

ARTICLE

Received 8 Mar 2015 | Accepted 11 Dec 2015 | Published 22 Jan 2016

DOI: 10.1038/ncomms10456

OPEN

Genetic basis for glandular trichome formation in cotton

Dan Ma^{1,*}, Yan Hu^{1,*}, Changqing Yang^{2,*}, Bingliang Liu¹, Lei Fang¹, Qun Wan¹, Wenhua Liang¹, Gaofu Mei¹, Lingjian Wang², Haiping Wang¹, Linyun Ding¹, Chenguang Dong¹, Mengqiao Pan¹, Jiedan Chen¹, Sen Wang¹, Shuqi Chen¹, Caiping Cai¹, Xiefei Zhu¹, Xueying Guan¹, Baoliang Zhou¹, Shuijin Zhu³, Jiawei Wang², Wangzhen Guo¹, Xiaoya Chen² & Tianzhen Zhang¹

Trichomes originate from epidermal cells and can be classified as either glandular or non-glandular. *Gossypium* species are characterized by the presence of small and darkly pigmented lysigenous glands that contain large amounts of gossypol. Here, using a dominant glandless mutant, we characterize *GoPGF*, which encodes a basic helix-loop-helix domain-containing transcription factor, that we propose is a positive regulator of gland formation. Silencing *GoPGF* leads to a completely glandless phenotype. A single nucleotide insertion in *GoPGF*, introducing a premature stop codon is found in the duplicate recessive glandless mutant (*gl₂gl₃*). The characterization of *GoPGF* helps to unravel the regulatory network of glandular structure biogenesis, and has implications for understanding the production of secondary metabolites in glands. It also provides a potential molecular basis to generate glandless seed and glanded cotton to not only supply fibre and oil but also provide a source of protein for human consumption.

¹State Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Hybrid R & D Engineering Center (the Ministry of Education), Nanjing Agricultural University, Nanjing 210095, China. ²National Key Laboratory of Plant Molecular Genetics, National Plant Gene Research Center, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China. ³Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Zhejiang 310029, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.Z. (email: cotton@njau.edu.cn) or to X.C. (email: xychen@sibs.ac.cn).

Trichomes are specialized structures that originate from epidermal cells, and can be classified into two main categories: non-glandular or glandular¹. The most remarkable feature of glandular trichomes is their unique capacity to synthesize, store and sometimes secrete a wide array of metabolites, including polysaccharides, organic acids, proteins, terpenoids (such as the antimalarial drug artemisinin), alkaloids and polyphenols^{1,2}. These compounds give plants a distinctive smell, the natural barrier to protect plants against herbivorous insects and pathogens, or have a significant commercial value as drugs, fragrances, food additives and insecticides^{1,3}. As such, glandular trichomes have been described as bio-factories for the production of high-value natural products such as volatile oils, resins, mucilages and gums. Despite extensive study of the development and fine structure, and physiological and molecular metabolisms with regard to chemicals synthesized, stored and secreted in glandular trichomes^{4–7}, almost nothing is known about the genetics underlying their development.

Cotton (*Gossypium* spp) is not only the leading natural fibre resource, but also the third largest field crop in terms of edible oilseed tonnage. In addition to 21% oil, cottonseed has a relatively high protein content (23%). For every kilogram of fibre produced, 1.65 kg of seed is also collected. According to this ratio, the global cottonseed production could potentially provide protein for half a billion people annually⁸. However, the *Gossypium* species are characterized by the presence of small darkly pigmented lysigenous glands containing gossypol deposits, which are toxic to humans and monogastric animals. Therefore, cottonseed cannot be used to produce edible proteins or oils directly. Gossypol is a yellowish phenolic compound that contributes to the defence of cotton against pests, diseases and abiotic stresses^{9,10}. Therefore, potentially the best approach would be to develop cotton that produces gossypol-free seed for human consumption and normal gossypol content in other tissues for protection from pests and pathogens by seed-specific genetic engineering of gland formation or gossypol biosynthesis. Ultra-low gossypol cottonseed lines have been developed using RNAi knockdown of δ -cadinene synthase gene(s) during seed development in *Gossypium hirsutum*⁸. The ultra-low gossypol trait is stable under field conditions and the foliage/floral organs contain wild-type levels of gossypol and related terpenoids, indicates it could be possible to enhance and expand the nutritional utility of the annual cottonseed output to fulfil the ever-increasing needs of humanity after the evaluation of its safety and nutritional efficacy⁹.

Understanding the molecular genetic basis of gossypol gland formation could provide additional methods to develop gossypol-free cotton seeds that could be produced efficiently and be adapted widely. The glands originate from a cluster of gland primodium cells, which differ from other cells in that they have a high-density cytoplasm and large nucleolus. The internal cells are then degraded and form a cavity known as a gland in the ground meristem¹¹. From a cross between a cultivated and a primitive race of *G. hirsutum*, McMichael^{12,13} recovered a recessive mutant that eliminates all glands on the aerial plant parts and seeds (Fig. 1). This offered the possibility of cultivating glandless cotton, which could potentially help to alleviate hunger and protein shortages¹⁴. Up to now, genetic research has revealed that cotton gland formation is determined by a combination of at least six independent loci, *gl*₁, *gl*₂, *gl*₃, *gl*₄, *gl*₅ and *gl*₆ (ref. 15). The glandless phenotype is controlled by two pairs of duplicate recessive genes (*gl*₂*gl*₃) on chromosome (chr.) A12 and D12, respectively^{15,16}. In addition, a single dominant glandless mutation in *G. barbadense* was discovered in Egypt following the irradiation of Giza 45 seeds with ³²P, which results in glandless plants and seeds (Figs 1 and 2a). Genetic analysis

revealed that it is a dominant allele at the *Gl*₂ locus that is epistatic to *Gl*₃, and the gene symbol *Gl*₂^g was proposed^{17,18}. A new strain of cotton that is homozygous for this new gene was released as Bahtim 110 in Egypt¹⁹ and Hai-1 in China¹⁸. These two mutants have been used to develop many glandless cultivars of both *G. hirsutum* and *G. barbadense* to produce little or no gossypol in seeds. For example, more than 20 glandless cultivars have been developed in China²⁰. However, the molecular genetic basis for the formation of pigment glandular trichomes, which are storage organs of gossypol in cotton, remains unknown. Here, through a map-based cloning approach, we identify *Gossypium* PIGMENT GLAND FORMATION GENE (*GoPGF*) as the likely causative gene for the phenotype of glandless mutant cotton. This helps to uncover the molecular basis for formation of glandular trichomes and secondary substances such as gossypol in cotton.

Results

Map-based cloning of the dominant glandless gene *Gl*₂^g. We previously anchored the dominant glandless gene *Gl*₂^g between two microsatellites (SSRs), NAU3778 and NAU2251b, with genetic distances of 9.27 and 0.96 cm, respectively in chr. 12A (ref. 21). On the basis of the genome sequence of *G. raimondii*²², *Gl*₂^g was delimited within a 1-Mb interval on scaffold D8. We mined and developed 316 SSRs to screen the polymorphism between Hai-1 and TM-1 within this region (Supplementary Data 1). We used an enlarged mapping population including 2,197 individuals (Supplementary Table 1) and seven polymorphic SSRs, narrowed the *Gl*₂^g locus to a 43-kb region flanked between w7954 and w5383 with genetic distances of 0.5 and 0.6 cM, respectively (Fig. 2b; Supplementary Data 1). Within this region, seven putative open reading frames (ORFs; Supplementary Table 2) are predicted based on our high-quality reference genome sequence of tetraploid cotton *G. hirsutum* acc. TM-1 (ref. 23). Quantitative RT-PCR (qPCR; Fig. 2c; Supplementary Fig. 1) revealed that only the expression level of ORF2 was altered, and was significantly lower in the dominant glandless mutant Hai-1 and the (Hai-1 × TM-1)_F₁ than in the glanded *G. hirsutum* acc. TM-1 and *G. barbadense* cv. Hai7124. Therefore, we considered ORF2 as a candidate gene for *Gl*₂^g.

We then isolated the full length genomic DNA of ORF2 from the glandless Hai-1 and its wild type, Giza 45, and found that the coding region is 1,428 bp in length with no intron and contains a predicted gene of unknown function. The protein is predicted to contain a conserved motif shared by bHLH and R2R3-MYB in the N-terminal motif (Pfam14215) from amino acid 20 to 202 and a helix-loop-helix DNA-binding domain (Pfam00010) from amino acid 293 to 350 at the C-terminal (<http://pfam.sanger.ac.uk/>), which is involved in DNA binding and protein oligomerization. The gene is phylogenetically closely related to bHLH transcription factor (Thecc1EG015640t1) in *Theobroma cacao* and belongs to bHLH transcription factor family. We named the gene as *Gossypium* pigment gland formation gene (*GoPGF*). Orthologous genes of *GoPGF* in *G. raimondii* and *G. arboreum* are Gorai.008G259000.1 and Cotton_A_01306, respectively. *GoPGF* protein was localized to the nucleus (Supplementary Fig. 2). Quantitative reverse transcriptase PCR (qRT-PCR) analysis revealed that the gene is expressed in a constitutive manner in most organs and tissues, including root, stem and leaf (Fig. 2c; Supplementary Fig. 3), consistent with the presence of glands. DNA sequence comparisons revealed that three single nucleotide polymorphisms (SNPs) exist between the dominant glandless Hai-1 and the glanded Giza 45 and other *G. barbadense* cultivars in *GbPGF* (Supplementary Fig. 4), but only result in one amino acid change from alanine to valine at residue 43 (Fig. 2d). On the basis of these SNPs, we found that the mutant allele (*GbPGFm*) from Hai-1 co-segregates with 1,624 glandless plants of three F₂s (Supplementary Data 1).

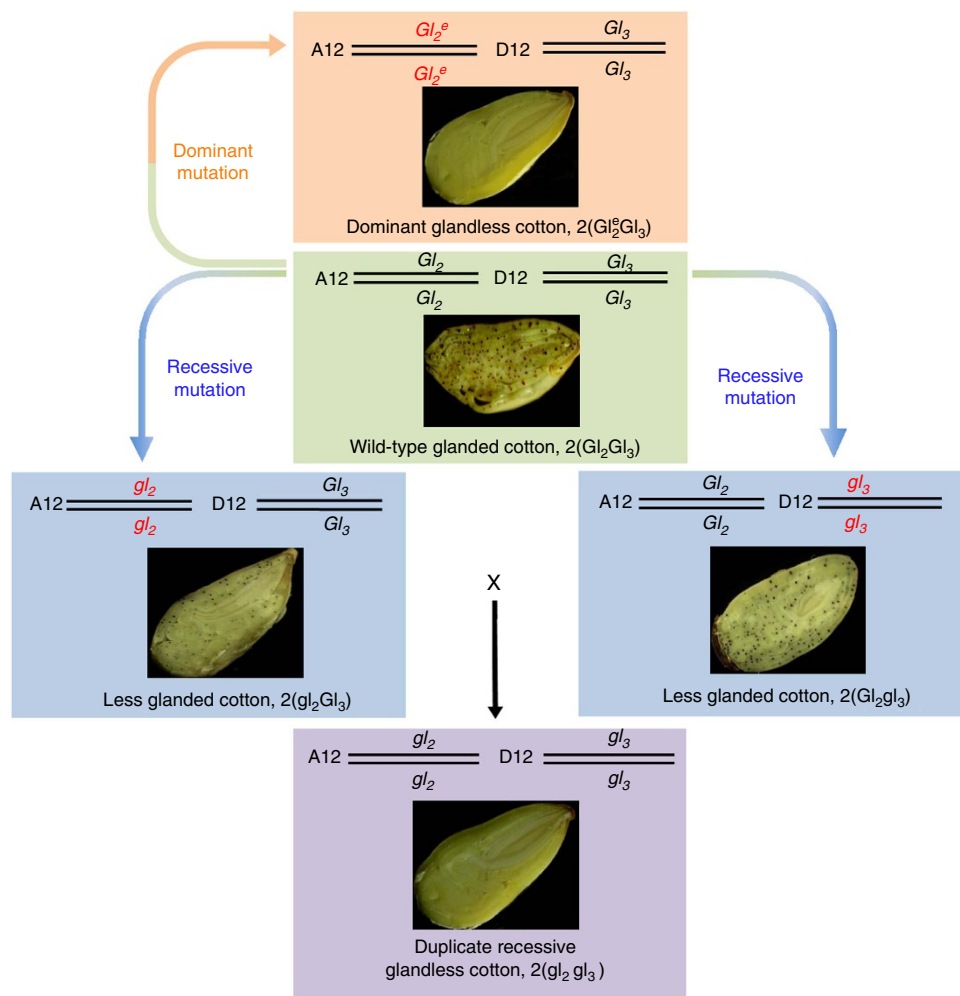


Figure 1 | Origins of the genotypes and phenotypes of glandless tetraploid cottons. The genotypes include a dominant mutation, $Gl_2^eGl_3$; two recessive mutations, gl_2Gl_3 and Gl_2gl_3 ; and a duplicated recessive cotton, gl_2gl_3 . Black points in each graph show the glands on the surface of seeds.

The difference in *GoPGF* mRNA levels between Hai-1 and TM-1 may possibly be determined by transcriptional regulation of the *GoPGF* gene. To investigate the differential expression of *GoPGF_A12* from TM-1 and Hai-1, we isolated 1.8-kb fragments upstream of the start codon of *GoPGF* from TM-1 and Hai-1. No difference in transient expression of GUS reporter constructs was detected after *Agrobacterium*-mediated transformation of tobacco leaves (Supplementary Fig. 5), indicating that their promoters likely do not confer the extra low expression of *PGF* in Hai-1. Whether the mutation reduces its expression in Bahtim 110 or Hai-1 or whether a causative mutation on a distant loci²⁴ exerts influence on *GbPGF* remains to be explored.

Nucleotide insertion results in two recessive mutant alleles. The whole plant glandless phenotype is controlled by a combination of two recessive mutant alleles, gl_2 and gl_3 (ref. 13) or one dominant gene Gl_2^e (refs 13,15; Fig. 1). So, if *PGF* is the candidate gene for Gl_2^e , we reasoned that *PGF_A12* and *PGF_D12* in the two recessive mutant alleles, gl_2 and gl_3 , should be non-functional. We isolated and sequenced *GoPGFs* from tetraploid cotton accessions and their extant progenitor species, *G. arboreum* (A_2) and *G. raimondii* (D_5). Sequence alignments show that there are homoeologous *PGF* gene pairs in the corresponding A subgenome (*GoPGF_A12*) and D subgenome (*GoPGF_D12*) of tetraploid cottons with seven SNPs between them

(Supplementary Fig. 4; Supplementary Table 3). We isolated the mutant *GoPGF* genes in the chr. A12 (*GhPGF_A12^m*) and D12 (*GhPGF_D12^m*), respectively, from the duplicate recessive glandless mutant, 2(gl_2gl_3), and the monomeric mutants 2(Gl_2gl_3) and 2(gl_2Gl_3) with a low number of glands (Fig. 1). Sequence alignments show that a single “T” nucleotide insertion occurs between 735 and 736 bp in the coding region of *GhPGF_A12^m* (gl_2) in chr. A12, and a single “A” nucleotide insertion occurs between 916 and 917 bp in the coding region of *GhPGF_D12^m* (gl_3) in chr.D12. This premature translation termination correlated with the production of fewer glands in monomeric mutants (Gl_2gl_3 and gl_2Gl_3) or completely glandless phenotype in the duplicate mutant (gl_2gl_3 ; Figs 1 and 2d; Supplementary Fig. 4). qPCR analysis revealed that *GhPGF_A12^m* (gl_2) and *GhPGF_D12^m* (gl_3) were expressed at very low levels compared to their corresponding homoeologous genes (Gl_3 and Gl_2) in the monomeric mutants 2(gl_2Gl_3) and 2(Gl_2gl_3), respectively, as well as in the duplicate recessive glandless mutant 2(gl_2gl_3) and in Hai-1 (Fig. 2e). These results strongly suggest that *GoPGF* is the gene that controls cotton pigment gland formation.

***GbPGF*-silencing results in a glandless phenotype.** To further assess its function, we cloned the 3'-end fragment (904–1,428 bp) of *GbPGF_A12* from Giza 45 and inserted it into pTRV2 for virus-induced gene silencing (VIGS)²⁵ to suppress the expression

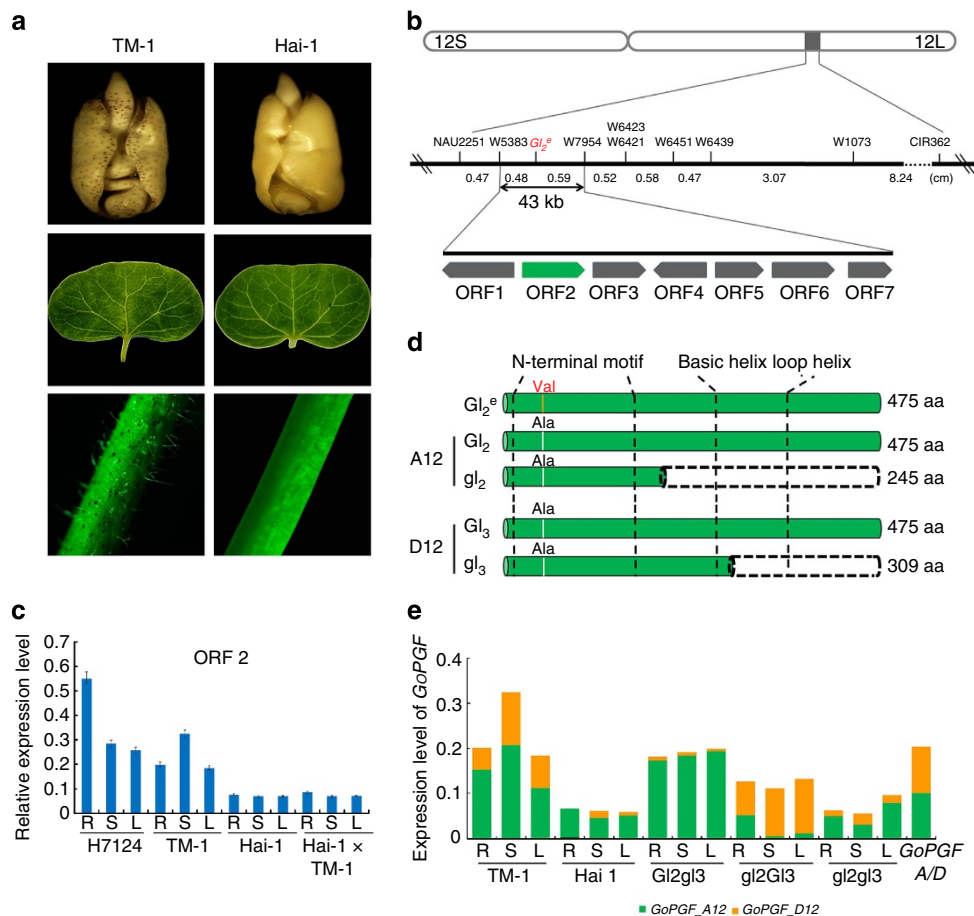


Figure 2 | Fine mapping of dominant glandless gene *Gf* and the expression pattern of *GoPGF*. (a) Gland traits of seeds, leaves and stems of TM-1 (glanded) and Hai-1 (glandless). (b) Fine mapping of the dominant glandless gene, *Gf*. Seven genes (ORF1-7) in the mapping region are indicated by boxes. (c) Real-time RT-PCR expression analysis of ORF2 in the root, stem and leaf of Hai7124, TM-1, Hai-1 and Hai-1 × TM-1. Error bars represent the s.d. of the mean values of three biological replicates. (d) Sequence diversity of PGF. An amino acid change from alanine to valine was observed in the dominant glandless gene (*Gf*). The premature translation termination resulted in the production of fewer glands (*Gl₂gl₃* and *gl₂Gl₃* mutant) or completely glandless (*gl₂gl₃* mutant) phenotypes. (e) Pyrosequencing analysis of the relative expression levels of *GoPGF_A12* and *GoPGF_D12* homoeologous alleles in the root, stem and leaf of TM-1, Hai-1, 2(*Gl₂gl₃*), 2(*gl₂Gl₃*) and 2(*gl₂gl₃*).

of endogenous *GbPGF* in two cultivated glanded tetraploid cottons. The *PGF*-silenced *G. barbadense* cv. Giza 45 (wild type of the Hai-1 mutant), Hai7124 and *G. hirsutum* acc.TM-1 plants all exhibited a glandless or low gland number phenotype in the newly emerging tissues 14 days post-agro-infiltration (Figs 3a and 4; Supplementary Fig. 6). We observe no or fewer visible glands in the new growing upper stems and leaves of the *PGF*-silenced plants, but the stems below the cotyledon node and cotyledons had many thickly dotted glands that had already formed before infiltration. The transcripts of *GbPGF* in the *PGF*-silenced leaves were significantly reduced compared to the untreated TM-1 and Hai7124, indicating that *GbPGF* was effectively silenced in VIGS plants (Figs 3b and 4a). In the wild-type cotton plants, large cavities of mature pigmented glands were present, however, no such gland cavities were observed in the mesophyll cells of *GbPGF*-silenced leaves in either Hai7124 or TM-1 (Fig. 3c). These data further suggest that *GbPGF* regulates the formation of glands that act as a storage organ for gossypol and other related sesquiterpenes in cotton. Consistently, by suppressing the *GoPGF* expression, gland cell differentiation was blocked, and hemigossypol and gossypol content was reduced by over 93% in the *GbPGF*-silenced leaves compared with the untreated TM-1 and Hai7124 leaves (Fig. 3d), suggesting that *PGF* is also involved in gossypol biosynthesis, directly or indirectly. Interestingly,

however, the non-glandular hair trichomes developed normally (Supplementary Fig. 7) in the *PGF*-silenced *G. hirsutum*, suggesting that development of hair or non-glandular and glandular trichomes may use different genetic machinery in cotton.

Gossypol biosynthesis is not linked to gland formation.

Through an antisense strategy, we developed transgenic cotton plants with seed-specific silencing of (+)- δ -cadinene-8-hydroxylase (*CYP706B1*), a P450 monooxygenase in the gossypol biosynthesis pathway²⁶. Gossypol levels were significantly reduced in the transgenic seeds, but lysigenous glands still formed as in non-transgenic plants (Supplementary Fig. 8). This result is similar to a previous report, which showed that silencing of (+)- δ -cadinene synthase gene (*CDN*)²⁷ that catalyses the first step in gossypol biosynthesis did not affect gland formation, suggesting the terpenoid aldehyde synthesis and gland formation are uncoupled.

Phylogenetic analysis shows a distinct 'glandular trichome formation' clade. Together with other bHLH members in the vascular plants species covered by glandular trichome such as *Nicotiana tabacum*, *Solanum lycopersicum* and *Artemisia annua* (Supplementary Fig. 9; Supplementary Data 2), *GoPGF* is

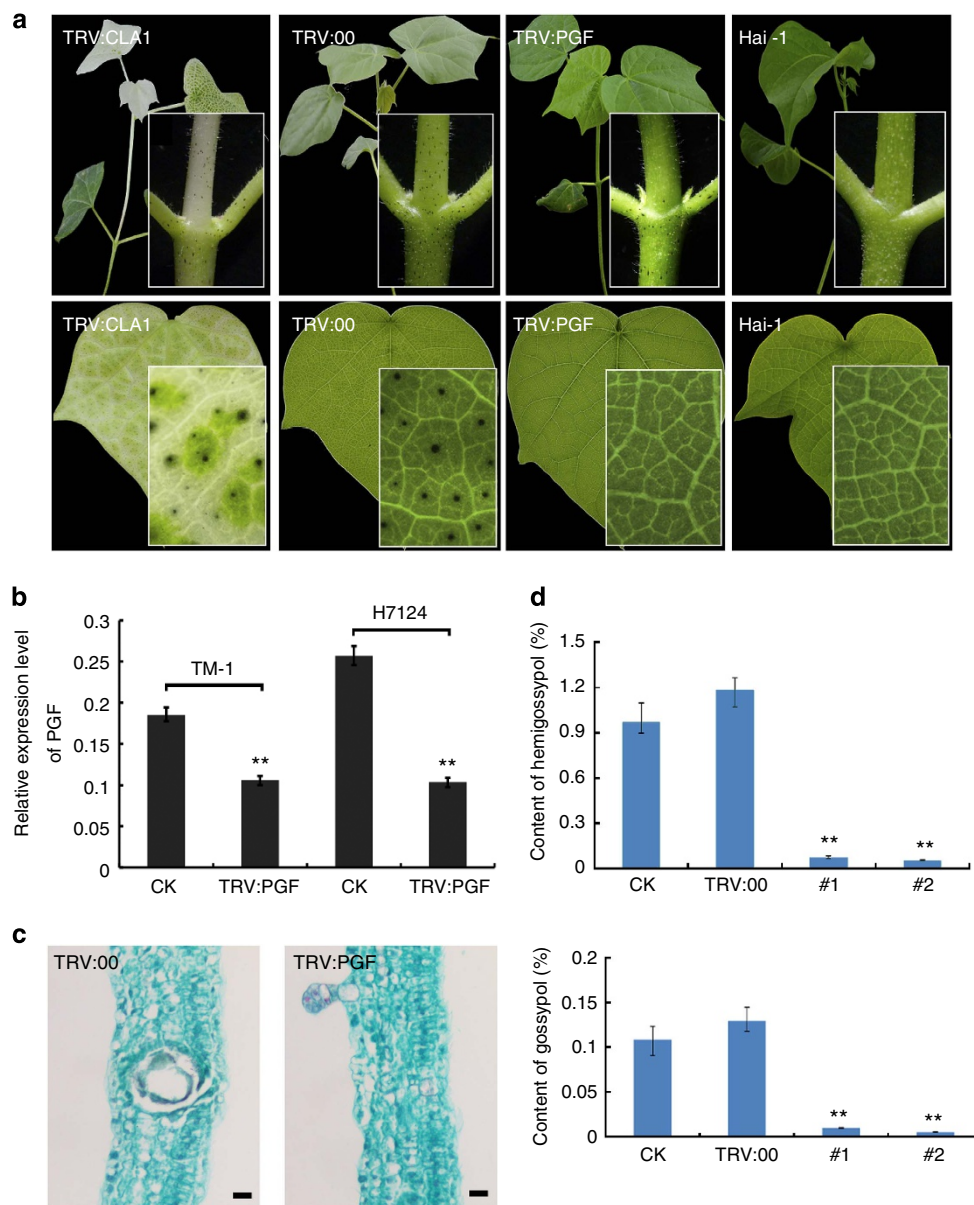


Figure 3 | Functional characterization of *GoPGF* by VIGS. (a) Phenotypes of TM-1 before and after *GoPGF* silencing by VIGS, showing the presence and the absence of the glands. **(b)** Transcript level of *GoPGF* in *GoPGF*-silenced leaves. **(c)** Cavity observed in the leaves of TM-1 but disappeared in the leaves emerged after VIGS. Scale bar, 10 μ m. **(d)** Hemigossypol and gossypol content in control (CK), empty vector (TRV:00) and in the *GoPGF*-silenced (#1 and #2) leaves of TM-1, determined by high-performance liquid chromatography. ** $P < 0.01$; Student's *t*-test, $n = 3$. Error bars are s.d. of three biological repeats.

classified in a distinct clade (clade II), which has 19 unknown function members. These bHLHs form a specific cluster involved in regulating glandular trichome formation, distinct from other bHLHs known to be involved in secondary metabolism such as *AtMYC2*, *AtMYC3* and *AtMYC4* in *Arabidopsis*²⁸ (Fig. 5a).

Differential gene expression in *GoPGF*-silenced plants. To gain insight into the regulatory networks that may underlie gland formation, we compared gene expression in leaves of *GoPGF*-silenced plants, the control TM-1 and the mutant Hai-1 by RNA sequencing (RNA-seq). We found that 3,582 genes were differentially expressed with 2,276 upregulated and 1,306 downregulated in the *GhPGF*-silenced leaves, including significantly reduced expression of 15 terpenoid synthase (*TPS*) genes, 18 *MYBs* and 31 *WRKYs* (Supplementary Data 3). The

upregulated genes include genes related to light reaction, cell wall degradation and protein synthesis. Strikingly, the downregulated genes include genes related to secondary metabolism and terpenoid biosynthesis²⁹, jasmonate (JA) signalling and genes in the *WRKY* and *MYB* transcription factor families (Supplementary Fig. 10). Subsequent qPCR confirmed decreased expression of some *TPS* and *WRKY* genes in *PGF*-silenced tissues (Supplementary Fig. 11). Yeast one-hybrid binding analysis suggests that *GoPGF* could specifically interact with the G-box motif (Supplementary Fig. 12), which is commonly found in *TPSs* and *WRKYs* promoter regions including *CDN-1* (ref. 29). Pathway analysis also revealed these differentially expressed genes included genes involved in secondary metabolism (Supplementary Table 4). We also found that expression of *GoPGF* was induced by JA treatment (Supplementary Fig. 13). We speculate that *GoPGF* protein could control the specification

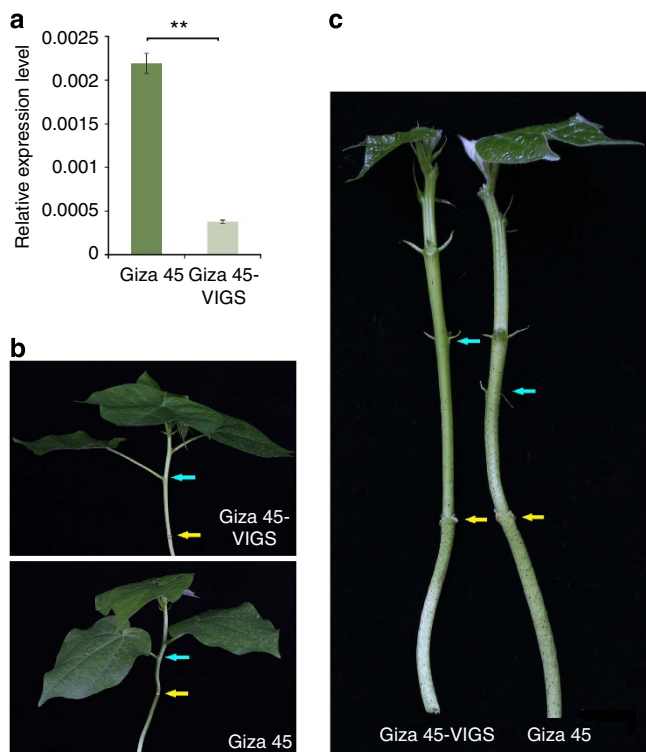


Figure 4 | Functional characterization of *GoPGF* by VIGS in *G. barbadense* cv. *Giza 45*. (a) Transcript level of *GoPGF* in normal *Giza 45* and corresponding *GoPGF*-silenced leaves. Error bars are s.d. of three biological repeats. ****** $P < 0.01$; Student's *t*-test, $n = 3$. (b,c) Phenotypes of *Giza 45* before and after *GoPGF* silencing by VIGS. Yellow and blue arrows indicate the cotyledonary node and the first vegetative branch, respectively.

and differentiation of gland cells, possibly through regulating the expression of *JAZ*, *WRKYs* or other genes. In addition, *PGF* may also regulate sesquiterpene biosynthesis via binding to the promoters of *TPSs* or *WRKYs* (Fig. 5b).

Discussion

We have provided evidence that *GoPGF*, encoding a bHLH transcription factor, is likely the causative gene for the glandless phenotype in cotton and appears to be a regulator of glandular trichome formation. It therefore likely represents the first gene to be successfully cloned in tetraploid cotton using a map-based cloning strategy. Its cloning helps to elucidate the molecular mechanism of the genetic control networks involved in secondary substances and formation of glandular trichome, which are storage organs in other plants such as *N. tabacum*³⁰, *S. lycopersicum*³¹, *A. annua*³² and so on. We identified a bHLH cluster that may have diverged from other bHLH genes and be involved in glandular trichome formation. Glandular trichomes are metabolic hotspots for biosynthesis, regulation and release of numerous volatile and non-volatile phytochemicals used by plants for interacting with the biotic environment. Natural products synthesized in trichomes are also widely adapted as flavorants, perfumes and pharmaceuticals. The distinct glandular trichome formation clade will provide us important candidates to control the biosynthesis of useful secondary compounds through genetic engineering.

Our results further clarified the close and complicated relationship between gland and gossypol. Our study showed three glands developed normally in *CYP706B1*-silenced transgenic cotton with reduced gossypol content, suggesting that gossypol was not required for gland morphogenesis, indicating they were uncoupled and controlled by different molecular machines. Nevertheless, gossypol content was markedly suppressed in the *GoPGF*-suppressed cotton without noticeable glands, suggesting that suppression of gland formation will feedback to gossypol

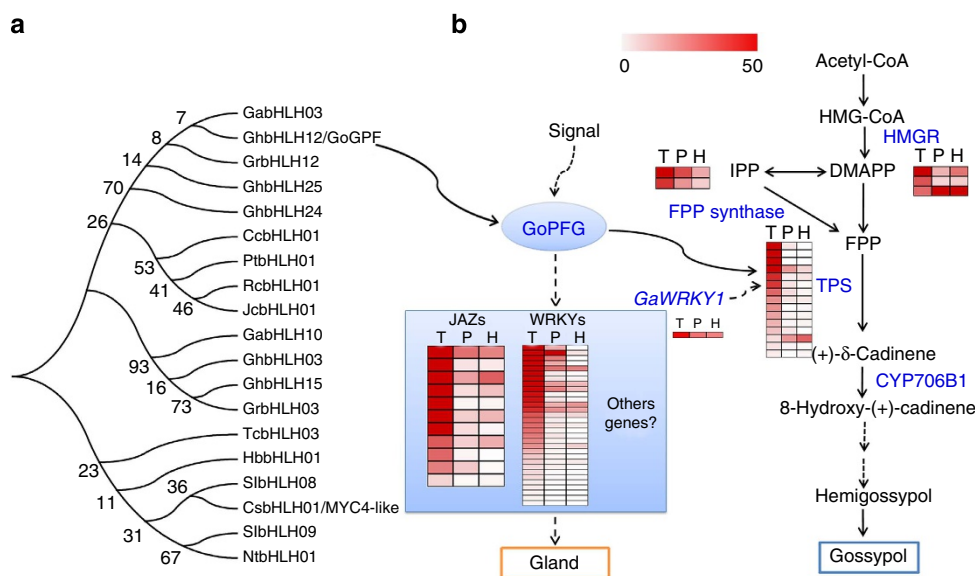


Figure 5 | A proposed model of gossypol biosynthesis and gland formation. (a) Among 111 bHLH members (Supplementary Data 2), 19 bHLHs from the distinct clade II may be responsible for glandular trichome formation. Numbers on branch nodes indicate percentage support in 1,000 bootstrap trials. (b) A proposed model of gossypol biosynthesis and gland formation. Blue text indicates genes downregulated in *GoPGF*-silenced TM-1. Heat map shows gene expression level with FPKM value in VIGS-free TM-1 (AT), *PGF*-silenced TM-1 (P) and Hai-1 (H), each block indicates one gene. First, the expression of *GoPGF* is induced by exogenous signal including JA. Then, *GoPGF* protein, as a regulator, controls the specification and differentiation of gland cells through regulating the expression of *JAZ*, *WRKYs* or other genes. On the other hand, *GoPGF* can specifically interact with the G-box motif, which is commonly found in *TPSs* and *WRKYs* promoter regions to feedback on gossypol biosynthesis pathway.

biosynthesis possibly through regulating the expression of gossypol-related genes such as *TPSs* and *WRKYs* by binding to their promoters. Taken together, cloning and functional assessment of PGF will open novel opportunities to decipher the mechanism of glandular trichome development in plants.

Australian *Gossypium* species such as *G. australe*, *G. bickii* and *G. sturtianum*, contain immature lysigenous glands but no terpenoid aldehydes, so the pigment glands only appear after seed germination; thus, the dormant seeds of these species lack gossypol^{33,34}. This distinguishing characteristic, known as the delayed gland morphogenesis trait, has the potential to enable the large scale, direct usage of cottonseed. Various efforts have been made to introduce this unique characteristic of wild Australian cotton species into cultivated tetraploid cotton³⁵, but such cultivars with the delayed gland morphogenesis trait have not been developed in cultivated tetraploid cottons by traditional breeding methods. The cloning of *GoPGF* may offer the opportunity to develop such cultivars as we investigate the role of *GoPGF* in the delayed gland formation in Australian species. These discoveries may help further to improve the productivity and economic value of cotton.

Methods

Plant materials. A mutant, “Hai-1” (*G. barbadense*), with a glandless trait controlled by a dominant gene, *G₁^g* (refs 17,19), was gifted by the Cotton Research Institute, Chinese Academy of Agricultural Science (CAAS). N1 and N7 are near isogenic lines of dominant glandless traits in Upland cotton. Three F₂ mapping populations of 1,599, 244 and 354 individuals, respectively, were developed by crossing TM-1 as a female parent to three glandless lines; Hai-1, N1 and N7. The gland traits of leaves were investigated in 2,197 individuals. TM-1 (*G. hirsutum*), Giza 45 and Hai7124 (*G. barbadense*), which have glands throughout the plants, and three other Upland cotton germplasms, 2(*G₁g₁*) and 2(*g₁g₁*), which have a few glands, and 2(*g₁g₁*), which has no glands, were used in this study. All cultivars were planted at Jiangpu Experiment Station and in green houses at Nanjing Agriculture University. Fresh leaves from the cultivars served as the source of genomic DNA, and other vegetative and reproductive tissues were collected for total RNA extraction. RNA was also extracted from developing embryos excised from ovules obtained from each boll at 3, 0, 3, 5, 10 and 20 days postanthesis. All collected plant materials were immediately frozen in liquid nitrogen and stored at -70 °C.

Map-based cloning of *G₁^g* gene. A previous study reported that the *G₁^g* locus lies between the molecular markers CLR362, NAU2251b, NAU3860b and STV033, with genetic distances of 9.27 and 0.96 cm²¹, respectively. The genetic markers were plotted to the scaffolds of *G. raimondii*²². The *G₁^g* locus was mapped to a 1-Mb region flanked by CLR362 and NAU2251b on scaffold D8, namely chr. 12, based on the linkage map constructed by JoinMap 3.0 (ref. 36). By using 598 F₂ mutant plants with additional molecular markers developed in this work (Supplementary Data 1), the *G₁^g* locus was further mapped to a 43-kb region between SSR marker w7954 and w5383. Then, the genomic DNA sequence of the candidate gene was amplified from *G. arboreum* cv. Jianglinzhongmian (A genome), *G. raimondii* (D genome), TM-1, Giza 36, Giza 45, Giza 67, Giza 80, Junhai-1, Hai7124 and Hai-1 (A_Dt genome) using primers (Supplementary Data 1), and the PCR products were confirmed by sequencing.

Subcellular localization of *GoPGF*. To examine the subcellular localization of *GoPGF* in cells, the PCR fragment amplified from cDNA from Hai-1 using the primers K0016F and K0016R (Supplementary Data 1) and inserted into transient expression vector pBinGFP4 (ref. 37) and generated the constructs pBinGFP4::*GoPGF*. The constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation. The recombinant plasmid was introduced into tobacco (*Nicotiana benthamiana*) leaf cells by *A. tumefaciens* infiltration³⁷. Green fluorescent protein signals in the tobacco epidermal cell were examined and photographed using a ZEISS LSM 710 confocal microscope (Zeiss Microsystems) with a ×20 objective lens (Zeiss) at the specific excitation and emission wavelengths 488 and 495–530 nm.

Quantitative RT-PCR analysis and pyrosequencing. RNA was extracted from the different tissues (root, stem, leaf and embryos) from plants with and without glands using a BioFlux kit. First-strand cDNA was generated using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech Co., Ltd.) according to the manufacturer’s instructions. Quantitative RT-PCR and pyrosequencing were performed with the primers listed in Supplementary Data 1. The pyrosequencing reaction was performed using a PyroMark sequencer (Qiagen, Valencia, CA).

Promoter activity analysis. The *GoPGF* promoter was isolated with the designed primers H3996F and H3996R from the A subgenome of TM-1 and Hai-1 (Supplementary Data 1). The promoters cover 1,818 and 1,809 bp in length upstream of the transcription start site. The binary vector pBI121 was used as a basic expression vector to make constructs. All the constructs were introduced into *A. tumefaciens* strain GV3101. Individual *Agrobacterium* colonies were grown on Luria-Bertani (LB) plates with kanamycin (50 µg ml⁻¹) for 48 h at 28 °C. A single positive colony was used to inoculate a 5-ml culture (LB with 50 µg ml⁻¹ kanamycin). Bacteria were pelleted, resuspended in infiltration medium (10 mM MgCl₂, 10 mM MES and 150 µM acetosyringone (pH 5.6)) to an OD₆₀₀ of 0.5–0.6, then incubated at room temperature for 3 h. The bacterial suspension was infiltrated into the abaxial side of fully expanded 6-week-old *N. benthamiana* leaves using a needleless 1-ml syringe. For each experiment, the positive control (pBI121 intron with 35S-GUS), the negative control (TATA-GUS in pMDC162 intron), and the constructs under investigation were infiltrated in areas of the same leaf or different leaves. After infiltration, the plants were kept in the greenhouse for 72 h for inoculation. For histochemical staining, the plant tissues were incubated at 37 °C overnight (12 h) in the dark in 1 mM X-Gluc (5-bromo-4-chloro-3-indolyl-b-D-glucuronide) in 100 mM sodium phosphate (pH 7.0), 10 mM EDTA, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.3% (v/v) Triton X-100 and 20% (v/v) methanol to eliminate endogenous GUS activity. After 12 h of staining, tissues were destained in an ethanol series (50, 70 and 95%) to remove chlorophyll, stored in 70% (v/v) ethanol, and photographed with a digital camera³⁸.

Virus-induced gene silencing assay. To knockdown the expression of *PGF* gene, a 415-bp fragment of *GbPGF_A12* cDNA from TM-1 corresponding to bases 904–1,319 of the *PGF* gene was PCR-amplified using Pfu DNA polymerase (Promega) and primers H1238F and H1238R (Supplementary Data 1). The resulting PCR product was cloned into *Xba*I-*Bam*H I-cut pTRV2 (ref. 25) to produce a VIGS vector named pTRV2-PGF_A12. The vectors pTRV1 and pTRV2-PGF_A12 were introduced into the *Agrobacterium* strain GV3101 by electroporation (Bio-Rad, Hercules, CA, USA). For the VIGS assay, the transformed *Agrobacterium* colonies containing pTRV1 and pTRV2-PGF_A12 were grown overnight at 28 °C in an antibiotic selection medium containing rifampicin and kanamycin 50 mg l⁻¹ each. *Agrobacterium* cells were collected and resuspended in infiltration medium (10 mM MgCl₂, 10 mM MES and 200 µM acetosyringone), adjusted to an OD₆₀₀ 0.5. *Agrobacterium* strains containing TRV1 and TRV2 vectors were mixed at a ratio of 1:1. Seedlings with mature cotyledons but without a visible rosette leaf (7 days after germination) were infiltrated by inserting the *Agrobacterium* suspension into the cotyledons via a syringe. The plants were grown in pots at 25 °C in a growth chamber under a 16-h light per 8-h dark cycle with 60% humidity.

Histochemistry and microscopy. Leaves detached from the seedlings before and after the treatment were cut into 1 mm² pieces and fixed in 0.1 mol l⁻¹ phosphate buffer (pH 7.0) containing 2.5% glutaraldehyde at 4 °C overnight. After three 30-min rinses in 0.1 mol l⁻¹ phosphate buffer (pH 7.0), samples were postfixed in 0.5% osmium tetroxide solution in buffer at 4 °C for 3 h. The samples were then rinsed in 0.1 mol l⁻¹ phosphate buffer (pH 7.0) for 30 min three times. Samples were dehydrated through a series of ethanol solutions (30, 40, 50, 65, 80, 90 and 100%; 30 min in each) twice, with a final change to 1, 2-epoxypropane, and were then embedded in Epon 812. Semi-thin sections (1–2 µm) were cut using a Reichert-Jung ultramicrotome and stained with toluidine blue O or methylene blue. Sections were examined and digitally recorded on a microscope.

Gossypol detection and analysis. The total gossypol concentration in leaves from TM-1, *PGF*-silenced TM-1, Hai-1, 2(*G₁g₁*), 2(*g₁g₁*) and 2(*g₁g₁*) plants was determined by high-performance liquid chromatography (HPLC)³⁹. Each 100 mg of freeze-dried plant sample was dissolved with 1 ml leaf extraction (acetonitrile/water/phosphoric acid = 80:20:0.1) for 1 h. The leaf extraction was centrifuged at low speed for 5 min and then transferred the supernatant carefully at room temperature. The eluent was filtered through a 0.22-µm nylon filter into a vial for HPLC analysis with Agilent Zorbax eclipse XDB-C18 analytical column (150 × 4.6 mm, 5 micron). The column was eluted with buffer (EtOH/MeOH/IPA/ACN/H₂O/EtOAc/DMF/PPAcD = 16.7:4.6:12.1:20.2:37.4:3.8:5.1:0.1) and kept at 35 ± 1 °C during the procedure. The determination wave length was 272 nm. Standards of gossypol, hemigossypolone, were used to assess the retention time of the individual terpenoids. The concentration of these compounds was calculated using Agilent 1100 system by comparing to the gossypol standard curve. All the reagents were of analytical grade and were made in China, with the exception of gossypol, which was purchased from Sigma Chemical Co.

RNA-seq. Total RNA samples were quality-checked using RNA Pico Chips on an Agilent 2100 bioanalyzer. All RNA samples were quantified and qualified with an RNA integrity number >8. RNA-seq libraries were constructed following the Illumina kit recommendations. The constructed libraries, indexed with six nucleotide sequences (barcode), were pooled together with equimolar amounts (2 nM) and were sequenced on the Illumina HiSeq 2000 sequencer with 2 × 100 bp. Raw data have been deposited in GenBank under the accession PRJNA265955. The raw FASTQ

format data generated from CASAVA v1.8.2 were first assessed for quality using FASTQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and FASTX toolkit v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). Poor quality reads (Phred score < 20) were trimmed at both ends with SolexaQA packages v2.2 (<http://solexaqa.sourceforge.net/>); only the reads with lengths ≥ 25 bp on both sides of the paired-end format were subjected to further analysis. The data were then aligned with the *G. hirsutum* (TM-1) genome (PRJNA248163). The software Cufflinks v2.2.1 (<http://cufflinks.cbc.umd.edu/>) was used to accurately quantify the abundance of genes and calculate the fragments per kilobase of genes per million mapped reads (FPKM). Differential expression was defined as a gene with a minimum of a twofold change (TM-1 versus PGF-silenced TM-1 and TM-1 versus Hai-1) with RPKM > 1 in either TM-1 or PGF-silenced TM-1 or Hai-1. K means clustering was performed with the open-source program, Cluster3.0 (<http://bonsai.hgc.jp/~mdehoo/software/cluster/software.htm>). The genes in each cluster were then classified into MapMan functional categories⁴⁰. Changes in the significance of expression were investigated in functional categories of the MapMan annotation through the application of a two-sided Wilcoxon rank test with a Benjamini-Yekutieli correction for multiple tests. Pathway analysis was mainly based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database⁴¹.

Development of the CYP706B1 RNAi transgenic line. To unravel the relationship between gossypol and glands, we cultivated the *CYP706B1* RNAi transgenic lines. A 408 bp fragment of *CYP706B1* (GenBank: AF332974.1) was amplified from the leaf cDNA library with dsP1-F-*Bam*HI: 5'-GGATCCTCAGC TCGTATTTCATGGCTG-3' and dsP1-R-*Xba*I: 5'-TCTAGACAAATACAATATC ATTGAGG-3' primers, and dsP1-F-*Sac*I: 5'-GAGCTCTCAGCTCGTATTTCATGG CTG-3' and dsP1-R-*Not*I: 5'-GCGGCCGCCAAATACAATATCATTGAGG-3' primers. The dsRNA construct was inserted into the *Bam*HI-*Sac*I site of binary vector pCAMBIA2301. The 1,143 bp globulin promoter, Proglobulin, was amplified with PG-F-*Hind*III: 5'-AAGCTTCTATTTTCATCTATTAG-3' and PG-R-*Bam*HI: 5'-GGATCCGATTACGATAAGCTCTGTAT-3' primers from cotton genomic DNA and inserted to the *Hind*III/*Bam*HI site to control dsRNA construct expression. The resultant constructs of Proglobulin::dsCYP706B1 (P1) was transferred into *A. tumefaciens* strain LBA4404 and then used for cotton transformation. Cotton seeds (R15) were surface-sterilized with 70% ethanol for 1 min, and 15% H₂O₂ for 4 h, followed by washing with sterile water five times. The sterilized seeds were germinated on Murashige and Skoog (MS) medium under dark conditions at 28 °C for 7 days. Hypocotyls were cut into 1-cm fragments and incubated with the overnight culture of *Agrobacterium* for 20 min. After 2 days of co-cultivation, hypocotyl explants were transferred to MS medium containing 1 $\mu\text{g ml}^{-1}$ 2,4-Dichlorophenoxyacetic acid (2,4-D), 50 $\mu\text{g ml}^{-1}$ kanamycin and 400 $\mu\text{g ml}^{-1}$ cefalotin for 2 months. The resistant calli were transferred to hormone- and antibiotic-free MS medium for somatic embryogenesis and plant regeneration. The resultant transgenic plants were transferred into soil for growing. Seeds from transgenic line P1-13-8 and progeny plants showed low gossypol contents. Genomic DNA from P1-13-8 and R15 were prepared for Southern blot analysis. *Hind*III- or *Eco*RI-digested genomic DNAs were separated on a 1% agarose gel, transferred to a Hybond-N⁺ nylon membrane (GE Healthcare Life Sciences) and probed with *NPTII* gene. The DIG High Primer DNA Labeling and Detection Starter kit II (Roche Applied Science, USA) were used for labelling and hybridization according to the manufacturer's protocol.

Yeast one-hybrid assay. The yeast one-hybrid assay was performed using the MATCHMAKER one-hybrid system (Clontech). Fragments containing four tandem copies of G-box (5'-CACGTG-3') (4 × G-box WT) and G-box mutant (5'-CATAGA-3'; 4 × G-box mutant) were synthesized by GenScript Biotechnology Co., Ltd. These two fragments were ligated into the *Hind*III-*Xho*I sites of pAbAi. The bait constructs were linearized with *Bst*BI and integrated into the yeast genome (strain Y1H). Various concentrations of aureobasidin A (AbA; Clontech, cat. no. 630446) on SD-Ura medium were used to identify the basal expression of AUR1-C. The ORF of *GoGPF* was ligated to the GAL4 activation domain in pGAD424. Yeast transformants were tested on SD-Ura medium containing 20 ng ml⁻¹ AbA.

Phylogenetic analysis. Alignment of the amino acid sequences of the *GoGPF* and bHLH families in other species was performed using the CLUSTALX program⁴². A neighbour-joining (NJ) method was then applied to produce a phylogenetic tree. The relative degree of branch support was determined within the NJ framework using the bootstrap procedure. The original data set was resampled 1,000 times.

References

- Hallahan, D. L., Callow, J. A. & Gray, J. C. Eds. *Plant Trichomes* Vol. 31 (Academic Press, San Diego, CA, USA, 2000).
- Graham, I. A. *et al.* The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. *Science* **327**, 328–331 (2010).
- Wagner, G. J. Secreting glandular trichomes: more than just hairs. *Plant Physiol.* **96**, 675–679 (1991).
- Schillmiller, A. L., Last, R. L. & Pichersky, E. Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.* **54**, 702–711 (2008).
- Schillmiller, A. L. *et al.* Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. *Proc. Natl Acad. Sci. USA* **106**, 10865–10870 (2009).
- Tissier, A. Glandular trichomes: what comes after expressed sequence tags? *Plant J.* **70**, 51–68 (2012).
- McDowell, E. T. *et al.* Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.* **155**, 524–539 (2011).
- Sunilkumar, G. *et al.* Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc. Natl Acad. Sci. USA* **103**, 18054–18059 (2006).
- Stipanovic, R. D., Bell, A. A. & Benedict, C. R. in: *Biologically Active Natural Products: Agrochemicals* (eds Cutler, H. G. & Cutler, S. J.) 211–220 (CRC Press, 1999).
- Mao, Y. B. *et al.* Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* **25**, 1307–1313 (2007).
- Tong, X. H., Zhu, S. J. & Ji, D. F. The expression and anatomical observation of the delayed pigment gland morphogenesis in an Upland cotton germplasm. *Cotton Sci.* **17**, 137–140 (2005).
- McMichael, S. C. Hopi cotton, a source of cottonseed free of gossypol pigments. *Agron. J.* **51**, 630 (1959).
- McMichael, S. C. Combined effects of glandless genes gl2 and gl3 on pigment glands in the cotton plant. *Agron. J.* **52**, 385–386 (1960).
- Lusas, E. W. & Jividen, G. M. Glandless cottonseed: a review of the first 25 years of processing and utilization research. *J. Am. Oil Chem. Soc.* **64**, 839–854 (1987).
- Endrizi, J. E., Turcotte, E. L. & Kohel, R. J. Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* **23**, 271–375 (1985).
- Lee, J. A. The genomic allocation of the principal foliar-gland loci in *Gossypium hirsutum* and *Gossypium barbadense*. *Evolution* **19**, 182–188 (1965).
- Kohel, R. J. & Lee, J. A. Genetic analysis of Egyptian glandless cotton. *Crop Sci.* **24**, 1119–1121 (1984).
- Tang, C. M. *et al.* Genetic analysis for Hai 1 strain of glandless cotton (*G. barbadense* L.): interaction between *Gl*₂ and *Gl*₁. *Cotton Sci. Sin.* **8**, 138–140 (1996).
- Afifi, A., Bary, A. A., Kamel, S. A. & Heikal, I. Bahtim 110, a new strain of Egyptian cotton free from gossypol. *Empire Cotton Growing Rev.* **43**, 112–120 (1966).
- Huang, Z. K. *The Cultivars and Their Pedigree of Cotton in China* (China Agriculture Press, 2007).
- Dong, C. G., Ding, Y. Z., Guo, W. Z. & Zhang, T. Z. Fine mapping of the dominant glandless gene *Gl*₂ in Sea island cotton (*Gossypium barbadense* L.). *Chin. Sci. Bull.* **52**, 105–1109 (2007).
- Paterson, A. H. *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).
- Zhang, T. *et al.* Asymmetric subgenomic evolution and domestication of allotetraploid cotton (*Gossypium hirsutum* L.). *Nat. Biotechnol.* **33**, 524–530 (2015).
- Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**, 594–597 (2006).
- Gao, X. *et al.* Silencing *GhNDR1* and *GhMCK2* compromises cotton resistance to *Verticillium wilt*. *Plant J.* **66**, 293–305 (2011).
- Luo, P. *et al.* Molecular cloning and functional identification of (+)- δ -cadinene-8-hydroxylase, a cytochrome P450 monooxygenase (*CYP706B1*) of cotton sesquiterpene biosynthesis. *Plant J.* **28**, 95–104 (2001).
- Benedict, C. R., Martin, G. S., Liu, J., Puckhaber, L. & Magill, C. W. Terpenoid aldehyde formation and lysisogenous gland storage sites in cotton: variant with mature glands but suppressed levels of terpenoid aldehydes. *Phytochemistry* **65**, 1351–1359 (2004).
- Hong, G. J., Xue, X. Y., Mao, Y. B., Wang, L. J. & Chen, X. Y. *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *Plant Cell* **24**, 2635–2648 (2012).
- Xu, Y. H., Wang, J. W., Wang, S. J., Wang, Y. & Chen, X. Y. Characterization of *GaWRKY1*, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)- δ -cadinene synthase-A. *Plant Physiol.* **135**, 507–515 (2004).
- Leng, L. & Lu, Y. G. Analysis of trichome density and the content of glandular secretion on different flue-cured tobacco leaves. *Appl. Mech. Mater.* **651–653**, 200–206 (2014).
- McDowell, E. T. *et al.* Comparative functional genomic analysis of Solanum glandular trichome types. *Plant Physiol.* **155**, 524–539 (2011).
- Weathers, P. J. *et al.* Artemisinin production in *Artemisia annua*: studies in planta and results of a novel delivery method for treating malaria and other neglected diseases. *Phytochem. Rev.* **10**, 173–183 (2011).

33. Zhu, S. & Daofan, J. I. Inheritance of the delayed gland morphogenesis trait in Australian wild species of *Gossypium*. *Chin. Sci. Bull.* **46**, 1168–1174 (2001).
34. Brubaker, C. L. Occurrence of terpenoid aldehydes and lysigenous cavities in the ‘glandless’ seeds of Australian *Gossypium* species. *Aust. J. Bot.* **44**, 601–612 (1996).
35. Dilday, R. Development of a cotton plant with glandless seeds, and glanded foliage and fruiting forms. *Crop Sci.* **26**, 639–641 (1986).
36. Ooijen, J. W. & Voorrips, R. E. JoinMap version 3.0: software for the calculation of genetic linkage maps. *Plant Res. Int.* (Wageningen, the Netherlands, 2001).
37. Liu, T. *et al.* Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nat. Commun.* **5**, 4686 (2014).
38. Chai, C. *et al.* Identification and functional characterization of the Soybean *GmaPPO12* promoter conferring *Phytophthora sojae* induced expression. *PLoS ONE* **8**, e67670 (2013).
39. Stipanovic, R. D., Altman, D. W., Begin, D. L., Greenblatt, G. A. & Benedict, J. H. Terpenoid aldehydes in Upland cottons: analysis by aniline and HPLC methods. *J. Agric. Food Chem.* **36**, 509–515 (1988).
40. Thimm, O. *et al.* MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939 (2004).
41. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, 480–484 (2008).
42. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405 (1998).

Acknowledgements

This work was financially supported in part by grants from NSFC(31201250) the Major State Basic Research Development Program of China (973 Program, 2011CB109300), the Priority Academic Program Development of Jiangsu Higher Education Institutions and the 111 Program.

Author contributions

T.Z. and X.C. conceptualized the research programme. T.Z., X.C. and Y.H. designed experiments and coordinated the project. D.M., Y.H., C.Y., B.L., Q.W., W.L., L.W., H.W., L.D., C.D., S.W., S.C., C.C., X.Z., X.G., B.Z., S.Z. and W.G. conducted laboratory experiments, and were involved in data analysis. B.L., L.F., D.M., G.M., M.P. and J.C. performed bioinformatic analyses. D.M., T.Z., Y.H., X.C., L.F. and J.W. participated in preparing and revising the manuscript. T.Z. and X.C. supervised data generation and analysis. All authors discussed results and commented on the manuscript.

Additional information

Accession codes: Sequences have been deposited at DDBJ/EMBL/GenBank under the accessions KP072743 (*GoPGF_A12* of TM-1), KP072744 (*GoPGF_D12* of TM-1), KP072745 (*GoPGF_A12* of Hai-1) and KP072746 (*GoPGF_D12* of Hai-1). Transcriptome data have been deposited in GenBank under the accession PRJNA265955.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Ma, D. *et al.* Genetic basis for glandular trichome formation in cotton. *Nat. Commun.* **7**:10456 doi: 10.1038/ncomms10456 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>