Genetic basis for glandular trichome formation in cotton

Dan Ma1,*, Yan Hu1,*, Changqing Yang2,*, 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 ginned 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 glandular. 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. 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. 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, 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, 0.65 kg of seed is also collected. According to this ratio, the global cottonseed production could potentially provide protein for half a billion people annually. However, the *Gossypium* species are characterized by the presence of small darkly pigmented lasing 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. 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.

Results

Map-based cloning of the dominant glandless gene *GbPGF*. We previously anchored the dominant glandless gene *GbPGF* 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*, *GbPGF* 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 *GbPGF* 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)F<sub>1</sub> than in the glanded *G. hirsutum* acc. TM-1 and *G. barbadense* cv. Hai7124. Therefore, we considered ORF2 as a candidate gene for *GbPGF*.

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 (Thecc1EG015640) 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 *F<sub>3</sub>*s (Supplementary Data 1).
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, gl2Gl3 and GbGl3; and a duplicated recessive cotton, gl2gl3. Black points in each graph show the glands on the surface of seeds.

Figure 1 | Origins of the genotypes and phenotypes of glandless tetraploid cottons. The genotypes include a dominant mutation, GbGl3; two recessive mutations, gl2Gl3 and GbGl3; and a duplicated recessive cotton, gl2gl3. Black points in each graph show the glands on the surface of seeds.

*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.
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-granulal and granulal 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 pathway28. 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
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 (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 GoPGF-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 biosynthesis, 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 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 μm. (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.
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, S. lycopersicum, A. annua 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 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 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.

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 TP5S and WRKYs by binding to their promoters. Taken together, cloning and functional assessment of PGF may open novel opportunities to decipher the mechanism of glandular trichome development in plants.

Australian Gossypium species such as G. australis, 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 gossypol33,34. 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 breeding methods 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, GoPGF (ref. 17), 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 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.

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 the PCR products were confirmed by sequencing.

Virus-induced gene silencing assay. To knockdown the expression of Gossypol gene, a 415-bp fragment of GoPGF_A12 DNA from TM-1 corresponding to bases 904–1,319 of the Gossypol gene was PCR-amplified using Pfu DNA polymerase (Promega) and primers H1238F and H1238R (Supplementary Data 1). The resulting PCR product was cloned into Xcel-BamHI 1-cut pTRV2 (ref. 25) to produce a VIGS vector named pTRV2-GPF_A12. The pTRV2-GPF_A12 was introduced into the Agrobacterium strain GV3101 by electroporation (Bio-Rad, Hercules, CA, USA). For the VIGS assay, the transformed Agrobacterium cells were cultured to an OD600 of 0.5–0.6, then incubated at room temperature for 3 h. The bacterial suspension was infiltrated in 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 introduced with 35S-GUS), the negative control (TATA-GUS in pMDKC162), 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 in 30% (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.

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 30 min in 0.1 mol l−1 phosphate buffer (pH 7.0) for fixation, the leaves were washed and dehydrated through a series of ethanol solutions (30, 40, 50, 65, 80, 90 and 100%), followed by 1,2-epoxypropane, and were embedded in Epon 812. Thin sections (1–2 μm) were cut using a Reichert-Jung ultramicrotome and stained with toluidine blue O or methyl blue. Sections were examined and digitally recorded on a microscope.

Gossypol content analysis. The total gossypol concentration in leaves from TM-1, Gossypol silenced TM-1, Hai-1, 2Glgl, 2GlgL, and 2GlgL plants was determined by high-performance liquid chromatography (HPLC)39. Each 100 mg of freeze-dried plant sample was dissolved with 1 ml leaf extract (acetone/tetramethylguanidine/water/phosphoric acid = 80:20:0:1) for 1 h. The leaf extract was centrifuged at 10,000 × g for 5 min and then transferred the supernatant carefully to a 50 ml ultracentrifuge tube. 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 (50 mm, 4.6 mm, 5 micron). The mobile phase consisted of 50 mM sodium acetate/ACN/H2O/PA/ACP/ACN/H2O/EO/ToOAc/DMF/PAAAd = 16:7:4.6:12:1.2:0.2:3.7:4.3:8.5:1:0.1 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 the standard 1100 system by comparing the gossypol standard curve. All the exogenous were of analytical grade and were made in China, except for the 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 PRJNA2659955. The raw FASTQ
format data generated from CASAVA v1.8.2 were first assessed for quality using FASTQC v10.1.0 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and FASTX toolkit v0.19.2 (https://hannonlab.cshl.edu/fastx_toolkit/). Fastq files were trimmed (Phred score <20) 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 (PRNA2A01863). The software Cuffdiff 2.2.1 (Iowa) was used to accurately calculate the abundance of genes and calculate the fragments per kilobase of gene 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 CYP706B1) with RPMK>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/~mdehoon/software/cluster/software.htm). The genes in each cluster were then classified into MapMan functional categories. Changes in the expression were investigated in functional categories of the MapMan annotation through the application of a two-sided Wilcoxon rank test with a significance of expression.

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 dp1-1-F BamH1: 5′-GGATCTCGAGGTGAAAGTCTTCCAGTCGATGCTTCTAGGC-3′ and dp1-1-R: XbaI: 5′-TCTAGAAGCTTCTATTTTCATCCTATTTAGA-3′ primers, and dp1-1-F: SacI: 5′-GACCTCGAGGAAGCTTCTATTTTCATCCTATTTAGA-3′ and dp1-1-R: NotI: 5′-GGCGGGCCCACTTAACTTATCATGTGGAGCTCTCAGCTCGTATTCATGGGT-3′ primers. The dsRNA construct was inserted into the BamH1-Sacl site of binary vector pCAMBIA2301. The 1.143 kb gus plasmid promoter, Progusul, was amplified with PG-F-HindIII: 5′-AACTCTCTTATTTCTATTTATGAGCTTCTATTTAGA-3′ and PG-R-BamH1: 5′-GGATCTCGAGGAAGCTTCTATTTTCATCCTATTTAGA-3′ primers from cotton genomic DNA and inserted into the HindIII/BamHI site to control dsRNA construct expression. The resultant constructs of Progusul: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 20 min. After 2 days of dark conditions at 28°C, 1 cefalotin for 2 months. The resistant calli were transferred to 1:2,4-Dichlorophenoxyacetic acid (2,4-D), 50 μM/C0176 for 20 min. After 2 days of Agrobacterium strain LBA4404 and then used for cotton transformation. Cotton seeds (R15) were surface-sterilized with 70% ethanol for 20 min. After 2 days of dark conditions at 28°C, 1 cefalotin for 2 months. The resistant calli were transferred to 1:2,4-Dichlorophenoxyacetic acid (2,4-D), 50 μM/C0176 for 2 months. The resistant calli were transferred to 16. 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 Gl2 and Gl. Cotton Sci. 8, 138–140 (1996).

Affi, 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., Gao, W. Z. & Zhang, T. Z. Fine mapping of the dominant glandless gene Gl in Sea island cotton (Gossypium barbadense L.). Chin. Sci. Bull. 51, 103–107 (2006).

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. & Doebler, J. A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescence architecture. Nat. Genet. 38, 594–597 (2006).

Gao, X. et al. Silencing GhNDR1 and GhMK2 compromises cotton resistance to Verticillium wilt. Plant J. 66, 293–305 (2011).

Luo, P. et al. Molecular cloning and functional identification of (+)-3-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 lysigenous gland storage sites in cotton: variant with mature glands but suppressed levels of terpenoid aldehydes. Phytochemistry 65, 1351–1359 (2004).

Hong, G. J., Xue, Y. X., Yao, M. 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 GhNDR1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-3-cadinene synthase-A. Plant Physiol. 135, 507–516 (2005).

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).

McWoll, 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. Jeannotte, 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., I.F., D.M., G.M., M.P. and J.C. performed bioinformatic analyses. D.M., T.Z., Y.H., X.C., I.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/