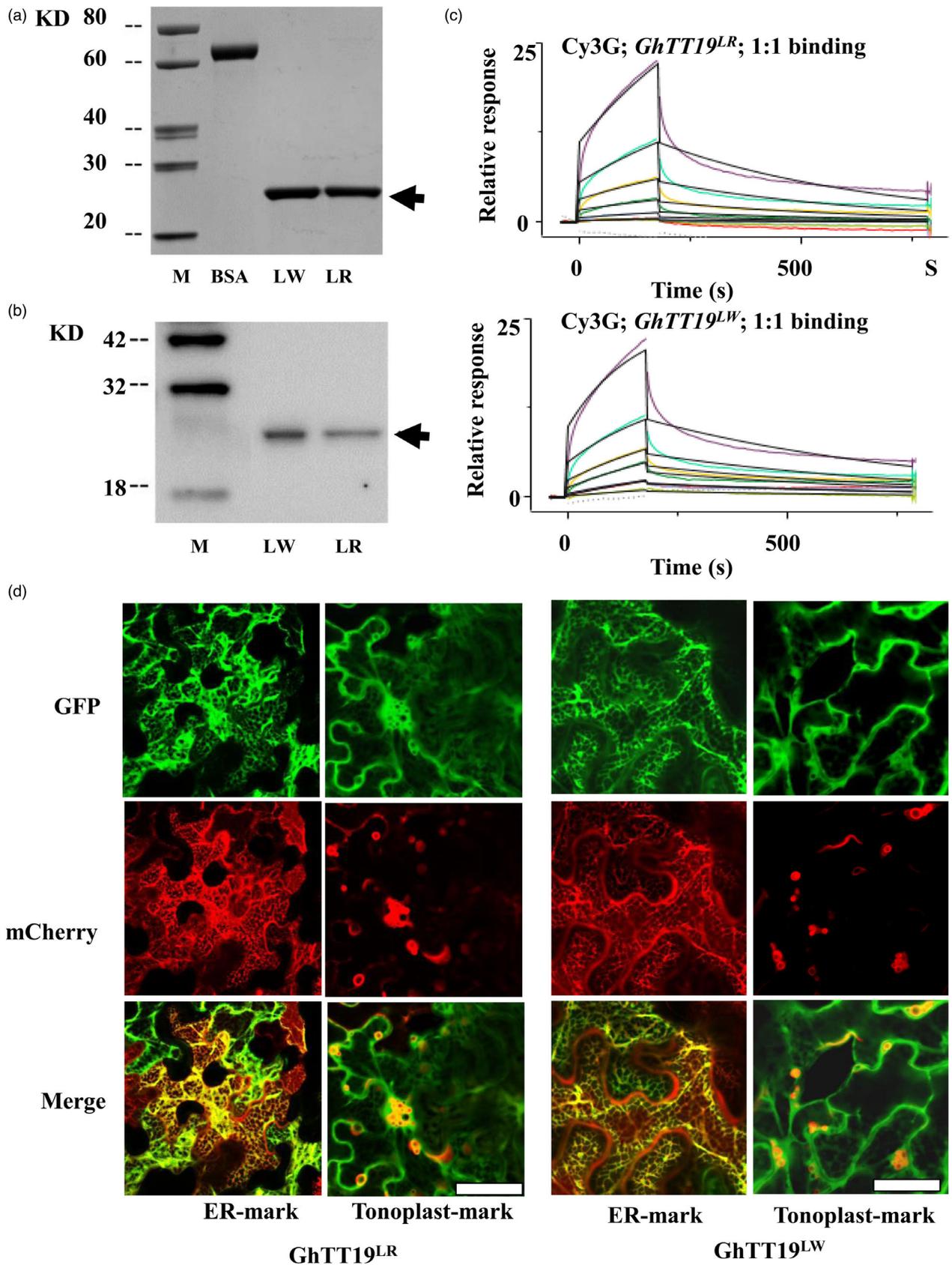
Figure 4 GhTT19 binds to Cy3G *in vitro*. Both GhTT19^{LR} and GhTT19^{LW} proteins are able to bind Cy3G, and have the same subcellular localization. (a) Expression and purification of GhTT19^{LR} and GhTT19^{LW} proteins in *E. coli*. Arrow indicates the IPTG-induced and purified target proteins, which were resolved in a standard SDS–PAGE. (b) Detection of the purified GhTT19 proteins by immunoblotting. (c) The binding affinity of Cy3G to GhTT19^{LR} and GhTT19^{LW}. The binding affinity was assessed using a Biacore 8000 biosensor system. (d) Subcellular localization of GhTT19^{LR} and GhTT19^{LW}. GhTT19 fusion proteins colocalize with ER, tonoplast marker-mCherry fusion and show yellow signals after merging. GhTT19 protein localize in ER and vacuolar membrane. Bars = 50 mm.

(Figure 5e). The transactivation assay showed that *GhHY5* did not activate both *proGhTT19*^{LR} and *proGhTT19*^{LW} (Figure 5g). There has been reported that *Arabidopsis* HY5 can bind to the *AtTT19* promoter (Lee *et al.*, 2007) but GhHY5 bind to the *GhTT19* promoter requires other partners, such as *MYB* transcription factors (An *et al.*, 2017). In our study, GhPAP1 strongly bonded to *proGhTT19*^{LR} and activated its transcription, and the light-regulated anthocyanin accumulation in the red petals was strongly correlated to the expressions of both *GhHY5* and *GhPAP1*. Thus, we concluded that the differential expressions of *GhTT19*^{LR} and *GhTT19*^{LW} were caused by both *GhHY5* and *GhPAP1*. Anthocyanin biosynthetic genes, including *GhCHS, GhCHI, GhF3H, GhF3'H, GhDFR, GhANS1* and *GhUFGT*, and anthocyanin regulatory genes, including *GhTT8, GhGL3* and *GhWD40*, had no significant difference in their expression levels (Figure S15). Anthocyanin transport genes, including *GhTT12, GhGFSa, GhGFSb, GhAHA10* and *GhABCC1* showed no significant difference in FPKM (Figure S16). Anthocyanin biosynthetic pathway genes were also not enriched in the GO enrichment analysis for the differentially expressed genes in the petals between LR and LW plants (Figure S17).

In summary, our study unveiled that the red-petal phenotype of cotton cultivar LR containing one of the anthocyanins, Cy3G, was controlled by a single dominant locus, *R3*^{bic}, in which a candidate *GST* gene, *GhTT19*, was detected via map-based cloning. The genetic experiments through both segregating populations and transgenic lines confirmed that *GhTT19* functioned for anthocyanin accumulations in different genotypes of cotton, such as LR with red flowers and LW with white flowers, this was relative to the GhTT19 localizing mainly on both ER and tonoplast. We found that an allelic variation, an *MYB* binding site EEC element (GAAATTC), exists in *proGhTT19*^{LR}, but not in *proGhTT19*^{LW}. The EEC element was able to enhance the interaction of *GhTT19*^{LR} to GhPAP1 and hence resulted in higher expressions of *GhTT19*^{LR} than *GhTT19*^{LW}. This higher expression of GhTT19^{LR} was positively correlated with the Gy3G contents and petal colours in LR flowers. In addition, *GhHY5*, a major light signal component in cotton, was found to have a special interaction with *GhTT19*^{LR} and promoted its transcription under stronger light intensity. Taken together, our study verified that the *GhHY5*- and *GhPAP1*-regulated expression of *GhTT19*^{LR} played essential roles in light-mediated anthocyanin pigmentation in cotton red flower phenotypes (Figure 7). This allele is an excellent target to breed new cultivars of cotton.

Experimental procedures

Plant materials

Gossypium hirsutum L. acc. LR (hereafter called LR) with red flowers originated from distant hybridization between *G. hirsutum* and wild diploid *G. bickii*. *G. hirsutum* L. acc. LW (here after called LW) with pale white flowers was the near-isogenic line of LR. *G. hirsutum* L. acc. W0, with white petals, was the wild-type

receptor for overexpression *GhTT19*^{LR}. Three F$_2$ populations consisting of 352, 1191 and 136 individuals were developed from a cross between LW as the female parent and LR as the male parent. All offspring plants were grown in Linqing Breeding Station and in the greenhouse, the artificial climate chamber of Jinan. For low light intensity treatment, cotton plants were grown in an artificial climate growth chamber where light intensity was 3200 *Lux*. Genomic DNA was extracted from fresh leaves for genotyping. Wild-type *Arabidopsis* Col-0, *tt19* mutant and transgenic *Arabidopsis* were grown in a 23/21 °C (day/night) chamber. All plant tissues were harvested and immediately frozen in liquid nitrogen, and stored at −80 °C for anthocyanin and RNA analyses.

Anthocyanin analysis

Around 200 mg of frozen petals were homogenized in liquid nitrogen and suspended in acidic methanol (99% methanol and 1% hydrochloric acid) for 1 h. After centrifugation, the absorbance of 200 μL supernatant at 530 and 657 nm was determined with a microplate reader (Omega, S/N 415-2809, Germany). The relative concentration of anthocyanin was calculated using the equation: $Q = [A530 - (0.25 \times A657)] \times M^{-1}$, where M and Q represent the fresh weight and anthocyanin relative concentration, respectively.

High-performance LC–tandem electrospray–MS (LC–ESI–ToF–MS/MS) analysis

To identify anthocyanin molecules, ~2.0 g petals were ground in liquid nitrogen into powder and suspended in 20 mL methanol:formic acid (9 : 1) for extraction three times by ultrasonication. The extracts were concentrated to 5 mL and purified with C18 Sep-pak solid phase extraction (500 mg/8 mL, GRACE, USA), with a mixed solution of methanol:water:formic acid (4 : 5 : 1) as eluent (10 mL). Elution solutions were subjected to Agilent 1260 UPLC-DAD-6530 ESI-QTOF MS (Agilent Technologies, Palo Alto, CA) equipped with Agilent Poroshell 120 SB-Aq C18 column (4.6 × 100 mm, 2.7 μm). LC–ESI–ToF–MS/MS analyses were carried out in accordance with a method described previously (Zhao *et al.*, 2011). MS data were managed with the MassHunter B0.05.0 Workstation.

Map-based cloning of *R$_3$*^{bic} gene

To map the gene responsible for the phenotype of red petals in LR, an F$_2$ mapping population with 352 individuals was developed by a cross between LR and LW in 2015. Based on the genome sequence of TM-1, we developed simple sequence repeat molecular markers (Wang *et al.*, 2015). The *R$_3$*^{bic} locus was mapped to a 33.2 kb region flanked by markers of A07-0594 and A07-0592 on Chromosome A07. The ORF and promoter sequences of the candidate genes were amplified from LR and LW varieties, respectively, and confirmed by sequencing. SNP primer S273 (Table S1), which could specifically amplify the 288th-A genotype, but not the 288th-G genotype, was

(a) GhTT19

(b)

(c)

(d) GhPAP1

(e)

(f)

(g)
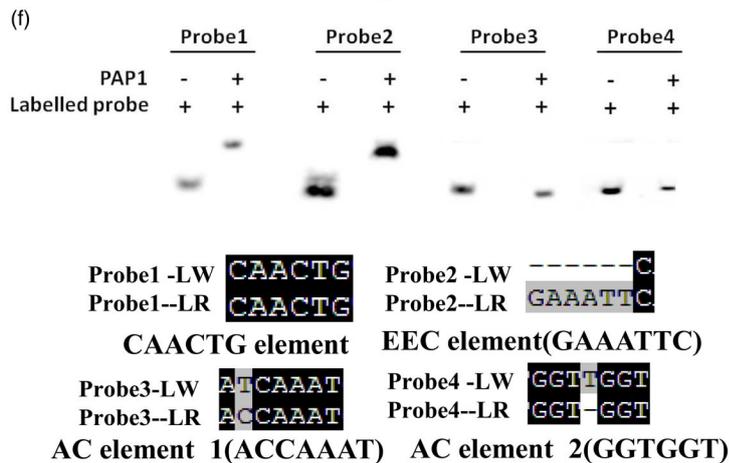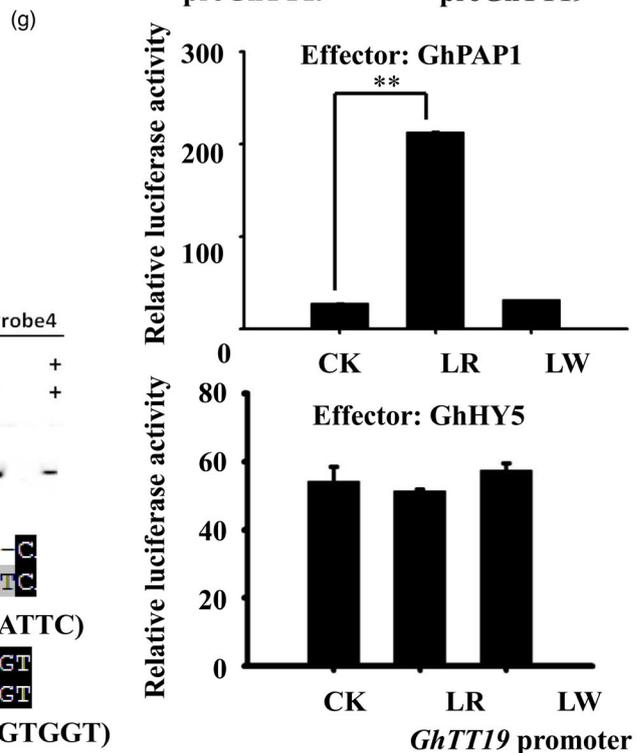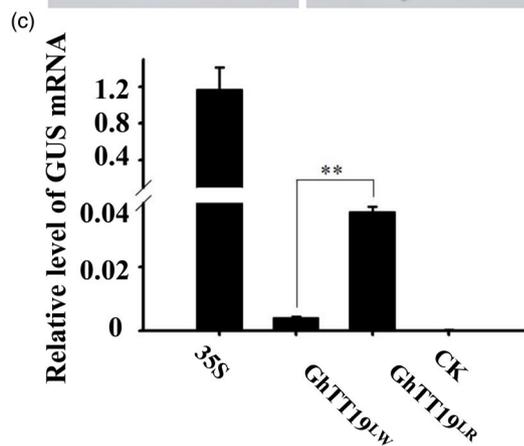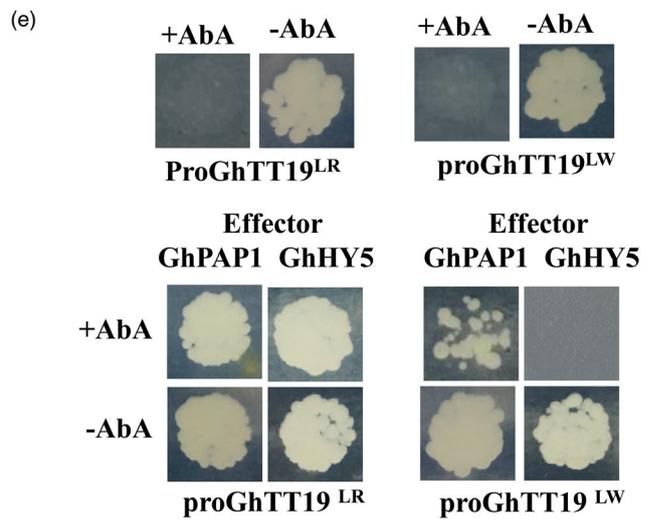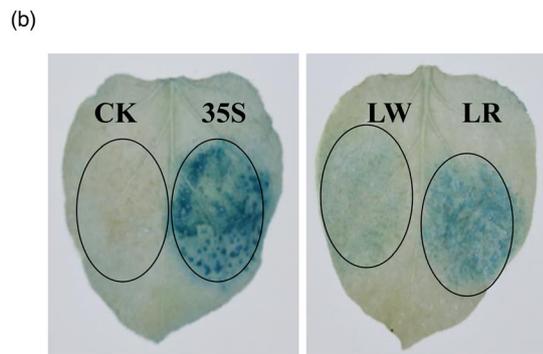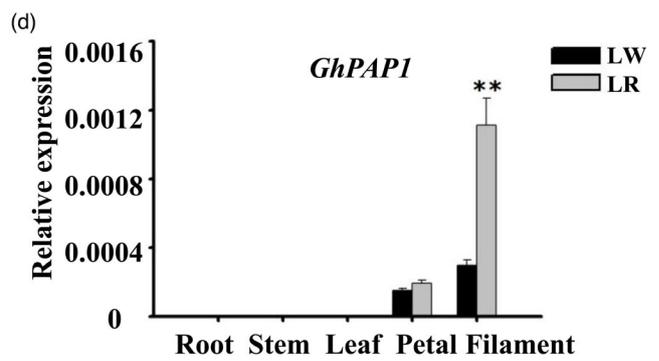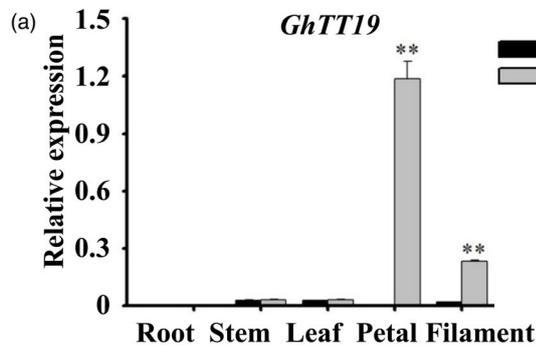
Figure 5 GhPAP1 bound to the *GhTT19^LR* promoter but not efficiently bond to the *GhTT19^LW* promoter. (a) Expression patterns of *GhTT19* in tissues of LR and LW plants. (b) Promoter activity of *GhTT19^LR* and *GhTT19^LW* genes. Vectors were introduced into *Agrobacterium* strain GV3101 and infiltrated into the *Nicotiana benthamiana* leaves. The activity of β-glucuronidase (GUS) was examined using GUS staining. (c) The level of GUS gene expression using qRT-PCR. (d) Expression of *GhPAP1* in different tissues of LR and LW plants. (e) Y1H assay of the binding of *GhPAP1* and *GhHY5* to promoters *pGhTT19^LR* and *pGhTT19^LW*. (f) Electrophoretic mobility shift assay (EMSA) result showing the interaction between GhPAP1 and labelled probes in the promoter of *GhTT19^LR* and *GhTT19^LW*. GhPAP1 could bind to CAACTG element and EEC element. While the EEC element only exists in the promoter of *GhTT19^LW*. (g) Relative luciferase activities were determined in the transient expression assays. Data are presented as means ± SE from at least three biological replicates. **P < 0.01, *P < 0.05 at Student's *t*-test.

developed based on point mutation and used in genotyping the enlarged F_2 population with 1191 individuals.

Genetic complementation of *Arabidopsis tt19* mutant

The ORFs of *GhTT19* genes from LR and LW cotton varieties (*GhTT19^LR* and *GhTT19^LW*, respectively) were cloned into the pBI121 vector under the control of a 35 S promoter. The vectors 35 S::*GhTT19^LR* and 35 S::*GhTT19^LW* were transformed into *Agrobacterium* strain GV3101 for the transformation of *Arabidopsis tt19* mutant using the floral dip method. Seeds from the T1 transgenic *Arabidopsis* were screened on 1/2 Murashige and Skoog medium containing 50 μg/mL kanamycin and 4% (w/v) sucrose. T_3 plants were used for phenotyping.

Expression and purification of recombinant GhTT19^LW and GhTT19^LR

The ORFs of *GhTT19^LW* and *GhTT19^LR* were subcloned into a pET-30a expression vector. The bacterially expressing and purification of GhTT19^LW and GhTT19^LR recombinant proteins were conducted as described previously (Chai *et al*., 2017). Proteins were analysed using SDS–PAGE and immunoblotting. Horseradish peroxidase-conjugated anti-His antibody was used for immunoblotting.

Cy3G-GhTT19 binding assay using Biacore 8000

Cy3G binding affinity to GhTT19 was assessed using a Biacore 8000 biosensor system (GE Healthcare). A CM5 sensor chip was activated with a 1 : 1 mixture of EDC (0.4 M 1-ethyl-3-3-dimethylaminopropyl carbodiimide) and NHS (0.1 M N-hydroxysuccinimide in H_2O). Bacterially expressed and purified GhTT19^LW and GhTT19^LR proteins were diluted in 10 mᴍ sodium acetate (pH 4.5) and were immobilized at a flow rate of 10 μL/min for 7 min. The remaining binding sites were blocked with 1 M ethanolamine hydrochloride (pH 8.5) at a flow rate of 10 μL/min for 7 min. Affinity measurement was performed at 25 °C with HBS-EP (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% Tween 20, pH 7.4) as running buffer. A series of Cy3G concentrations were injected over the ligand surface consecutively as an association phase (contact time 180 s). Running buffer was injected as the dissociation phase (contact time 600 s). The flow rate was 30 μL/min. All data were processed using the Biacore Evaluation software version 1.1.

Promoter analysis

The promoters (about 1.6 kb) of *GhTT19^LW* and *GhTT19^LR* genes were isolated using primers S128F/R (Table S1). The promoter fragments were cloned into vector pBI121 by replacing the 35 S promoter and then sequencing-confirmed constructs were introduced into *Agrobacterium* strain GV3101. The remaining steps were performed as described (Wan *et al*., 2016). For the transactivation assay, the ORFs of *GhPAP1* and *GhHY5* were recombined into the P2GW7 effector expression vector. The promoters

of *GhTT19^LR*, *GhTT19^LW* and *GhPAP1* were cloned into the pGreen-0800-LUC reporter vector. The effector expression and reporter constructs were transformed into *Arabidopsis* protoplasts, and the transactivation assay was performed as described previously (Zhao *et al*., 2021).

## RNA extraction, quantitative RT-PCR analysis

Total RNA was extracted from different tissues using the Biospin Plant Total RNA Extraction Kit (Hangzhou Bori Technology, China). First-strand cDNA was generated using HiScrip Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme, China) according to the manufacturer's instructions. Real-time PCR was performed using a LightCycler FastStart DNA Master SYBR Green I kit (Roche) in a LightCycler 480 II (Roche) detection system. Histone 3 gene was selected as the internal reference. The relative expression levels were calculated using the cycle threshold (Ct) $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). All primers are listed in Table S1.

## Genome resequencing analysis and transcriptome sequencing

Young leaves of LR and LW were collected for DNA isolation. Genomic DNA was extracted using the CTAB method. The task of genome resequencing was performed on an Illumina HiSeq2000™ by Biomarker Technologies (Beijing, China). The sequences of clean reads were mapped to the TM-1 genome (http://cotton.zju.edu.cn/). The difference sequences from the A07 chromosome of LR and LW were mapped to *G. bickii* genome (www.cottongen.org) (Sheng *et al*., 2022). Red and white petals at full bloom day from three different plants were chosen for transcriptome profiling. Sequencing was performed on the Illumina HiSeq™ system (LC Sciences). The sequences of clean reads were mapped to the TM-1 genome (http://cotton.zju.edu.cn/). The expression levels of genes were calculated by quantifying the cDNA fragments per kilobase of transcript per million fragments mapped (FPKM).

Yeast-one-hybrid (Y1H) assay

The ORFs of *GhHY5* and *GhPAP1* were cloned into a yeast one-hybrid vector pGADT7 in the reading frame with AD domain to make constructs AD-*GhHY5* and AD-*GhPAP1* fusions, respectively. Meantime, 1652 bp, 1694 bp and 1600 bp of promoter regions of *GhTT19^LW*, *GhTT19^LR* and *GhPAP1* genes were cloned into a yeast vector pAbAi to generate Matchmaker™ Gold Yeast One-Hybrid assay constructs *proGhTT19^LW*, *proGhTT19^LR* and *proGhPAP1* constructs, respectively. Y1H Gold strain was used for the transformation of the above-mentioned constructs for yeast one-hybrid assays with URA3 as a selectable marker. The positive transformants were further texted on the YNB medium lacking uracil and leucine and containing 100 ng/mL Aureobasidin A (AbA). The empty vector was used as a control for excluding the self-activation and false-positive results.
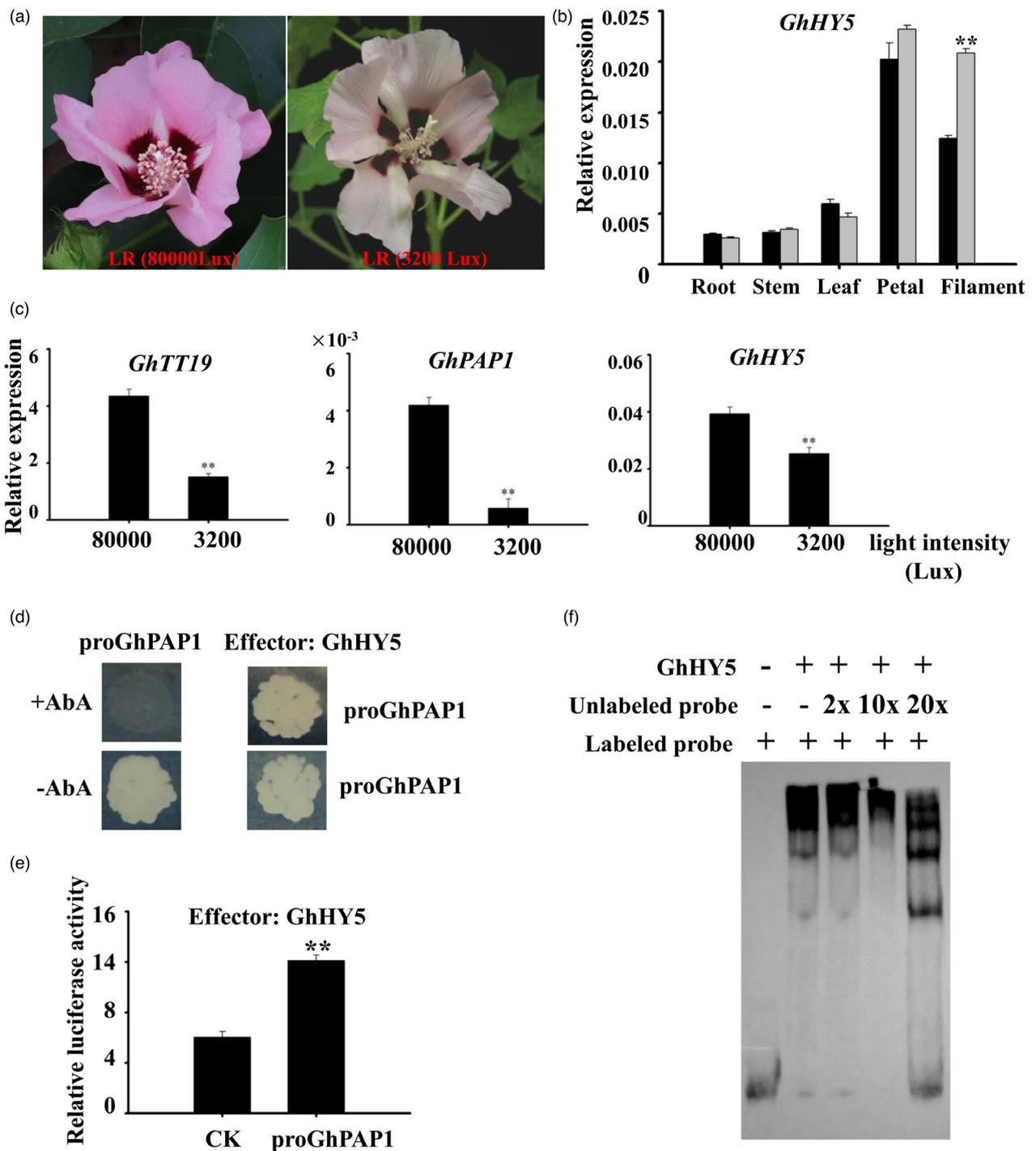
Figure 6 *GhHY5* can regulate *GhPAP1*, resulting in the change of petal colour under different light conditions. (a) The flower phenotypes of LR and LW under different light intensities (in the field or in the climatic chamber) on flowering day. (b) Expression patterns of *GhHY5* in tissues of LR and LW plants. (c) Expression of *GhTT19*, *GhPAP1* and *GhHY5* in LR plant under different light intensities intensity (in the field or in the climatic chamber) on a flowering day. (d) Y1H assay of the binding of GhHY5 to the promoter of *GhPAP1*. (e) Relative luciferase activities were determined in the transient expression assays. GhHY5 can activate the promoter of *GhPAP1*. (f) Electrophoretic mobility shift assay (EMSA) result. Each of the DNA probes, which were generated by PCR, was labelled at one of its 5′-ends with biotin incorporated during primer synthesis. Prob: 332 bp. For competitive experiments, DNA probes without biotin labeling. Data are presented as means ± SE from at least three biological replicates. **$P < 0.01$, *$P < 0.05$ at Student's t-test. The representative photos were shown in Y1H assays.
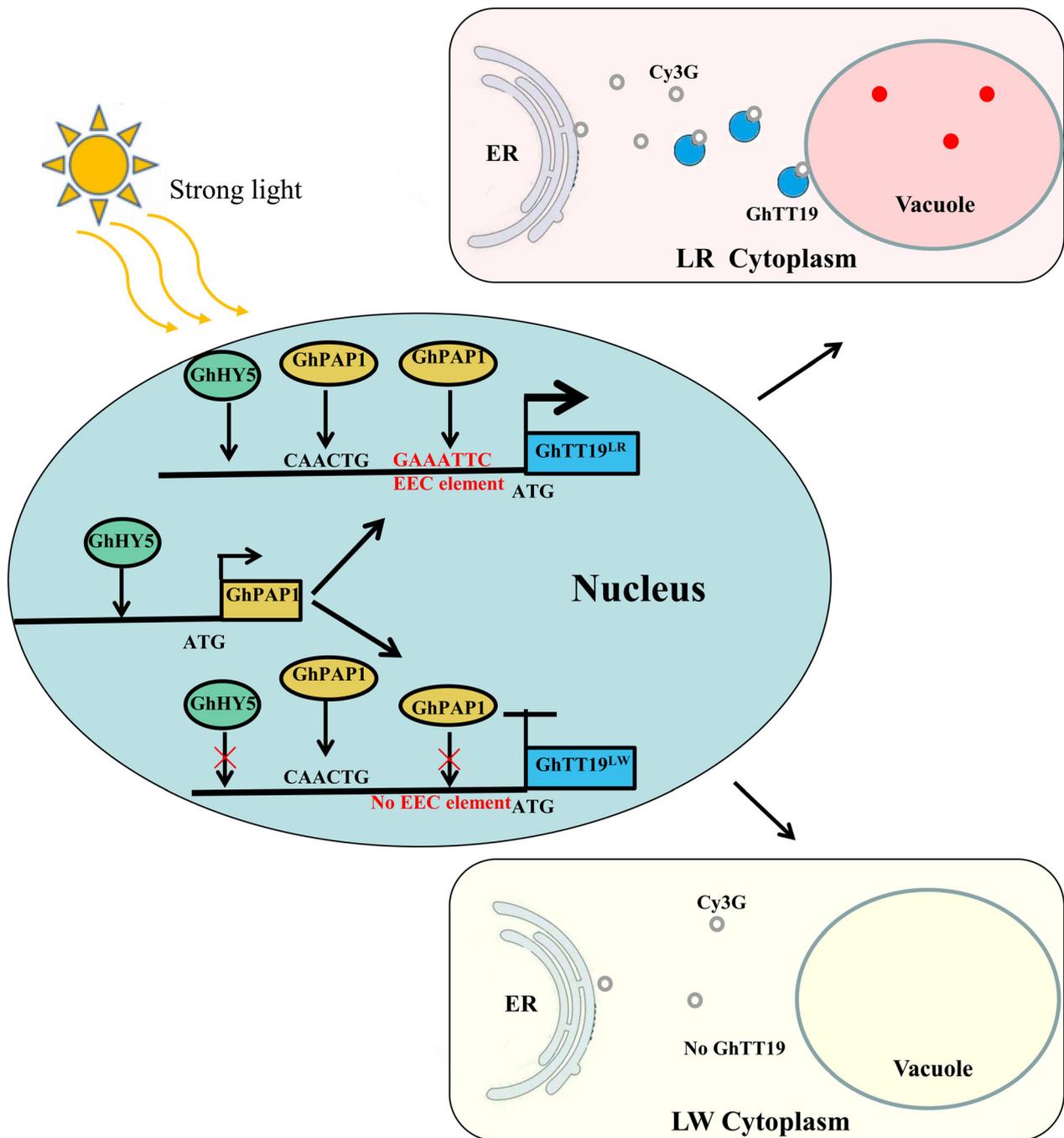
Figure 7 Proposed model for transcriptional regulation of *GhTT19* and its role in anthocyanin accumulation. GhPAP1 bind to the CAACTG element and the EEC element (GAAATTC) from the promoter of *GhTT19^{LR}* and activate its expression. GhTT19 binds to Cy3G and facilitates its transport to the vacuole for storage and coloration. However, in LW plants, the EEC element is missing. GhPAP1 binds weakly to the promoter of *GhTT19^{LW}* and could not activate *GhTT19^{LW}* expression. Strong light irradiation (e.g., in the field) mediates the activation of *GhHY5*, which further induces the expression of the anthocyanin biosynthesis regulator *GhPAP1*. While, under weak light conditions, for example, in the climatic chamber, the *GhHY5* expression level becomes lower than that under strong light, *GhPAP1* expression is also reduced, and causes lower expression of *GhTT19^{LR}*, and thereby reduced red pigmentation in petals.

EMSA assay

The ORF of *GhPAP1* and *GhHY5* were cloned into the pGS-21a and pET-30a respectively. The recombinant protein expression and purification were carried out as described (Chai *et al.*, 2017). All DNA probes were obtained from Tsingke Biotechnology Company (Beijing, China). Images were obtained using a Chemiscope 3300 (Clinx Science Instruments Company).

Subcellular localization of *GhTT19*

The ORF of *GhTT19^{LR}* and *GhTT19^{LW}* was fused with *GFP* in the pBinGFP4 vector. The vector was transiently expressed in tobacco (*Nicotiana benthamiana*) leaf cells as described (Chai *et al.*, 2017). An endoplasmic reticulum (ER) marker HDEL protein (pt-rk-CD3-999) and a tonoplast intrinsic protein fused with mCherry (Nelson *et al.*, 2007) were used as ER and tonoplast markers. Subcellular

localization of GhTT19-GFP was observed under a confocal laser scanning microscope (Zeiss LSM 780) with excitation and emission at 488 and 495–530 nm, respectively, for GFP and at 587 and 600–650 nm, respectively, for mCherry.

Virus-induced gene silencing

A 407-bp fragment (corresponding to bases 370–776) of *GhTT19* was cloned into pCLCrVa and named pCLCrVa: *GhTT19*. pCLCrVa: *GhTT19* and pCLCrVb were introduced into *Agrobacterium* strain LBA4404. The transformed *Agrobacterium* were grown overnight at 28 °C and then resuspended in infiltration media (10 mM MES, 10 mM MgCl$_2$ and 200 μM acetosyringone) and adjusted to OD600 = 1.0. Standing the cells at room temperature for 3 h. The *Agrobacterium* strain containing pCLCrVa: *GhTT19* and pCLCrVb mixed at a ratio of 1 : 1. The LR cotton seedlings were infiltrated with the *Agrobacterium* cells on cotyledons. The plants were grown in an artificial climate chamber at 26 °C/23 °C (day/night) on a 16/8 h light–dark period. The light intensity was 22 000 *Lux*.

Cotton transformation

The full length of *GhTT19^{LR}* was cloned into the pBI121 vector under the control of the 35S promoter. The construct was introduced into wild-type *G. hirsutum* L. acc.W0 (white petal) using *Agrobacterium* tumefaciens-mediated transformation as described (Li *et al*., 2009).

Pyrosequencing analysis

The pyrosequencing analysis of the LR cDNA samples were performed according to the manufacturer's recommendations. Bisulphite-converted cDNA were subjected to PCR amplification. The primers were designed using PyroMark Assay Design 2.0. One of the primers was biotin labelled. The primer sequences are listed in Table S1. KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA KK5701) was used to perform the PCR reaction. To confirm the quality, the PCR products were checked by 1.5% agarose gel electrophoresis. A 25 μL of each PCR product was then subjected to quantitative pyrosequencing analysis using a PyroMark Q48 system (Qiagen) according to manufacturer's instruction.

Statistical analysis

All the experiments were repeated at least three times on three biological replicates. The data were subjected to Student's *t*-test to detect statistical differences. *$P < 0.05$ and **$P < 0.01$.

Accession numbers

All sequences are available at DDBJ/EMBL/GenBank under the accession numbers: MN444040 (*GhTT19-At* of *G. hirsutum* acc. LW), MN4444041(*GhTT19-Dt* of *G. hirsutum* acc. LW), MN4444042 (*GhTT19-At* of *G. hirsutum* acc. LR) and MN4444043 (*GhTT19-Dt* of *G. hirsutum* acc. LR). The Illumina RNA-seq and genome resequencing data are available at the Sequence read Archive under accession number PRJNA878950.

Competing interests

The authors declare no conflict of interest.

Author contributions

JSZ and QC designed the research; QC, XW, MG, XZ, YC, CZ, HJ, JW, YW, MZ, AMB and JZ performed research; QC, XW and XZ analysed the data; JSZ, QC and JZ wrote the paper. All authors discussed the results and commented on the manuscript.

References

Abid, M.A., Wei, Y.X., Meng, Z.G., Wang, Y., Ye, Y., Wang, Y., He, H. *et al*. (2022) Increasing floral visitation and hybrid seed Production mediated by beauty mark in *Gossypium hirstum*. *Plant Biotechnol. J.* **20**, 1274–1284.

An, J.P., Qu, F.J., Yao, J.F., Wang, X.N., You, C.X., Wang, X.F. and Hao, Y.J. (2017) The bZIP transcription factor MdHY5 regulates anthocyanin accumulation and nitrate assimilation in apple. *Hortic. Res.* **4**, 17023.

Ballester, A.R., Molthoff, J., de Vos, R., Hekkert, B., Orzaez, D., Fernández-Moreno, J.P., Tripodi, P. *et al*. (2010) Biochemical and nolecular analysis of pinktomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiol.* **152**, 71–84.

Chai, Q., Shang, X., Wu, S., Zhu, G., Cheng, C., Cai, C., Wang, X. *et al*. (2017) 5-aminolevulinic acid dehydratase gene dosage affects programmed cell death and immunity. *Plant Physiol.* **175**, 511–528.

Chen, J., Yi, Q., Cao, Y., Wei, B., Zheng, L., Xiao, Q., Xie, Y. *et al*. (2016) ZmbZIP91 regulates expression of starch synthesis-related genes by binding to ACTCAT elements in their promoters. *J. Exp. Bot.* **67**, 1327–1338.

Conn, S., Curtin, C., Bezier, A., Franco, C. and Zhang, W. (2008) Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins. *J. Exp. Bot.* **59**, 3621–3634.

Espley, R.V., Hellens, R.P., Putterill, J., Stevenson, D.E., Kutty-Amma, S. and Allan, A.C. (2007) Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* **49**, 414–427.

Francisco, R.M., Regalado, A., Ageorges, A., Burla, B.J., Bassin, B., Eisenach, C., Zarrouk, O. *et al*. (2013) ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-Glucosides. *Plant Cell*, **25**, 1840–1854.

Gangappa, S.N. and Botto, J.F. (2016) The multifaceted roles of HY5 in plant growth and development. *Mol. Plant*, **9**, 1353–1365.

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verriès, C., Souquet, J.M. *et al*. (2009) Grapevine MATE-type proteins act as vacuolar H+-dependent acylated anthocyanin transporters. *Plant Physiol.* **150**, 402–415.

Gomez, C., Conejero, G., Torregrosa, L., Cheynier, V., Terrier, N. and Ageorges, A. (2011) In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J.* **67**, 960–970.

Grotewold, E. (2006) The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* **57**, 761–780.

Hinchliffe, D.J., Condon, B.D., Thyssen, G., Naoumkina, M., Madison, C.A., Reynolds, M., Delhom, C.D. *et al*. (2016) The GhTT2_A07 gene is linked to the brown colour and natural flame retardancy phenotypes of Lc1 cotton (*Gossypium hirsutum* L.) fibres. *J. Exp. Bot.* **67**, 5461–5471.

Ke, L.P., Yu, D.L., Zheng, H.L., Xu, Y., Wu, Y., Jiao, J., Wang, X. *et al*. (2022) Function deficiency of GhOMT1 causes anthocyanidins over-accumulation and diversifies fibrecolours in cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* **20**, 1546–1560.

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W. *et al.* (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell*, **19**, 731–749.

Li, F.F., Wu, S.J., Chen, T.Z., Zhang, J., Wang, H.H., Guo, W.Z. and Zhang, T.Z. (2009) Agrobacterium-mediated co-transformation of multiple genes in upland cotton. *Plant Cell Tissue Organ Culture*, **97**, 225–235.

Li, X., Ouyang, X., Zhang, Z., He, L., Wang, Y., Li, Y., Zhao, J. *et al.* (2019) Over-expression of the red plant gene R1 enhances anthocyanin production and resistance to bollworm and spider mite in cotton. *Mol. Genet. Genomics*, **294**, 469–478.

Li, S.Y., Zuo, D.Y., Cheng, H.L., Ali, M., Wu, C., Ashraf, J., Zhang, Y. *et al.* (2020) Glutathione S-transferases GhGSTF1 and GhGSTF2 involved in the anthocyanin accumulation in *Gossypium hirsutum* L. *Int. J. Biol. Macromol.* **165**, 2565–2575.

Liang, Z., Jiang, R. and Zhong, W. (1997) New red flower germplasm lines of cotton selected from hybrid of *Gossypium hirsutum. bickii*. *Sci. China Series C, Life Sci.* **40**, 284–292.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{(-\Delta\Delta CT)}$ method. *Methods*, **25**, 402–408.

Marrs, K.A., Alfenito, M.R., Lloyd, A.M. and Walbot, V. (1995) A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature*, **375**, 397–400.

Nelson, B.K., Cai, X. and Nebenführ, A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant J.* **51**, 1126–1136.

Perez-Diaz, R., Madrid-Espinoza, J., Salinas-Cornejo, J., Gonzalez-Villanueva, E. and Ruiz-Lara, S. (2016) Differential roles for VviGST1, VviGST3, and VviGST4 in proanthocyanidin and anthocyanin transport in *Vitis vinifera*. *Front. Plant Sci.* **7**, 1166.

Satoh, R., Fujita, Y., Nakashima, K., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004) A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the *ProDH* gene in *Arabidopsis*. *Plant Cell Physiol.* **45**, 309–317.

Shao, D.N., Li, Y.J., Zhu, Q.H., Zhang, X.Y., Liu, F., Xue, F. and Sun, J. (2021) *GhGSTF12*, a glutathione S-transferase gene, is essential for anthocyanin accumulation in cotton (*Gossypium hirsutum* L.). *Plant Sci.* **305**, 110827.

Shao, D.N., Zhu, Q.H., Liang, Q., Wang, X., Li, Y., Sun, Y., Zhang, X. *et al.* (2022) Transcriptome analysis reveals differences in anthocyanin accumulation in cotton (*Gossypium hirsutum* L.) induced by red and blue light. *Front. Plant Sci.* **13**, 788828.

Sheng, K., Sun, Y., Liu, M., Cao, Y., Han, Y., Li, C., Muhammad, U. *et al.* (2022) A reference-grade genome assembly for *Gossypium bickii* and insights into its genome evolution and formation of pigment gland and gossypol. *Plant Commun.* **10**, 100421. https://doi.org/10.1016/j.xplc.2022.100421

Shin, D.H., Choi, M., Kim, K., Bang, G., Cho, M., Choi, S.B., Choi, G. *et al.* (2013) HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in *Arabidopsis*. *FEBS Lett.* **587**, 1543–1547.

Su, W., Tao, R., Liu, W. *et al.* (2019) Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* **18**, 479–490.

Sun, Y., Li, H. and Huang, J.R. (2012) *Arabidopsis* TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol. Plant*, **5**, 387–400.

Tan, J., Tu, L., Deng, F., Hu, H., Nie, Y. and Zhang, X. (2013) A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiol.* **162**, 86–95.

Tanaka, Y., Sasaki, N. and Ohmiya, A. (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* **54**, 733–749.

Wan, Q., Guan, X., Yang, N., Wu, H., Pan, M., Liu, B., Fang, L. *et al.* (2016) Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. *New Phytol.* **210**, 1298–1310.

Wang, Q., Fang, L., Chen, J., Hu, Y., Si, Z., Wang, S., Chang, L. *et al.* (2015) Genome-wide mining, characterization, and development of microsatellite markers in gossypium species. *Sci. Rep.* **5**, 10638.

Wang, N., Zhang, B.B., Yao, T., Shen, C., Wen, T., Zhang, R., Li, Y. *et al.* (2022) Re enhances anthocyanin and proanthocyanidin accumulation to produce red foliated cotton and brown fiber. *Plant Physiol.* **189**, 1466–1481.

Winkel-Shirley, B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* **126**, 485–493.

Xu, D. (2020) COP1 and BBXs-HY5-mediated light signal transduction in plants. *New Phytol.* **6**, 1748–1753.

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., Lepiniec, L. *et al.* (2014) Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *New Phytol.* **202**, 132–144.

Xu, W., Dubos, C. and Lepiniec, L. (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **20**, 176–185.

Yan, Q., Wang, Y., Li, Q., Zhang, Z., Ding, H., Zhang, Y., Liu, H. *et al.* (2018) Up-regulation of GhTT2-3A in cotton fibres during secondary wall thickening results in brown fibres with improved quality. *Plant Biotechnol. J.* **16**, 1735–1747.

Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C.M., Zhang, C. *et al.* (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* **10**, 1494.

Zhang, S.J., Chen, J., Jiang, T., Cai, X., Wang, H., Liu, C., Tang, L. *et al.* (2022) Genetic mapping, transcriptomic sequencing and metabolic profiling indicated a glutathione S-transferase is responsible for the red-spot-petals in *Gossypium arboreum*. *Theor. Appl. Genet.* **135**, 3443–3454.

Zhao, J. (2015) Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* **20**, 576–585.

Zhao, J. and Dixon, R.A. (2010) The 'ins' and 'outs' of flavonoid transport. *Trends Plant Sci.* **15**, 72–80.

Zhao, J., Huhman, D., Shadle, G., He, X.Z., Sumner, L.W., Tang, Y. and Dixon, R.A. (2011) MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. *Plant Cell*, **23**, 1536–1555.

Zhao, X., Zeng, X., Lin, N., Yu, S., Fernie, A.R. and Zhao, J. (2021) CsbZIP1-CsMYB12 mediates the production of bitter-tasting flavonols in tea plants (*Camellia sinensis*) through a coordinated activator-repressor network. *Hortic. Res.* **8**, 110.

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., Yang, Z., Xie, X. *et al.* (2017) Development of "purple endosperm rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Mol. Plant*, **10**, 918–929.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Cotton flowers from *G. bickii*, LW and LR.

Figure S2 Alignment of the CDS sequences of *GhTT19*.

Figure S3 Alignment of the promoter sequences of *GhTT19* from LR, LW and *G. bickii*.

Figure S4 SNP detection and prediction of variability between LR and LW using genome resequencing.

Figure S5 FPKM expression values and GO enrichment analysis of genes, from *G. bickii* introgressed segment of A07 regions, in the petals of LR and LW.

Figure S6 *GhTT19^{LR}* and *GhTT19^{LW}* can complement the anthocyanin deficiency phenotype of *Arabidopsis tt19* mutant.

Figure S7 Amino acid alignments of homologue proteins of AtTT19 from different species.

Figure S8 The phenotype of LR after *Sulphur* (*Su*) silenced by VIGS.

Figure S9 The phenotype of *GhTT19* overexpression lines.

Figure S10 qRT-PCR analysis of *GhTT19* in the F_2 population.

Figure S11 Alignment of the promoter sequences of *GhTT19^{LR}* and *GhTT19^{LW}*.

Figure S12 Identification and characterization of cotton *PAP1* genes.

Figure S13 The CAACTG element and EEC element, the *PAP1* binding sites, on the *GhTT19* promoter of different cotton with diverse genetic backgrounds.

Figure S14 Transcriptional profiling of bZIP genes in different tissues in *G. hirsutum* acc.TM-1.

Figure S15 Expression level of genes involved in anthocyanin biosynthesis, and regulation analysed by qRT-PCR.

Figure S16 Transcript abundance of genes involved in antho-cyanin transport analysed by RNA-seq.

Figure S17 GO enrichment analysis of the differentially expressed genes in petals between LR and LW.

Table S1 All primers developed and used in the present research.

Table S2 Putative functions of candidate genes for red petal phenotype.

Table S3 ORFs and its Dt homologue gene sequences in LR, LW and *G. bickii*.

Table S4 Differential genes on chromosome A07 between LR and LW and their expression patterns in petals.