Genetic mapping, transcriptomic sequencing and metabolic profiling indicated a glutathione S-transferase is responsible for the red-spot-petals in *Gossypium arboreum*

Sujun Zhang¹ · Jie Chen³ · Tao Jiang² · Xiao Cai¹ · Haitao Wang¹ · Cunjing Liu¹ · Liyuan Tang¹ · Xinghe Li¹ · Xiangyun Zhang¹ · Jianhong Zhang¹

Received: 12 May 2022 / Accepted: 31 July 2022 / Published online: 19 August 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

*Key message* A GST for red-spot-petals in *Gossypium arboreum* was identified as the candidate under the scope of multi-omics approaches.

*Abstract* Colored petal spots are correlated with insect pollination efficiency in *Gossypium* species. However, molecular mechanisms concerning the formation of red spots on *Gossypium arboreum* flowers remain elusive. In the current study, the Shixiya1-R (SxyR, with red spots) × Shixiya1-W (SxyW, without red spots) segregating population was utilized to determine that the red-spot-petal phenotype was levered by a single dominant locus. This phenotype was expectedly related to the anthocyanin metabolites, wherein the cyanidin and delphinidin derivatives constituted the major partition. Subsequently, this dominant locus was narrowed to a 3.27 Mb range on chromosome 7 by genomic resequencing from the two parents and the two segregated progeny bulks that have spotted petals or not. Furthermore, differential expressed genes generated from the two bulks at either of three sequential flower developmental stages that spanning the spot formation were intersected with the annotated ones that allocated to the 3.27 Mb interval, which returned eight genes. A glutathione S-transferase-coding gene (*Gar07G08900*) out of the eight was the only one that exhibited simultaneously differential expression among all three developmental stages, and it was therefore considered to be the probable candidate. Finally, functional validation upon this candidate was achieved by the appearance of scattered petal spots with inhibited expression of *Gar07G08900*. In conclusion, the current report identified a key gene for the red spotted petal in *G. arboreum* under the scope of multi-omics approaches, such efforts and embedded molecular resources would benefit future applications underlying the flower color trait in cotton.

Introduction

*Gossypium arboreum* L. is a diploid species introduced from India, which possesses many favorable traits that upland cotton cultivars lack, including drought tolerance as well as resistance to diseases and insect pests (Liu et al. 2006). In addition, some *G. arboreum* accessions produce strong fibers as well as high-oil-content seeds (Liu et al. 2006). In most of *G. arboreum* lines, red spots develop at the basal part of petals during flower development and persist lifelong (Zhou 2011). In contrary, only sparse cases have been reported in *G. hirsutum* flowers (Jones 1996; Zhou 2011; Abid et al. 2022). These spots are easily recognizable features, which could enhance the attractiveness for insect pollinators (Jones 1996; Abid et al. 2022). In addition to the red spots, there exist other two petal spot phenotypes in *G. arboreum* accessions: spotless and ghost spot. It was indicated that red pigmented spots were conferred by a single completely dominant gene by crossing an accession had the ghost spot phenotype with another one that had red petal spots (Erpelding 2022). Meanwhile, the petal spots could be readily observed on the adaxial side rather than the abaxial side of petals (Cavallini-Speisser et al. 2021). Despite the universally pronounced phenomenon observed, little researches regarding the
molecular mechanisms or the genetic determinants have been conducted in *G. arboreum*.

Of the many internal and external factors influencing flower coloration, anthocyanin types and concentrations are considered to be major determinants in many species (Weiss 1995; Tanaka et al. 2008). Anthocyanins, which are water-soluble pigments, are a class of flavonoids responsible for the red, purple, and blue coloration of flowers, fruits, and leaves (Zhang et al. 2018). Although anthocyanins are structurally diverse chemicals, there are only six common chromophore forms (glycone) of anthocyanins: cyanidin, delphinidin, pelargonidin, peonidin, petunidin, and malvidin (Jaakola, 2013). The anthocyanin biosynthesis pathway has been well characterized and the genes encoding proteins mediating anthocyanin biosynthesis have been identified in many plant species, including *Arabidopsis thaliana* (Li and Strid 2005), *Petunia hybrida* (Chen et al. 2021), spinach (Cai et al. 2018), alfalfa (Duan et al. 2020) and many fruits (Zhang et al. 2018). Meanwhile, the regulatory transcription factors of anthocyanin biosynthesis have also been identified (Saito et al. 2013). Specifically, the MYB transcription factors are crucial for the accumulation of anthocyanins in fruits, including apple, pear, peach, sweet cherry and the fruits of *Citrus* species (Zhang et al. 2018). In *Theobroma cacao* L., the green/red pod coloration is likely regulated by the MYB transcription factor encoded by *TcMYB113* (Motamayor et al. 2013). In *G. barbadense*, an R2R3 MYB113 transcription factor could control spot formation at the base of flower petals (Abid et al. 2022).

Aside from the biosynthetic genes and the regulatory factors, the anthocyanins are, after being synthesized in the cytosol, generally transported into vacuoles via specific transporters (Zhang et al. 2018), wherein the glutathione S-transferases (GSTs) are likely the most important. Earlier researches revealed visible changes to pigmentation in mutants with loss-of-function mutations of GST-encoding genes, including *tt19* in *A. thaliana*, *bz2* in maize, *an9* in petunia and *fl3* in carnation (Marrs et al. 1995; Alfenito et al. 1998; Larsen et al. 2003; Kitamura et al. 2004; Sun et al. 2012). A recent study has demonstrated one GST that functions downstream of FvMYB10 is the main transporter of the anthocyanins responsible for strawberry foliage and fruit coloration; it can be modified to change strawberry fruit from white to red (Luo et al. 2018).

Although considerable researches on flower coloration have been performed in various species (Sasaki et al. 2015; Jiao et al. 2020; Morimoto et al. 2020; Pu et al. 2020), little is known regarding the formation of red spots on *G. arboreum* flowers. One way to narrow down the scope of candidate genes for downstream sifting and identification is the combination of genomic resequencing and bulked segregant analysis (BSA), which has been utilized to dissect color traits in numerous species. For instance, the *BrTT1* responsible for the seed coat coloration in *Brassica rapa* was determined by combining genomic resequencing and a BSA (Wang et al. 2016). Similarly, the anthocyanin accumulation-related genes in pepper fruits (Wang et al. 2018) and the gene affecting the foliage and fruit coloration in strawberry was also identified (Luo et al. 2018). Additionally, a next-generation sequencing technique for transcriptome analyses (RNA-seq) has facilitated the rapid and cost-effective identification of candidate genes for specific traits (Zhang et al. 2017; Hou et al. 2018; Meng et al. 2019; Jiang et al. 2020; Li et al. 2020). In this study, the red-spot-petal phenotype in *G. arboreum* was related to the content variation of anthocyanin chemicals by metabolic profiling, it was mapped onto a dominant locus on chromosome 7 through the combination of genomic resequencing and BSA. Next, a GST-coding gene was pinpointed as the likely candidate by further integrating the transcriptomic data that spans sequential developmental stages of flowers wherein contrast visual characters regarding colored spots are formed. At last, functional validation was fulfilled by the observation of scattered color spots on petals with inhibited expression via virus-induced gene silencing (VIGS) methodology. To this end, we have successfully identified a key gene involved in the red spot formation of *G. arboreum* flowers. The whole process represents a first, to the best of our knowledge, systematic evaluation of this trait in *G. arboreum*.

### Materials and methods

#### Plant materials

*Gossypium arboreum* inbred line Shixiya1-R (abbreviated as SxyR, same below) produces flowers with red spots at the base of petals, whereas the flowers of inbred line Shixiya1-W (SxyW) lacks red spots at the corresponding position (Fig. 1a–j). SxyR and SxyW were used as the female and male parents, respectively. The segregating population was obtained by self-pollinating. Both parents were planted in three rows from 2014 to 2016 at Xiaoxianhe Experimental Station, Hebei Academy of Agricultural and Forestry Sciences, Hebei, China. Twenty to twenty-two plants per 6-m-long row were retained, and the row spacing was 0.75 m. *F*₁ 

#### Petal metabolite extraction, detection and analysis

Petal samples were collected in the morning (at 9:00 to 10:00 am) when flowers blossom (stages IV as indicated
in Fig. 1). Specifically, the red spotted areas of each flower were quickly cut and snap-frozen in liquid nitrogen, then stored in a −80 °C freezer. Afterwards, these samples were freeze-dried and then ground into fine powder using zirconia beads and a mixer mill (MM 400, Retsch) (1.5 min at 30 Hz). Metabolites were extracted by adding 1.0 mL 70% methanol into 100 mg powder, followed by 4 °C incubation, centrifugation and filtration as previously described (Chen et al. 2020). The sample extracts were subjected to a commercial company (Metware Biotechnology, Wuhan, China) for anthocyanin detection on the 6500 Q TRAP LC–MS/MS system (AB Sciex, CA, USA), using the widely-targeted metabolomics protocol (Chen et al. 2013). Anthocyanin metabolites were identified by comparing the detection signals with the metabolite database (Metware Biotechnology, Wuhan, China), and then were quantified using the multiple reaction monitoring method (Chen et al. 2013; Wang et al. 2017).

Whole genome sequencing and BSA-seq analysis

Four DNA libraries for genomic resequencing were prepared, which included the two bulks containing each of 20 F₂ plants with (Pr) or without (Pw) spotted flowers, and the two parents (SxyR and SxyW). Samples were delivered to a commercial company (Biomarker, Beijing, China) to construct the paired-end and mate-pair Solexa libraries according to the manufacturer’s instructions (Illumina, CA, USA). The resulting DNA libraries were subjected to Solexa sequencing on the Illumina HiSeq™ 2500 platform (Biomarker, Beijing, China, http://www.biomarker.com.cn). Data have been uploaded to NCBI with accessions SAMN28187015 to SAMN28187018.

After the barcodes were trimmed from the high-quality reads, the clean reads for the four libraries were mapped to the G. arboreum reference genome WHU-updated v1 (Huang et al. 2020) using the default parameters of the Burrows–Wheeler Aligner (Li and Durbin 2009). The SAMtools program (Li et al. 2009) was used to mark duplicates, after which GATK (McKenna et al. 2010) was used for the local realignment and base recalibration. A range of filters was applied to obtain highly accurate SNP and InDel sets (Reumers et al. 2012).

To identify candidate regions, a SNP-index association analysis and a Euclidean distance (ED) association analysis were conducted (Abe et al. 2012; Zhang et al., 2022). The LOESS method was used for fitting the ED and delta SNP-index values. The median + 3 standard deviations of the fitted values of all sites were used as the correlation threshold for the ED analysis. The SNP index was calculated after combining the InDel and SNP markers. The 99th percentile of the fitted delta SNP-index values was set as the threshold for the SNP-index association analysis. And the analysis was completed using the BMKCloud platform (http://www.biocloud.net).
RNA extraction, sequencing and analysis

Basal part of petals from at least three Pr and Pw flowers were collected at each of the continuous stages (stages II, III, and IV for RNA-seq, and all five stages for qRT-PCR) when spots develop. All samples were immediately frozen in liquid nitrogen and stored at −80 °C before RNA or metabolites extraction. All experiments were performed using three biological replicates.

Total RNA was extracted using the RNAprep Pure Plant Kit (Tiangen Biotech, Beijing, China). Poly(A) mRNA was enriched from the total RNA using oligo (dT) magnetic beads. The poly(A) mRNA was subsequently fragmented using the RNA Fragmentation Kit (Ambion, TX, USA) and then transcribed into first-strand cDNA using reverse transcriptase and random hexamer primers (NEB, MA, USA). Second-strand cDNA was synthesized using DNA polymerase I and RNase H (Invitrogen, CA, USA). Finally, fragments of suitable lengths were isolated and ligated to adapters and then sequenced on the Illumina HiSeq™ 2500 platform.

Clean reads were obtained from the raw data by removing adapter sequences, low-quality reads and reads containing poly-N sequences, and all clean reads were mapped to the G. arboreum reference genome WHU-updated v1 (Huang et al. 2020). Gene expression levels were estimated by RSEM version 1.2.26 (Li et al. 2011). The differential expressed genes (DEGs) between two samples were identified and analyzed using the DESeq R package (version 1.10.1) at an adjusted p-value < 0.05. The RNA-seq data have been deposited in NCBI with accessions SAMN28185984 to SAMN28185989.

Expression analysis of the genes involved in anthocyanin biosynthesis

The qRT-PCR primers for candidate genes involved in anthocyanin biosynthesis and the G. arboreum histone 3 gene (internal control) are listed in Table S1. The qRT-PCR analysis was performed using SYBR® Premix Ex Taq™ II (Takara, Beijing, China) and the ABI Prism 7500 system (Applied Biosystems, CA, USA). The 2−ΔΔCT method was used to quantify gene expression levels (Schmittgen and Livak 2008).

Virus induced gene silencing (VIGS) assay

Two VIGS vectors, designated as pTRV2:00 and pTRV2:Gar07G08900, were respectively transformed into Agrobacterium tumefaciens strain GV3101, wherein pTRV2:Gar07G08900 harbors a 300-bp product of Gar07G08900 amplified from Pr cDNA and pTRV2:00 is the control. The transformed GV3101 was injected into the spotted parent (i.e., Sxy-R) and descendant (e.g., F2 lines) flower buds around stage II (Fig. 1b) according to previous protocols (Wang et al., 2018; Yang et al., 2019), and the corresponding buds or flowers after inoculation were sampled for qRT-PCR to evaluate the expression levels of the target gene Gar07G08900. Primers are listed in Table S1. Four biological replicates that collected from different plants were applied for the inoculation and subsequent quantification.

Statistical analysis

Data were analyzed using SigmaPlot 14.0. Data are herein presented as the mean ± standard error. The significance of any differences was assessed using p < 0.05 as the threshold.

Results

The red petal spots were levered by a single dominant locus affecting anthocyanin contents

SxyR (Shixiya1-R, with red spots on petals, as indicated in Fig. 1a–e) and SxyW (Shixiya1-W, without spots, Fig. 1f–j) were respectively used as the female and male parents in a hybridization. As a result, all 70 F1 hybrids had red petal spots, while in the F2 segregating population, 161 plants had flowers with red spots and 53 plants had no red petal spots (i.e., 3:1 segregation ratio). Therefore, the red petal spots trait of G. arboreum was likely levered by a single dominant locus.

The color phenotype was largely linked to the accumulation of anthocyanins. To determine if this is the issue in our case, a UPLC/ESI-Q TRAP-MS/MS system was utilized to detect the anthocyanin metabolites of red-spot G. arboreum flowers. A total of 17 anthocyanin metabolites were identified, including procyanidin, cyanidin, peonidin, delphinidin, and malvidin derivatives (Table 1). Of them, cyanidin and delphinidin derivatives were most abundant from the red spot petal site, accounting for more than 90% of the total anthocyanin metabolite contents (Fig. 1k). Meanwhile, four of the six cyanidin derivatives, and both the two delphinidin metabolites exhibited enormously higher contents in spotted SxyR petals than that in non-spotted SxyW positions (log10 (fold-change) > 1 and -log10 (p-value) > 2, as illustrated in Fig. 1k). Compared with this content variation, the rest detected metabolites generally showed less contents and tinier differentiations (Fig. 1k), suggesting the cyanidin and delphinidin metabolites are the major effectors for the red-spot petal phenotype (Fig. 1a–j) in G. arboreum, and candidate genes resided within the single dominant locus were logistically deduced to be responsible for the metabolism or transportation of designated anthocyanins.
Genomic resequencing and genetic mapping

Genotyping the lines that have red spot petals or not is a prerequisite for pinpointing the genetic locus responsible for this phenotype. Therefore, a total of four libraries including the two parents (SxyR and SxyW), and two bulks (Pr-bulk and Pw-bulk) involving F2 descendants that constitute opposite red petal phenotypes were subjected to genomic resequencing. As a result, a total of 70,893,040 (SxyR), 68,122,431 (SxyW), 135,892,656 (Pr-bulk) and 133,225,405 (Pw-bulk) clean reads were obtained respectively, with more than 95% of them from each library mapped to the *G. arboreum* reference genome (Table 2). Overall, a 10× coverage depth for the parental bulks and at least a 21× coverage depth for the two F2 progeny bulks were generated from the high-throughput sequencing (Table 2), which ensures downstream identification of single nucleotide polymorphisms (SNPs) and development of functional markers that linking to the red-spot-petal phenotype.

SNP calling was performed by comparing our resequencing data that successfully mapped to the *G. arboreum* reference genome, which resulted in 286,708 SNPs in the parents and 175,184 SNPs from the two bulked pools. Meanwhile, a total of 64,020 and 18,582 insertion/deletions (InDels) were respectively detected in the parents and the mixed pools (Table S2). The SNPs and InDels were distributed on all the 13 chromosomes of *G. arboreum* (Table S2 and Fig. S1).

The Euclidean distance (ED) methodology was applied to map the dominant locus that linked to the red-spot-petal phenotype, which allocated the QTL, under the threshold of 1.98, to a 10.78 Mb-14.05 Mb region on chromosome 7 (Fig. 2a). This range spans 3.27 Mb and contains 163

| Table 1 | Anthocyanin metabolites in cotton flowers with and without red spots |
| Index | Compounds | Pr contents | Pw contents |
|-------|------------|-------------|-------------|
| pmal590 | Peonidin O-hexoside | 5.8 × 10⁶ | 2.1 × 10⁶ |
| pmb0542 | Cyanidin 3-O-malonylhexoside | 2.4 × 10⁶ | 0.9 × 10⁵ |
| pmb0545 | Rosinidin O-hexoside | 1.1 × 10⁷ | 1.0 × 10⁵ |
| pmb0550 | Cyanidin 3-O-glucoside (Kuromanin) | 9.9 × 10⁷ | 3.0 × 10⁵ |
| pmb2957 | Cyanidin O-syringic acid | 1.5 × 10⁷ | 0.9 × 10⁵ |
| pem0430 | Procyanidin A1 | 6.5 × 10³ | 8.7 × 10³ |
| pem0433 | Procyanidin A2 | 4.5 × 10⁴ | 4.8 × 10⁴ |
| pem0434 | Procyanidin B2 | 4.7 × 10⁶ | 3.1 × 10⁶ |
| pem0436 | Procyanidin B3 | 2.8 × 10⁶ | 3.5 × 10⁶ |
| pem1398 | Delphinidin 3-O-glucoside (Mirtillin) | 5.8 × 10⁷ | 1.1 × 10⁶ |
| pem1773 | Cyanidin 3-O-rutinoside (Keracyanin) | 4.7 × 10⁷ | 6.3 × 10⁵ |
| pem1777 | Cyanidin 3,5-O-diglucoside (Cyanin) | 3.9 × 10⁶ | 1.7 × 10⁶ |
| pme0027 | Cyanidin 3-O-galactoside | 9.1 × 10⁷ | 2.7 × 10⁵ |
| pme0116 | Delphin chloride | 1.0 × 10⁶ | 0.9 × 10⁵ |
| pme0203 | Peonidin 3-O-glucoside chloride | 2.7 × 10⁸ | 9.8 × 10⁵ |
| pme0614 | Peonidin 3, 5-diglucoside chloride | 2.5 × 10⁶ | 0.9 × 10⁵ |
| pme0616 | Malvidin 3-acetyl-5-diglucoside | 9.0 × 10³ | 2.6 × 10⁴ |

Metabolites were extracted from basal parts of Pr (plants with red spotted flowers) or Pw (without spots) petals at stage IV, relative contents were presented by respective peak areas. For those undetected metabolites, 0.9 × 10¹ was assigned as their peak area values

| Table 2 | Summary of the sequencing data and alignment to the *G. arboreum* reference genome for the four DNA libraries |
| Sample name | Clean reads | Clean reads (%) | GC content (%) | Q20 (%) | Q30 (%) | Total alignment ratio (%) | Sequencing depth (×) |
|------------|-------------|-----------------|---------------|---------|---------|------------------------|-------------------|
| SxyR | 70,893,040 | 99.96 | 35.83 | 94 | 89 | 96.36 | 11.51 |
| SxyW | 68,122,431 | 99.96 | 35.76 | 94 | 90 | 95.89 | 10.76 |
| Pr-bulk | 135,892,656 | 99.96 | 35.58 | 93 | 89 | 96.28 | 22.18 |
| Pw-bulk | 133,225,405 | 99.97 | 35.72 | 95 | 90 | 96.15 | 21.34 |

The four DNA libraries constitute the two parents that have red-spotted petals (SxyR, Shixiya-R) or not (SxyW, Shixiya-W), and the two mixed F2 progenies that have such opposite phenotypes wherein Pr-bulk and Pw-bulk are for descendants with or without red spots, respectively. Q20 and Q30 are quality parameters of sequencing, which refer to the percentages of nucleotides having high confidence (99.0% for Q20 and 99.9% for Q30) to be correct.
annotated genes (Tables 3 and S3). Meanwhile, another mapping approach, designated as the SNP-index association algorithm, was also utilized, which returned a 12.04 Mb-12.91 Mb location under the threshold of 0.78 on chromosome 7 (Fig. 2b). The narrowed extent (0.87 Mb) contains only 39 candidate genes (Tables 3 and S3), and it resided within the 3.27 Mb range that was obtained from previous ED mapping approach. Therefore, the quantitative locus was tentatively assigned to the 0.87 Mb region on chromosome 7 (Table 3), based on the attempts described above.

**Transcriptomic sequencing to facilitate candidate assignment**

Although the mapping interval was narrowed to a 0.87 Mb range on chromosome 7 by combining two mapping approaches, candidate assignment from the involved 39 annotated genes was still too vague to fulfill. Meanwhile, more evidences were needed to exclude the peripheral genes that may highly-linked or coinherited with these 39 genes. Accordingly, petals from three stages wherein the spots develop (stages II, III and IV, as illustrated in Fig. 1) were correspondingly collected from the segregated progenies that with or without red spots. A total of 975,976,429 clean reads were obtained from the RNA-seq output, with 40,143,724 − 96,530,240 clean reads per sample. Afterwards, the differentially expressed genes (DEGs) from each of the three stages were analyzed, which generated a total of 227, 165 and 105 DEGs (Figs. 3a) from stages II (Table S4), III (Table S5) and IV (Table S6), respectively. In order to verify the accuracy of the RNA-seq output, twelve genes were randomly selected to perform qRT-PCR (Fig. S2) using primers listed in Table S1. The results suggested most of them conformed with the RNA-seq data.

We then compared the 485 DEGs with the 39 annotated genes that resided within the 0.87 Mb interval. Only two genes were retained in the intersection, which were annotated as a hypothetical protein (Gar07G09710) and an uncharacterized protein (Gar07G09750, Table 4), respectively. These two *G. arboerum* genes are less likely to be the candidate for the accumulation of anthocyanin metabolites, for none of their orthologs in model species (like *Arabidopsis*) were implied to be associated with the metabolism of flavonoids. Hence, the 485 DEGs were further compared with the 163 genes within the 3.27 Mb interval that obtained

---

**Table 3** Information regarding the QTL regions by the two statistical algorithms

| Association analysis methods | Chromosome ID | Start (bp) | End (bp) | Size (Mb) | Gene number |
|-----------------------------|--------------|------------|----------|-----------|-------------|
| Euclidean distance          | Chr07        | 10,783,922 | 14,055,900 | 3.27      | 163         |
| SNP_index                   | Chr07        | 12,043,803 | 12,914,627 | 0.87      | 39          |
| Combination of both         | Chr07        | 12,043,803 | 12,914,627 | 0.87      | 39          |
from the ED mapping methodology, which generated six more genes. For these eight DEGs that located within the 3.27 Mb QTL range (Table 4), two were postulated to be the candidate for the accumulation of anthocyanins (and further, the appearance of red spots on petals) based on the sequence similarities between the \textit{G. arboerum} genes and its \textit{Arabidopsis} orthologs. Specifically, \textit{Gar07G09390} was postulated to encode a MYB transcription factor, its best hit in \textit{Arabidopsis}, \textit{AT1G66370}, was involved in regulating anthocyanin biosynthetic pathway (Gonzalez et al. 2008). Another gene (\textit{Gar07G08900}) that encoded a glutathione S-transferase (GST) was also considered to be a probable candidate for its homolog in \textit{Arabidopsis} (AtTT19) was involved in the accumulation of proanthocyanidins in the

Table 4 The eight DEGs that resided within the QTL interval

| Gene ID        | log2FoldChange (Pr vs. Pw) | Stage II | Stage III | Stage IV | Gene annotation                                                                 |
|---------------|-----------------------------|----------|-----------|----------|--------------------------------------------------------------------------------|
| Gar07G08900   | 3.99                        | 6.48     | 3.06      |          | PREDICTED: glutathione S-transferase F12 [\textit{Gossypium arboerum}]          |
| Gar07G08980   | 3.95                        |          | 3.95      |          | PREDICTED: protein DETOXIFICATION 8-like [\textit{Gossypium arboerum}]          |
| Gar07G09260   | 1.52                        |          |           |          | Putative L-type lectin-domain containing receptor kinase S.5-like protein [\textit{Gossypium arboerum}] |
| Gar07G09390   | 2.70                        |          |           |          | PREDICTED: transcription factor MYB90-like [\textit{Gossypium arboerum}]        |
| Gar07G09710   | −3.46                       |          | 2.66      |          | Hypothetical protein GOBAR_AA02861 [\textit{Gossypium barbadense}]              |
| Gar07G09750   | −1.27                       |          |           |          | PREDICTED: uncharacterized protein LOC107955911 [\textit{Gossypium hirsutum}]    |
| Gar07G09860   | −3.46                       |          |           |          | Hypothetical protein GOBAR_DD04258 [\textit{Gossypium barbadense}]              |
| Gar07G09900   | 1.30                        |          |           |          | PREDICTED: 4-hydroxyphenylpyruvate dioxygenase [\textit{Gossypium arboerum}]    |
seed coat (Kitamura et al. 2010). Notably, Gar07G09390 was differentially expressed (Fig. 3a and Table S5) in stage III of flower development, whereas the differential expression for Gar07G08900 was universally observed in all three stages (Fig. 3a and Tables S4-S6). Noteworthy, both of these two candidates were located outside the 0.87 Mb interval that were co-mapped by the two mapping methodologies (Fig. 3a).

**Functional validation of Gar07G08900 through virus-induced gene silencing (VIGS)**

The GST-coding gene (Gar07G08900) was differentially expressed through all three flower developmental stages. Comparatively, the MYB-coding gene (Gar07G09390) exhibited tinier expression discrepancy (Tables 4, S4 to S6), which was experimentally confirmed by qRT-PCR (Fig. 3b). Therefore, the GST gene (Gar07G08900) was subjected to further functional validation via VIGS. The Agrobacterium strain GV3101 that harbors different VIGS vectors (i.e., pTRV2:Gar07G08900 and pTRV2:00) was injected into the flower buds around stage II when the flowers were about to be colored (Fig. 1). The relative expression of the target gene, Gar07G08900, in the petals after inoculation was monitored, which suggested the expression levels were intact when injected pTRV2:00 as control, while the expression was suppressed by pTRV2:Gar07G08900 (Fig. 3c). Consequently, we could discern the scattered red spot from VIGS petals (Fig. 3d), which indicated the inclusion of Gar07G08900 in the red-spot-petal phenotype, probably by transporting the anthocyanin metabolites in *G. arboreum* flowers.

**Discussion**

The red-spotted *G. arboreum* flowers are related to the content variation of anthocyanins

Anthocyanins are universally responsible for the pigmentation of land plants (Shang et al. 2011; Zhang et al. 2014). Overall, different partitions from the six main anthocyanin aglycon derivatives (i.e., decorations on pelargonidin, cyanidin, delphinidin, peonidin, petunidin and malvidin) may be detected from respective species (Saigo et al. 2020). For instance, the cyanidin decorates constituted the largest proportion within the Malvaceae species (Saigo et al. 2020), which was consistent with our metabolic output (Fig. 1k). Meanwhile, our metabolic constitution was similar with the spots within the yellow pansy flowers (Li et al. 2014a) and the *G. hirsutum* lines (Abid et al. 2022). However, we have only included 17 anthocyanins, which may not enough to conclude the full spectrum of anthocyanins in *G. arboreum* petals, but is representative to unveil the candidate genes responsible for the red-spotted-petal phenotype.

**Candidate gene identification under the scope of multi-omics approaches**

Preliminary mapping of candidate regions can be achieved quickly using BSA seq (Sheng et al. 2017; Han et al. 2018), through the combination of ED-based and SNP-index-based association analyses (Geng et al. 2016; Zhang et al. 2022). Next, a large segregated population and high-density markers were needed to further fine-map candidate genes (She et al. 2018; Liu et al. 2019). In the matter of current project, a large candidate range spans 3.27 Mb was obtained using the ED methodology (Fig. 2a), wherein contained 163 annotated genes. Subsequently, a small intersection (0.87 Mb) was generated by the two association methods (Fig. 2), and only 39 candidate genes were resided (Table S3). Following this methodology, an enlarged mapping population may further narrow down the interval and therefore shrink the number of candidates. Indeed, in a recent *G. barbadense* × *G. hirsutum* report (Abid et al. 2022), 5660 F2 plants were used to narrow down the candidate region responsible for the spotted flower to a ~298 kb genomic interval. Subsequently, the authors sifted the resided eight genes based on annotation information, and finally identified a responsible one that encodes a MYB transcription factor (Abid et al. 2022). In our case, we may similarly sift the involving candidate genes by using a larger population, and yet alternatively, we chose to intersect the DEGs during the petal spot formation with the candidate range, which accordingly retained 8 candidates out of 163 genes. At last, we successfully identified a GST-encoding gene as the candidate, which was unexpectedly located outside the 0.87 Mb interval that sanctioned by both mapping algorithms (Fig. 3a), although it was suggested that combination of ED-based and SNP-index-based association analyses could improve prediction accuracy of candidate regions (Geng et al. 2016). In fact, all of the three probable candidate genes (Gar07G08900, Gar07G09390 and Gar07G09900, which will be discussed below) are resided outside the 0.87 Mb range, which indicated the difficulty in fine mapping candidate genes by forward genetics approaches only, and in turn implied the effectiveness of our strategy.

Within the two probable candidates that sifted by annotation information (Fig. 3a), we chose to validate the GST (Gar07G08900) in the first place, for it was the only one that exhibited differential expression across all the three stages for spot formation based on transcriptomic output (Fig. 3a and Table 4), which was then validated by qRT-PCR that it showed a more enormous expression differentiation between the Pr and Pw petals alongside the flower development than that of the MYB (Fig. 3b). However, it is far from

 Springer
satisfaction to assure the GST is the only candidate for this trait, since the cotton flowers only presented scattered spots after the GST was repressed by VIGS (Fig. 3c–d) while the two parents and their descendants showed a presenta-to-devoid colorations (Fig. 1). This output (Fig. 3d), which seems dissuadable and was affirmed by numerous VIGSs, may be partly accused that the anthocyanins have already been placed at the timepoint when VIGS was applied. However, it was probably the most suitable timepoint for VIGS, for an earlier manipulation would cause a significantly higher rate of bud abortion (by our practical experience, data not shown), and a later injection may cause a less discernable color fade. Another plausible explanation would be the existence of more candidates, like the MYB transcription factors, that acted simultaneously with GST (Ageorges et al. 2006; Luo et al. 2018; Jiang et al. 2019; Cui et al. 2021a, b). Coincidently, another gene (Gar07G09390) aside from the GST (Gar07G08900) that was retained within the intersection between the transcriptome and genetic mapping (Fig. 3a) was also considered to be a highly probable candidate. It encodes a MYB transcription factor and expressed higher at stage III of the spotted petals than non-spotted ones (Fig. 3a–b). Meanwhile, its best hit in Arabidopsis, AT1G66370, was involved in regulation of anthocyanin biosynthetic pathway and affected the expression of enzymes involved in later steps of anthocyanin biosynthesis (Gonzalez et al. 2008). Moreover, its orthologous gene in G. barbadense (Gbar_A07G008330) could restore petal spot formation in G. hirsutum flowers (Abid et al. 2022). Given to this, the functional connection of this MYB-encoding gene with petal spot phenotype, and its possible regulatory potential onto Gar07G08900 or other genes, should be subjected to further evaluation.

All in all, more evidences are required to prove or exclude the connections between any of the eight genes (Table 4) and the spotted petals. For instance, Gar07G09900 was annotated to express a 4-hydroxyphenylpyruvate dioxygenase (HPPD), the enzymatic product of which is homogentisic acid (HGA, Martin et al. 2017). HGA is widely detected in ornamental flowers, which contributed the highest phenolic compound in China rose (Li et al. 2014b). It was considered as one of the potential precursors of melanin that constitutes the major pigment of sesame seeds (Dossou et al. 2022). Meanwhile, both the HGA compound and the HPPD gene are involved in the vitamin E biosynthetic pathway (Martin et al. 2017), wherein the involved gene expressions and the metabolite contents are related to the fruit color of tomato (Vela-Hinojosa et al. 2019). Collectively, the inclusion of HGA and maybe other vitamin E pathway metabolites between the spotted and non-spotted petals could be subjected to metabolic profiling, which may provide a preliminary clue onto the involvement of HPPD against the phenotype of interest.

Current uncertainty upon candidate gene assignment was partially because our efforts started at a macroscale point (i.e., presence versus absence of red spots based on visual discrimination), wherein the red spot or the pigmentation which results in coloration is not evenly distributed among the petal cells. From what we have discerned, the pigmentation only occurs at the adaxial rather than the abaxial side of petals. More specifically, it has been reported that the mature petals typically compose of an abaxial and an adaxial epidermal layer and varied layer of mesophyll cells, these cell layers have different origin and diversification (Cavallini-Speisser et al. 2021). Here comes the question: on what layer of petals do pigmentation occur, and how does it expand when petals initiate and develop? Some state-of-the-art technologies like the single-cell transcriptome or metabolome may help resolving such an issue in the future.

**Further research and utilization of the red-spot-petal phenotype in cotton**

The colored spot is widespread in wild cotton species, e.g., G. arboreum, G. darwinii, G. barbadense, and landraces of G. hirsutum (Zhou 2011; Zhang et al. 2016; Abid et al. 2022). However, it is hardly observed in most modern commercial G. hirsutum cultivars, possibly because of its negative relation to major economic traits (Zhang et al. 2016). It has been reported that most of the collections within the National Plant Germplasm System of America harbors red spotted petals (Erpelding 2022), which is consistent with our observations onto the G. arboreum germplasm accessions in China. It is noteworthy that contrary to our observations across the G.arboreum progenies, the F1 lines of G. hirsutum×G. barbadense had an intermediate color shade and size compared with the spotted parent (Abid et al. 2022), suggesting the petal spot is a semidominant trait, which was similarly observed for that of G. hirsutum×G. darwinii descendants (Zhang et al. 2016). It seems the extra genetic factors may be ascribed to G. hirsutum based on the above two cases and the facts that spotted petals are common in both G. darwinii and G. barbadense species while sparse in G. hirsutum cultivars (Zhang et al. 2016; Abid et al. 2022). Accordingly, we crossed G. hirsutum spotted petal lines with non-spotted ones, and all of the F1 and F2 lines were more like the current G. arboreum cases that no intermediate petal spots were appeared (data not published). Although the existence of red petal spots seems to be regulated by a single dominant locus in current study and previous report (Erpelding 2022), it is not always that way for petal color phenotypes. For instance, the flower corolla color of G. arboreum accessions supported a two-gene model with a 9:3:3:1 ratio (Erpelding 2021). In view of this, the regulatory network for this particular phenotype in various cotton subspecies is worth investigating.
It has been suggested that the colored spot could enhance the attractiveness of cotton flowers for insect pollinators, resulting in improved boll-setting in *G. barbadense* and *G. hirsutum* (Jones 1996; Abid et al. 2022). From this point of view, it makes sense to specifically introduce the colored petal spot into the main upland cotton cultivars. To achieve this, pivotal candidates should firstly be identified, and the regulatory network is about to be unveiled. Under this scope, we have pioneered in presenting the current study. Next, markers with tight linkage with these key genes can be utilized to fulfill the marker assisted selection during the cross. In case the introgression of spotted petal can’t bring any positive yield or quality gain to the hybrid, the petal spots itself could be used as an indicator to tell the authenticity and purity of hybrids if a pure line with red spots was one of the parents (invention patent CN 102057862 A).

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04191-z.

**Author Contribution statement** JZ, XZ and SZ conceived and designed the experiments. SZ, XC, HW, LT and XL performed the experiments. JC and TJ analyzed the data. JC and SZ wrote the paper. All authors read and approved the final version of the paper.

**Funding** This work was supported by HAAFS Science and Technology Innovation Special Project (2022KJCXXZ-MHS-1) and Youth Fund of Hebei Natural Science Foundation (C2021301039).

**Data availability** Genomic resequencing data were deposited into Sequence read archive (SRA) database under the accession numbers of SAMN28185984 to SAMN28185989, and the transcriptome data were under the accession numbers of SAMN28185984 to SAMN28185989.

**Declarations**

**Conflict of interest** The authors declare no conflict of interest.

**References**

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174–178. https://doi.org/10.1038/nbt.2095

Abid MA, Wei Y, Meng Z, Wang Y, Ye Y, Wang Y et al (2022) Increasing floral visitation and hybrid seed production mediated by beauty mark in *Gossypium hirsutum*. Plant Biotechnol J 20:1274–1284. https://doi.org/10.1111/pbi.1380

Ageorges A, Fernandez L, Viallet S, Merdinoglu D, Terrier N, Romieu C (2006) Four specific isogenes of the anthocyanin metabolic pathway are systematically co-expressed with the red colour of grape berries. Plant Sci 170:372–383. https://doi.org/10.1016/j.plantsci.2005.09.007

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R et al (1998) Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. Plant Cell 10:1135–1149. https://doi.org/10.1105/tpc.10.7.1135

Cai X, Lin L, Wang X, Xu C, Wang Q (2018) Higher anthocyanin accumulation associated with higher transcription levels of anthocyanin biosynthesis genes in spinach. Genome 61:487–496. https://doi.org/10.1139/gen-2017-0261

Cavallini-Speisser Q, Morel P, Monniaux M (2021) Petal cellular identities. Front Plant Sci 12:745507. https://doi.org/10.3389/fpls.2021.745507

Chen G, He W, Guo X, Pan J (2021) Genome-wide identification, classification and expression analysis of the MYB transcription factor family in petunia. Int J Mol Sci 22:4838. https://doi.org/10.3390/ijms22094838

Chen J, Hu X, Shi T, Yin H, Sun D, Hao Y et al (2020) Metabotype-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. Plant Biotechnol J 18:1722–1735. https://doi.org/10.1111/pbi.13335

Chen W, Gong L, Guo Z, Wang Z, Zhang H, Liu X et al (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. Mol Plant 6:1769–1780. https://doi.org/10.1093/mp/ssu080

Cui D, Zhao S, Xu H, Allan AC, Zhang X, Fan L et al (2021a) The interaction of MYB, bHLH and WD40 transcription factors in red pear (*Pyrus pyrifolia*) peel. Plant Mol Biol 106:407–417. https://doi.org/10.1007/s11103-021-01160-w

Cui Y, Fan J, Lu C, Ren J, Qi F, Huang H et al (2021b) ScGST3 and multiple R2R3-MYB transcription factors function in anthocyanin accumulation in *Senecio cruentus*. Plant Sci 313:111094. https://doi.org/10.1016/j.plantsci.2021.111094

Dossou S, Luo Z, Wang Z, Zhou R, Zhang Y et al (2022) The dark pigment in the sesame (*Sesamum indicum*) seed coat: isolation, characterization, and its potential precursors. Front Nutr 9:858673. https://doi.org/10.3389/fnut.2022.858673

Duan HR, Wang LR, Cui GX, Zhou XH, Duan XR, Yang HS (2020) Identification of the regulatory networks and hub genes controlling alfalfa floral pigmentation variation using RNA-sequencing analysis. BMC Plant Biol 20:110. https://doi.org/10.1186/s12870-020-2322-9

Erpelding JE (2021) Genetic characterization of the red coloured corolla phenotype for *Gossypium arboreum* accession PI 529731. Plant Breed 140:142–149. https://doi.org/10.1111/pbr.12884

Erpelding JE (2022) Genetic analysis of the Asiatic cotton (*Gossypium arboreum*) petal spot phenotypes. Plant Breed 141:71–76. https://doi.org/10.1111/pbr.12984

Geng X, Jiang C, Yang J, Wang L, Wu X, Wei W (2016) Rapid identification of candidate genes for seed weight using the SLAF-seq method in *Brassica napus*. PLoS ONE 211:e0147580. https://doi.org/10.1371/journal.pone.0147580

Gonzalez A, Zhao M, Leavitt J, Lloyd A (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J 53:814–827. https://doi.org/10.1111/j.1365-313X.2007.03373.x

Han Y, Zhao F, Gao S, Wang X, Wei A, Chen Z et al (2018) Fine mapping of a male sterility gene *ms-3* in a novel cucumber (*Cucumis sativus*) mutant. Theor Appl Genet 131:449–460. https://doi.org/10.1007/s00122-017-0261-6

Hou XG, Guo Q, Wei WQ, Guo LL, Guo DL, Zhang L (2018) Screen- ing of genes related to early and late flowering in tree peony based on bulked segregant RNA sequencing and verification by quantitative real-time PCR. Molecules 23:689. https://doi.org/10.3390/molecules23030689

Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski J et al (2020) Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. Nat Genet 52:516–524. https://doi.org/10.1038/s41588-020-0607-4
Sheng Y, Wang Y, Jiao S, Jin Y, Ji P, Luan F (2017) Mapping and preliminary analysis of aborted microspores (AMS) as the candidate gene underlying the male sterility (MS-5) mutant in melon (Cucumis melo L.). Front Plant Sci 8:902. https://doi.org/10.3389/fpls.2017.00902

Sun Y, Li H, Huang JR (2012) Arabidopsis TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. Mol Plant 5:387–400. https://doi.org/10.1093/mp/ssr110

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J 54:733–749. https://doi.org/10.1111/j.1365-313X.2008.03447.x

Vela-Hinojosa C, Escalona-Buendía HB, Mendoza-Espinoza JA, Villa-Hernández JM, Pérez-Flores LJ (2019) Antioxidant balance and regulation in tomato genotypes of different color. J Am Soc Hortic Sci 144:45–54. https://doi.org/10.21273/JASHS04525-18

Wang G, Chen B, Du H, Zhang F, Zhang H, Wang Y et al (2018) Genetic mapping of anthocyanin accumulation-related genes in pepper fruits using a combination of SLAF-seq and BSA. PLoS ONE 13:e0204690. https://doi.org/10.1371/journal.pone.0204690

Wang Y, Xiao L, Guo S, An F, Du D, Li M (2016) Fine mapping and whole-genome resequencing identify the seed coat color gene in Brassica rapa. PLoS ONE 11:e0166464. https://doi.org/10.1371/journal.pone.0166464

Wang ZR, Cui YY, Vainstein A, Chen SW, Ma HQ (2017) Regulation of Fig (Ficus carica L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. Front Plant Sci 8:1990. https://doi.org/10.3389/fpls.2017.01990

Weiss MR (1995) Floral color change: a widespread functional convergence. Am J Bot 82:167–185. https://doi.org/10.1002/j.1537-2197.1995.tb11486.x

Yang X, Xu Y, Yang F, Magwanga RO, Cai X, Wang X et al (2019) Genome-wide identification of OSCA gene family and their potential function in the regulation of dehydration and salt stress in Gossypium hirsutum. J Cotton Res 2:87–104. https://doi.org/10.1186/s42397-019-0028-z

Zhang K, Yuan M, Xia H, He L, Ma J, Wang M et al (2022) BSA-seq and genetic mapping reveals AhRt2 as a candidate gene responsible for red testa of peanut. Theor Appl Genet 135:1529–1540. https://doi.org/10.1007/s00122-022-04051-w

Zhang S, Lan Q, Gao X, Yang B, Cai C, Zhang T et al (2016) Mapping of genes for flower-related traits and QTLs for flowering time in an interspecific population of Gossypium hirsutum×G. darwini. J Genet 95:197–201. https://doi.org/10.1007/s12041-016-0617-2

Zhang X, Li R, Chen L, Niu S, Chen L, Gao J et al (2017) Fine-mapping and candidate gene analysis of the Brassica juncea white-flowered mutant Bjpc2 using the whole-genome resequencing. Mol Genet Genomics 293:359–370. https://doi.org/10.1007/s00438-017-1390-5

Zhang Y, Butelli E, Martin C (2014) Engineering anthocyanin biosynthesis in plants. Curr Opin Plant Biol 19:81–90. https://doi.org/10.1016/j.pbi.2014.05.011

Zhang YL, Fang ZZ, Ye XF, Pan SL (2018) Identification of candidate genes involved in anthocyanin accumulation in the peel of jaboticaba (Myrciaria cauliflora) fruits by transcriptomic analysis. Gene 676:202–213. https://doi.org/10.1016/j.gene.2018.07.039

Zhou ZL (2011) Study on genetic diversity of G. arboreum L. germplasm resources conserved in China. Master Dissertation of Chinese Academy of Agricultural Sciences, 20 (in Chinese). https://doi.org/10.7666/d.Y1932322

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Sujun Zhang1 · Jie Chen3 · Tao Jiang2 · Xiao Cai1 · Haitao Wang1 · Cunjing Liu1 · Liyuan Tang1 · Xinghe Li1 · Xiangyun Zhang1 · Jianhong Zhang1

Sujun Zhang
sujun1987_good@126.com

Jie Chen
lqlcj@126.com

Tao Jiang
jttaojiang@163.com

Xiao Cai
249692773@qq.com

Haitao Wang
495374062@qq.com

Cunjing Liu
245253837@qq.com

Liyuan Tang
451614258@qq.com

Xinghe Li
1240296052@qq.com

1 Institute of Cotton, Hebei Academy of Agriculture and Forestry Sciences/Key Laboratory of Biology and Genetic Improvement of Cotton in Huanghuaihai Semiarid Area, Ministry of Agriculture and Rural Affairs, Shijiazhuang 050051, Hebei, China

2 Institute of Cash Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050051, Hebei, China

3 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China