The Hairless Stem Phenotype of Cotton (*Gossypium barbadense*) Is Linked to a *Copia*-Like Retrotransposon Insertion in a *Homeodomain-Leucine Zipper* Gene (*HD1*)

Mingquan Ding,*,[1] Wuwei Ye,†,[1] Lifeng Lin,‡ Shae He,* Xiongming Du,† Aiqun Chen,*,§ Yuefen Cao,* Yuan Qin,† Fen Yang,* Yurong Jiang,* Hua Zhang,* Xiyin Wang, Andrew H. Paterson,**,[2] and Junkang Rong*,[2]**

*Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, College of Agriculture and Food Science, Zhejiang A&F University, Lin'an 311300, China, †State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China, ‡Nanosphere, Inc., 4088 Commercial Drive, Northbrook, Illinois 60062, §College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China, and **Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia 30605

ABSTRACT Cotton (*Gossypium*) stem trichomes are mostly single cells that arise from stem epidermal cells. In this study, a *homeodomain-leucine zipper* gene (*HD1*) was found to cosegregate with the dominant trichome locus previously designated as T1 and mapped to chromosome 6. Characterization of *HD1* orthologs revealed that the absence of stem trichomes in modern *Gossypium barbadense* varieties is linked to a large retrotransposon insertion in the ninth exon, 2565 bp downstream from the initial codon in the At subgenome *HD1* gene (*At-GbHD1*). In both the At and Dt subgenomes, reduced transcription of *GbHD1* genes is caused by this insertion. The disruption of *At-HD1* further affects the expression of downstream *GbMYB25* and *GbHOX3* genes. Analyses of primitive cultivated accessions identified another retrotransposon insertion event in the sixth exon of *At-GbHD1* that might predate the previously identified retrotransposon in modern varieties. Although both retrotransposon insertions results in similar phenotypic changes, the timing of these two retrotransposon insertion events fits well with our current understanding of the history of cotton speciation and dispersal. Taken together, the results of genetics mapping, gene expression and association analyses suggest that *GbHD1* is an important component that controls stem trichome development and is a promising candidate gene for the *T1* locus. The interspecific phenotypic difference in stem trichome traits also may be attributable to *HD1* inactivation associated with retrotransposon insertion.

KEYWORDS *Gossypium barbadense*; glabrous stem; *homeodomain-leucine zipper* gene (*HD1*); *T1* locus; *Ty1* retrotransposon element

TRICHOMES are mostly single-celled structures formed by the unidirectional extension of the outer layer of epidermal cells. They are of central importance to the *Gossypium* (cotton) genus, being its major economic product (lint fiber), an important taxonomic character, and an important determinant of pest resistance and other traits. The development of cotton fibers and *Arabidopsis* trichomes likely shares a similar regulatory mechanism involving closely related genes as well as a similar development and patterning signaling pathway, making trichomes an excellent system in which to study genetic mechanisms regulating cotton fiber development (Wan *et al.* 2014).

Trichomes are found on the stems of many *Gossypium* species and are even reflected in the species name of the most widely cultivated species, *Gossypium hirsutum* (whose name is derived from the Latin word *hirsutus*, meaning "hairy"). Stem trichomes exhibit striking phenotypic differences between different species and even among cotton varieties within the same species. Despite the fact that stem

doi: 10.1534/genetics.115.178236
Manuscript received May 16, 2015; accepted for publication June 25, 2015; published Early Online July 1, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178236/-/DC1
[1]These authors contributed equally to this work.
[2]Corresponding authors: Plant Genome Mapping Laboratory, 111 Riverbend Road, University of Georgia, Athens, GA 30605. E-mail: paterson@uga.edu; and College of Agriculture and Food Science, Zhejiang A&F University, Lin'an 311300, China. E-mail: jkrong@yahoo.com

trichomes show near-discrete genetic variations and are easier to detect than fibers, relatively few studies have focused on the development of cotton stem trichomes and their underlying regulation. Until now, through limited classical genetic studies and genetic mapping, at least five loci (T1–T5) (Lee 1985) and several additional genes and/or quantitative trait loci (QTLs) that modify the densities and distributions of stem trichomes (Wright *et al.* 1999; Lacape and Nguyen 2005) have been identified. The *T1* locus, which was mapped to chromosome 6 of the A genome, was predicted to contain major genes that regulate trichome initiation (Lacape and Nguyen 2005). This locus also has been reported frequently to cosegregate with multiple QTLs related to fiber fineness, length, elongation, color, and length uniformity (Wan *et al.* 2007), indicating that the underlying gene(s) in this locus is (are) involved in multiple epidermal trichome development processes.

Some important trichome developmental and patterning-related genes include *GLABRA1* (*GL1*), *GLABRA2* (*GL2*), *basic helix-loop-helix protein* (*GL3*), *ENHANCER OF GLABRA3* (*EGL3*), and WD40 protein *TRANSPARENT TESTA GLABRA1* (*TTG1*) in *Arabidopsis* (Larkin *et al.* 1994; Payne *et al.* 1999; Esch *et al.* 2003; Zhang *et al.* 2003; Wang *et al.* 2004; Humphries *et al.* 2005; Zhao *et al.* 2008; Cui *et al.* 2013; Zhou *et al.* 2014). Their homologous genes in cotton also have been identified as critical for trichome development. For example, *MYB109* encodes a product that works like the *Arabidopsis GL1* gene and is a key regulator in the development of epidermal hairs. *GaMYB2*, which was cloned from *G. arboreum* and is functionally homologous to *Arabidopsis GLABRA1* (*GL1*), can complement the *GL1* mutant phenotype involved in trichome formation (Wang *et al.* 2004). The overexpression of *GhRDL1* and *GhMYB2* increased *Arabidopsis* seed and leaf hairs by 8–10% (Guan *et al.* 2011). *GhTTG1* and *GHTTG3* could recover the trichome phenotype of the *Arabidopsis ttg1* mutant (Zhang *et al.* 2003). Recently, two cotton homologs of an *Arabidopsis* positive trichome formation regulator (*AtML1*), *GhHD1* and *GbML1*, were shown to be associated with cotton stem trichome formation (Zhang *et al.* 2010; Walford *et al.* 2012). *GhHD1*, which belongs to the HD-ZIP IV family, plays a vital role in regulating trichomes and fiber initiation as a member of an important protein complex, together with several key MYBMIXTA-like transcription factors (Zhang *et al.* 2010; Walford *et al.* 2012; Bedon *et al.* 2014).

As suggested by its species name, the vast majority of *G. hirsutum* accessions show pubescent stems and seeds, while *G. barbadense* varieties generally have glabrous stems and seeds. To date, the gene(s) responsible for these differences have remained unknown. In this study, we identified an insertion of a *Ty1-copia* retrotransposon into *GbHD1* in most *G. barbadense* varieties that correlated highly with their stem glabrous trichome phenotypes. We also determined that the retrotransposon insertion in *HD1* seems to be specific to the A subgenome locus in tetraploid cotton and almost invariably results in the loss of stem trichomes. To the best of our knowledge, this is the first report of a natural transposable element (TE) insertion causing a selectable phenotypic change in cotton.

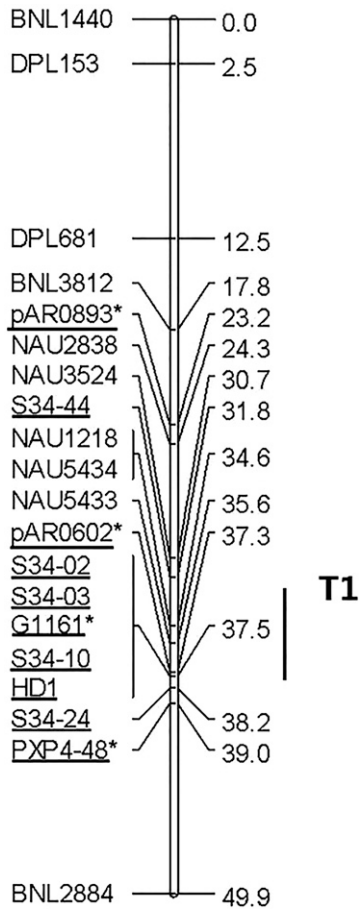

Figure 1 Genetic map of the chromosome fragment around a dominant trichome gene *T1* constructed by using an F_2 mapping population of T586 × Pima S6. *HD1* was mapped as a CAPS marker (PCR fragment was amplified using the primers UP2/Down2 and digestion by *Bst*z171I). The loci with lines below the name are SSCP markers, those with stars were developed from the RFLP map (Rong *et al.* 2004), and others are from the *G. raimondii* genome sequence (Paterson *et al.* 2012).

Materials and Methods

Plant materials and development of mapping populations

Sixty-eight *G. barbadense* varieties/accessions used in the association study were obtained from the Cotton Research Institute, Chinese Academy of Agricultural Sciences (CRI-CAAS), in Anyang, China, and the U.S. National Plant Germplasm System, in College Station, Texas (kindly provided by James Frelichowski) (Supporting Information, Table S1). The *G. barbadense* accessions include modern varieties (54/68) and primitive cultivated (14/68) forms (Table S1, primitive cultivated forms); the former were collected from different countries for broad representation of the gene pool and the latter from their native area (South America). In addition, some special germplasm was used, including T586

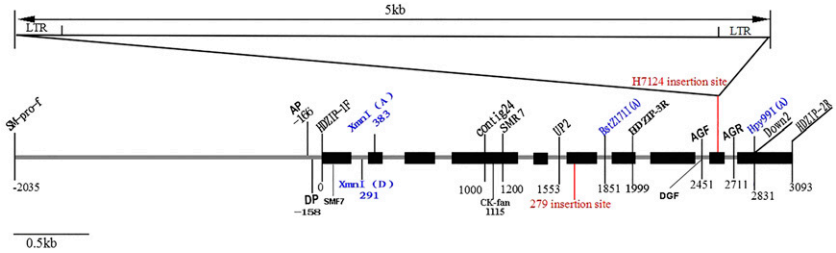

Figure 2 Structure of the *At-GbHD1* gene showing the position of some important restriction enzyme digestion sites, primers, and transposon insertions. Bold and thin lines represent exons and introns. Numbers are the positions of restriction enzymes, primers, and transposon insertions, indicated in base pairs, with the initiation codon assigned as zero.

(an upland cotton genetic marker line with several mutant phenotypes) and TM1 (a genetic standard line). In addition to the materials listed in Table S1, a *G. arboreum* accession (A2-117) and *G. raimondii* (D5) also were used for DNA sequencing analysis.

Two mapping populations were developed from crosses between T586 and Pima S6 to map the trichome gene and between T586 and Tu 75-37 to determine the association of the glabrous stem phenotype and retrotransposon insertion.

All cotton materials were grown on the campus farm of Zhejiang A&F University, Lin'an, China, except the primitive cultivated forms, which were grown in pots and kept in the greenhouse during the winter. Plants of each variety/accession were labeled for DNA and RNA extraction as well as for trichome observation.

Marker development and DNA extraction

Genetic marker candidates that are potentially linked to trichome phenotypes were chosen from both published A genome (He *et al.* 2013) and tetraploid genome mapping studies (Wright *et al.* 1999; Rong *et al.* 2007). Candidate RFLP markers were screened for polymorphisms between the parents of the mapping population in this study (T586 and Pima S6), as reported by He *et al.* (2013). In addition, the genome sequences of *G. raimondii* (Paterson *et al.* 2012), spanning 5624.28 kb ranging from 26,943 to 5,651,218 nt [trichome gene (*T1*) region] of chromosome 10 determined by synteny between genetic maps (Wright *et al.* 1999; Rong *et al.* 2007; Desai *et al.* 2008) and a physical map (Paterson *et al.* 2012), were used to develop additional DNA markers. Primer design and PCR and SSCP procedures were reported previously (He *et al.* 2013). In brief, primer design for SSCP marker development was based on the following criteria: the length of the amplified fragments should be 250–600 bp and cover the introns. PCR products were first checked on 1% agarose gels, and those showing one to two clear bands for both genotypes were used for SSCP employing the protocol described by Lee *et al.* (1992) with some modifications (He *et al.* 2013).

DNA from different genotypes including the plants of the mapping populations mentioned earlier was extracted as described by Paterson *et al.* (1993) and slightly modified as follows: about two to three leaf buds were picked from each plant, placed in 2-ml Eppendorf tubes, and homogenized with either an electric drill in a cotton nucleic lysis buffer or in liquid nitrogen. After incubation in 65° for 25–30 min, the supernatant was purified with 24:1 chloroform: isoamylol once. The supernatant was then precipitated with two volumes of 95% ethanol, and the pellets were dissolved in 300 μl H_2O and used as templates for PCR.

Trichome observation and map construction

To construct a genetic linkage map, 124 F_2 plants derived from a cross between T586 and Pima S6 were used. When the plants started to flower, trichomes on the stems were observed and scored as either present or absent (*i.e.*, as a dominant genetic marker). The linkage group was built using MapMaker 3.0, as described by Rong *et al.* (2004). SSCP markers were pooled together with trichome phenotype data, and linkage was determined at a level of LOD $>$ 3.0.

RNA extraction and subgenome origin assignments of amplicon

Young stems were collected and homogenized in liquid nitrogen. RNA was extracted following the CTAB protocol and purified using Qiagen RNeasy columns (Qiagen, Hilden, Germany). Reverse transcription (complementary DNA synthesis) was performed using the TAKARA PrimeScript 1st Strand cDNA Synthesis Kit (TAKARA BIO, Inc., Canton, MA). For RT-PCR assay, a pair of primers (SMF7/SMR7) was designed to amplify a 682-bp fragment including a SNP recognized by *Xmn*I to distinguish the A and D subgenome–specific genes. Digestion of PCR products by *Xmn*I will give rise to two fragments of 204 and 478 bp in size, respectively, when only the A subgenome–specific *GbHD1* is expressed. However, when only the D subgenome–specific *GbHD1* is expressed, the PCR products will not be cleaved by *Xmn*I, and only a single 682-bp fragment will be detected. When both A and D subgenome–specific *GbHD1* genes are expressed, DNA fragments of all three sizes will be present.

Cloning of the HD1 gene, restriction digestion, and sequence analysis

Primer design: A series of primers was designed using DNA sequences around the *HD1* gene ranging over 15,362 bp (from 5,518,596 to 5,532,963 bp) of chromosome 10 of *G. raimondii* as reference for PCR amplification (Table S2 and Figure S1). The primers can be divided into four groups according to their purpose. Group 1 has only one pair of primers, HDZIP1F/HDZIP2R, that was used to amplify the entire *HD1* gene. Group 2 consists of At and Dt subgenome–specific primers, including four forward (AP, AGF, DP, and

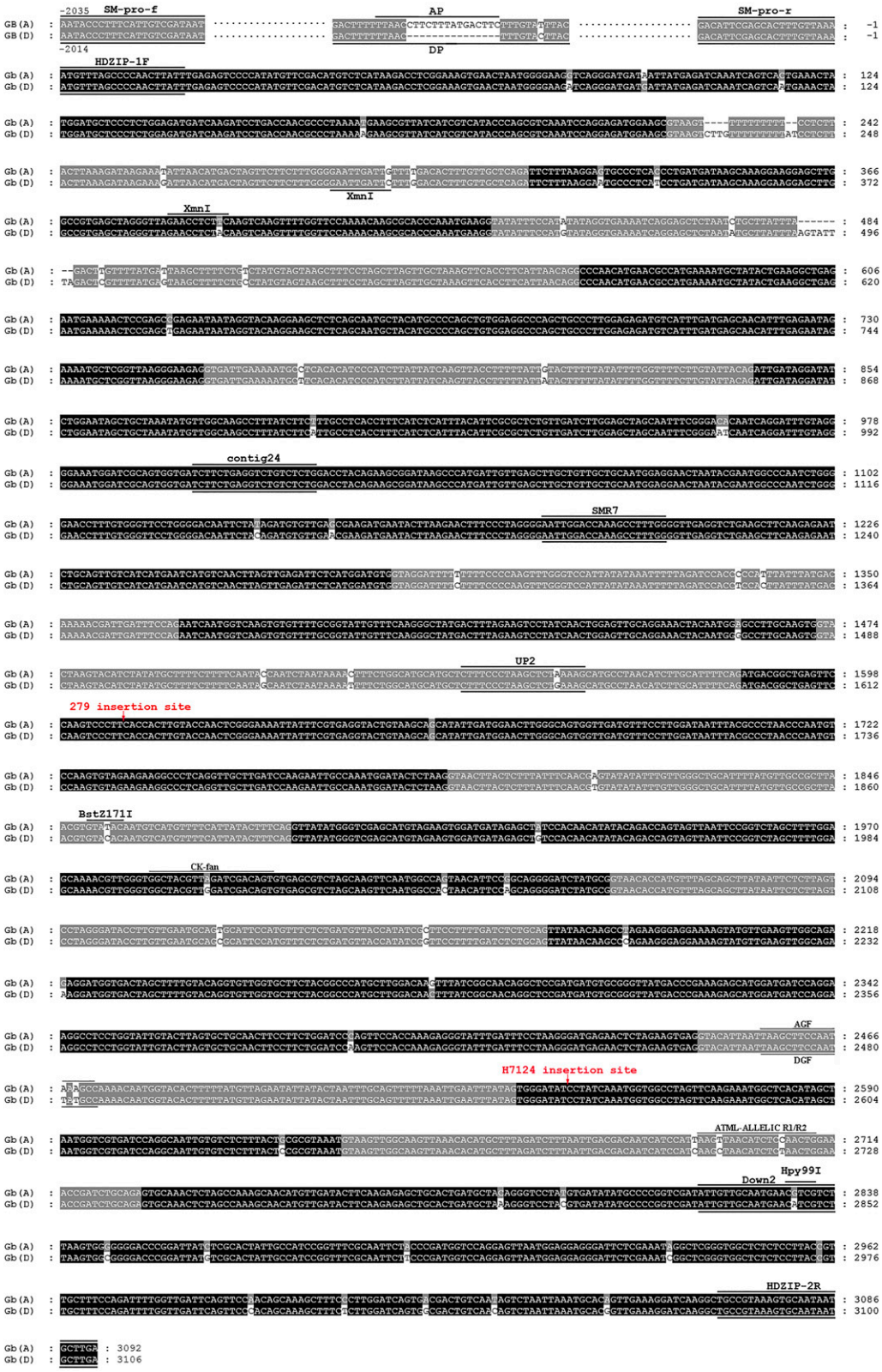

Figure 3 Sequence alignment of *At-GbHD1* and *Dt-GbHD1* genes showing the gene structure and positions of important primers and restriction enzyme digestion sites. Sequences highlighted in black and gray are exons and introns, respectively. Lines above and below the sequences indicate the primer sequences, with their names listed above and below these lines. Arrows indicate the positions of transposon insertions.

DGF) and two reverse (AGR and DGR) primers that were used to amplify a fragment from a specific subgenome. Group 3 primers were designed to amplify a fragment including a restriction site and used to distinguish the genomic identity of PCR products. Group 4 primers were used to delimit the site where the DNA change occurred in the At subgenome *HD1* gene that resulted in no amplification of PCR product.

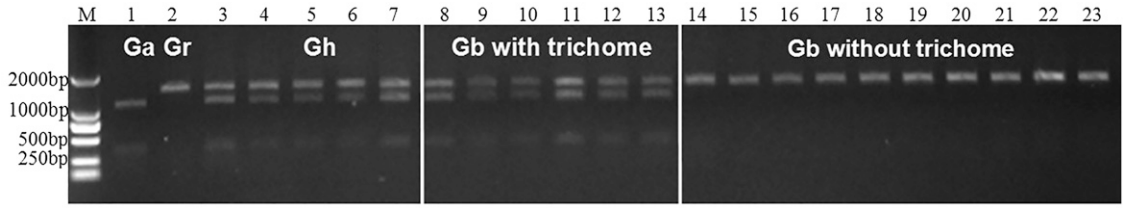

Figure 4 Different digestion pattern of fragments amplified from different cotton species using primers UP2/Down2 and digestion with *Bstz*17I. Ga, *G. arboreum* (diploid A genome); Gr, *G. raimondii* (diploid D genome); Gh, *G. hirsutum* (AD1), lanes 3–7; Gb, *G. barbadense* (AD2), lanes 9–23.

PCR reaction: Each PCR reaction generally contains 50 ng of cotton genomic DNA, 10X buffer (Mg^{2+} Plus), 2.5 mM deoxynucleotide triphosphates (dNTPs), 5 μM of each primer, and 0.75 U of Taq DNA Polymerase (TAKARA BIO) in a total volume of 15 μl. PCR reactions were typically run using the following program: predegenerate at 94° for 5 min, then run the cycle of denature at 94° for 45 sec, annealing at 52–58° for 30 sec, and extension at 72° for 1 min for 30 cycles, followed by a final extension at 72° for 7 min. This program may be modified according to the size of the PCR products. For example, TAKARA LA Taq was used to amplify fragments of over 3 kb in size using the primer pair HDZIP1F/HDZIP2R, for which a 3-min extension time was used. When attempting to amplify a fragment with the TE insertion using the subgenome-specific primer such as AF2 and DF2, TAKARA LA Taq with GC buffer was applied, and a 3-min extension time also was used.

Sequencing of PCR products: For sequencing, PCR products were first checked on agarose gels, and the targeted bands were excised. The fragments were purified using an AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA) and cloned into a TA cloning vector to transform *Escherichia coli*. Positive clones were identified by colony PCR and sent out for sequencing by the Beijing Genomics Institute (BGI). To assess the attribution of A and D subgenomes in tetraploid cotton, about 10 clones were first sequenced for T586 (*G. hirsutum*) and Pima S6 (*G. barbadence*), together with A2-113 and D5 as progenitor genome controls using the primer pair of HDZIP1F/HDZIP2R (Figure 2 and Table S2). DNA sequences from different clones were aligned using Vector NTI Advance 10.0, and sequencing errors were manually inspected to obtain the correct sequence. The cDNA sequences from *G. arboreum* accessions A2-113 and T586 were used to determine the gene structure (exon and intron).

Restriction digestion: To distinguish the A and D subgenomes in a tetraploid cotton, restriction sites were checked according to the preceding alignment. PCR product was cut by the restriction enzyme and checked on 1% agarose gels. The genome attribution of the genes was determined using diploid genomes as control.

Data availability

Table S1 contains detailed descriptions of all used cotton materials. Table S2 contains all of the primers designed for this research. Table S3 and S5 contains all the HD1 related sequencing results for each cotton variety and individual.

Results

Trichome phenotypes in G. barbadense

Six-hundred and fifty-four individual plants from 68 varieties and accessions of *G. barbadense* (54 modern cultivars and 14 primitive cultivated forms) were observed for their stem trichome phenotypes. Among these, 12 (100 plants) showed pubescent phenotypes, 48 (444 plants) showed glabrous phenotypes, and 8 (110 plants) contained individuals with or without trichomes. The third group could be further classified into two subgroups based on the frequency of plants having trichomes in the progeny population. Six varieties (NLD31, -33, -35, -40, -42, and -47; Table S1) defined the first subgroup, in which only a low frequency [up to 11.3% (or 9/80)] of plants with stem trichomes was observed. The other two varieties (NLD48 and -51; Table S1) defined the second subgroup, in which over 90.0% (or 27/30) of individual plants showed stem trichomes (Table S1).

Close correlation between HD1 and trichome phenotypes revealed by genetic mapping in cultivated tetraploid cottons

The presence or absence of trichomes is a single-gene trait and follows Mendelian inheritance in *G. arboreum* and *G. hirsutum* (Wan *et al.* 2007; Desai *et al.* 2008). This trait cosegregates with the loss-of-fiber mutant *sma-4* (Desai *et al.* 2008; He *et al.* 2013). Recently, map-based cloning of *sma-4* was conducted by our research group using a large-scale high-resolution fine-mapping population containing 2500+ individuals, which suggested strongly that the underlying causative gene for the *sma-4* phenotype is *HD1* on chromosome LG A03.

To explore the linkage between the *HD1* locus and trichome phenotypes in tetraploid cotton, we developed a mapping population including 124 F_2 plants from a cross between T586 and Pima S6. Using this population, a local linkage map was constructed comprising 10 simple sequence repeat (SSR), 9 single-strand conformation polymorphism (SSCP),

Table 1 Relationship of trichome phenotypes to *At-HD1* genotypes in *G. barbadense*

At-HD1	Variety[a]	No trichomes	With trichomes
No	33 (351)	33 (351)	0(0)
Yes	10(92)	0(0)	10(92)
No/yes[b]	8(110)	8(74)	8(36)
Total	51(553)	41(425)	18(128)

[a] The numbers in parentheses are the total plant number studied.
[b] Varieties showing segregation of the *HD1* gene on the A genome [with and without the *HD1* gene on chromosome 6(A) and trichome growth (with and without stem trichomes].

Table 2 Relationship of trichome phenotypes to segregation of the *At-HD1* gene in *G. barbadense* varieties

At-HD1	Variety[a]	No trichomes	With trichomes
No	8(74)	8(74)	0(0)
Yes	8(36)	0(0)	8(36)

[a] The numbers in parentheses are the total plant number studied.

and 1 cleaved amplified polymorphic sequence (CAPS) markers for *HD1*, spanning 49.9 cM, with an average interval of 2.26 cM between markers (Figure 1). The trichome phenotype was found to cosegregate with *HD1*, a restriction fragment length polymorphism (RFLP)–derived SSCP marker (G1161), and three other SSCP markers developed based on the *G. raimondii* genome sequence (Paterson *et al.* 2012), which were flanked by pAR0602 and S34-24 at 0.2 and 0.7 cM, respectively (Figure 1).

Sequence characterization of the HD1 gene in cultivated tetraploid cottons

A pair of primers (HDZIP1F/HDZIP2R, group 1) was designed to amplify the entire coding region (from the start codon to the stop codon) of *HD1* orthologs from both subgenomes for sequencing to study the allelic variations of *HD1* in *G. barbadense* (Figure 2 and Figure 3, Figure S1, and Table S2). A total of 185 clones from 18 plants of 15 *G. barbadense* varieties, including six primitive cultivated forms, was sequenced (Table S3). Sequencing results confirmed the identity of the *HD1* gene, which consists of 10 exons and 9 introns, by the presence of basic HD domains in the predicted protein (Zhang *et al.* 2010) (Figure 2 and Figure 3). The A genome gene amplicon is 3105 bp long, while the D genome amplicon is 3104 bp long, showing 97.6% identity with 16 predicted amino acid changes. Sequence alignment revealed that all *G. barbadense* varieties fall into three groups: the first group of varieties retained *HD1* orthologs from both At and Dt subgenomes. L-7009 (numbered as NLD52) is an example of this group. The second group comprises varieties whose *At-HD1* is missing, and only *Dt-HD1* could be detected (Table S3). Examples of this group are 9078N (NLD75-1), Pima S6, 25686 (NLD45-1), and Gb-Yumian. The third group showed a mixed result among individual plants from the same variety, some resembling group 1 with both homologs intact and others resembling group 2 containing only the Dt homolog. For example, NLD51-1 contains only the Dt homolog, while NLD51-2 contains both copies. Interestingly, no plants containing only the At homolog were identified.

Association between trichome growth and an intact HD1 gene on the A subgenome

The presence and absence of stem trichomes (pubescent phenotype) perfectly correlated with the presence and absence of *At-HD1* in the 15 *G. barbadense* varieties studied (Table S4). In all plants examined, among both uniform and segregating varieties, all plants having trichomes also had both *At-HD1* and *Dt-HD1* orthologs, whereas plants having no trichomes lacked a copy of *At-HD1*.

To further verify this correlation, we extended the study of the retention patterns of *At-HD1* genes to a larger population of 116 plants from 52 *G. barbadense* varieties using a homeo-SNP (SNP between genomes, T/C) at a position 1851 bp downstream of the start codon. This site can be used to distinguish *At-* and *Dt-HD1* in tetraploid cottons by digestion with the restriction enzyme *Bstz17I* (Figure 2 and Figure 3). As expected, the amplified sequences of *At-HD1* (primer: UP2/Down2) (Figure 3 and Figure S2) could be cut into two fragments with expected sizes of 985 and 315 bp, respectively, while the *Dt-HD1* amplicons remained intact, with the expected size of 1300 bp (Figure 4).

Exploring the sequencing and enzyme-cutting results indicated that among 51 *G. barbadense* varieties, including the 12 primitive cultivated forms for which trichomes have been scored, 10 contained both *At-* and *Dt-HD1* orthologs (like group 1), 33 exhibited only the *Dt-HD1* allele (like group 2), and 8 showed mixed patterns of loss of *At-HD1* among individuals (like group 3) (Tables 1 and 2). Again, there was a perfect correlation between the presence of trichomes and *At-HD1*, lending further support to the hypothesis that lack of *At-HD1* is highly related to lack of stem pubescence in *G. barbadense*.

Lack of PCR amplification of At-HD1 in glabrous plants is caused by Ty1 LTR-retrotransposon insertions

To further investigate the reason for the lack of amplification of *At-HD1* in glabrous *G. barbadense* varieties, a series of overlapping primer pairs (group 4) was designed from both the upstream and downstream regions of the *HD1* gene for PCR (Figure S1 and Table S2). Using five pairs of primers (SM1–5) in the promoter regions, the PCR results suggested that the maximum left border of the missing amplification of *At-HD1* was near the junction between the *At-HD1* promoter and coding regions. In this experiment, we sequenced 92 clones of the promoter region, with a length of 2006–2036 bp upstream from the start codon of 4 *G. hirsutum* and 10 *G. barbadense* varieties using the primers sm-pro-F/R (Table S5). These clones could be divided into two groups based on their sequence similarities to either the diploid *At-HD1* or *Dt-HD1* gene, respectively, using the sequences of eight clones from A2-47 and *G. raimondii* (D5) genome sequences as references (Table S3). A deleted 15-bp DNA

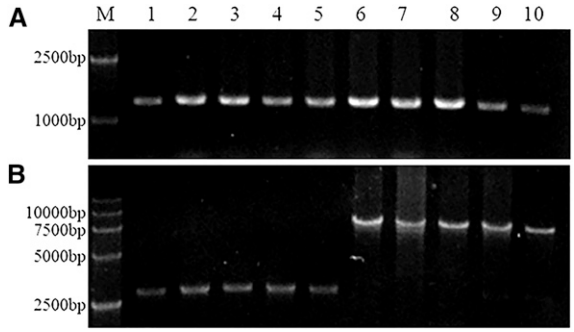

Figure 5 PCR amplification of *GbHD1* genomic DNA fragments of different sizes in the *G. barbadense* varieties with glabrous or pubescent stems. A. All varieties (lanes 1–10), whether they have glabrous or pubescent stems, produced a PCR amplicon with the same size using AP/SMR7 primers. B. Varieties (lanes 1–5) with pubescent stems produced a 3007-bp PCR fragment; varieties (lanes 6–10) with glabrous stems produced a PCR fragment of ~8 kb using AP/Down2 primers.

fragment in the *Dt-HD1* promoter region, which could easily distinguish At and Dt copies, was used to design forward primers specific to the At copy (named as AP, groups 2 and 4) compared with the Dt copy (designed based on the sequences flanking the specific deleted region and named DP) (Figure 3 and Figure S1 and Figure S2). A series of reverse primers (group 4) also was designed based on the downstream regions of *HD1* (Table S2). When the reverse primers were used with UP2-rvs, a fragment of the same size was amplified in varieties both with and without trichomes (Figure 5A), indicating the presence of *HD1* on both chromosome 6(A) and chromosome 25(D). However, when the reverse primer was moved down to Down2 or HDZIP2R, a fragment of ~8 kb was amplified in the varieties without trichomes (and without amplification of *At-HD1* under normal PCR conditions), but in the varieties with trichomes (with amplification of *At-HD1* under normal PCR conditions), an expected fragment was amplified (Figure 5B). These results suggested that a DNA fragment was inserted into the region between primers UP2 and Down2 (Figure 2) causing failure to amplify the At copy with primers UP2/Down2 and HDZIP1F/HDZIP2R in the initial PCR reactions. Further sequencing results revealed an insertion of a 4999-bp fragment derived from a *Ty1-copia* LTR retrotransposon at a position 2565 bp downstream from the *HD1* start codon. All components of a typical *Ty1-copia* element, such as long terminal repeats (LTRs), gag and pol, were readily identifiable in the insertion sequence (accession no. KF740825) (Cao *et al.* 2015).

To reveal whether all *G. barbadense* plants lacking *At-HD1* amplification have the same *Ty1* LTR-retrotransposon insertion (hereafter called the *TE insertion*), three pairs of primers (AGF/AGR, AGF/LTR-R, and LTR-L/AGR) were used. One was a pair of At-genome-specific primers that flanked the insertion point closely and could amplify a 261-bp fragment in *HD1* without TE insertion or a 5260-bp fragment including the 4999-bp insertion and 261 bp of the gene (Figure 6). The other two pairs of primers were a combination of the preceding forward primer and the left LTR primer (LTR-R, designed based on the end sequence of LTR), as well as the preceding reverse primer and right (forward) LTR primer. All modern *G. barbadense* varieties were found to carry a fragment of 5260 or 261 bp or both (heterozygote) when the AGF/AGR primer pair was used (Figure 6) or a strong band of 5092 bp and a weak band of 471 bp when AGF/LTR-R and LTR-L/AGR were used. This confirms that the modern varieties with glabrous stems have the same LTR retrotransposon at the same location in *At-HD1* (Table S1).

To further illustrate the correlation between glabrous stem and TE insertion in *At-HD1*, the segregation of trichome phenotype and the TE insertion were analyzed in a population of 96 F_2 plants derived from a cross between T586 and Tu 75-37 (a *G. barbadense* variety with glabrous stem and TE insertion). Twenty-seven plants were found with no stem trichomes, like Tu 75-37, and the remaining 69 plants had varied numbers of trichomes, with the ratio of plants with and without trichomes being not significantly different from the 3:1 segregation of a single gene. The TE insertion was found to perfectly correlate with trichome growth; *i.e.*, all 27 plants with glabrous stems had the TE insertion, and all 69 plants having trichomes on their stems either did not have TE insertions (33 plants) or were heterozygous for TE insertion (36 plants) (Figure S2). This result indicated that trichome initiation is perfectly associated with the *At-HD1* gene and that the pubescent stem trait is dominant to the glabrous stem.

Ty1 transposon insertion occurred repeatedly in the At-HD1 gene during history

The timing of the TE insertion in *At-HD1* of *G. barbadense* was investigated by identifying both the trichome phenotype and *At-HD1* of 14 primitive cultivated cottons originating from Central and South America (Table S1). All these varieties had abundant vegetative growth with many branches and strong main stems 2–2.5 m high. Only three of them bloomed—but almost 2 months later than day-neutral cultivars in this temperate latitude; they also set few bolls. Three accessions (GB81, GB413, and GB656) had many trichomes during the seedling stage, while two others (GB413 and GB1600) grew smaller numbers of trichomes around 2 months after planting. The other nine accessions had few (GB102) or no trichomes.

The preceding *HD1* gene primers were applied to study the *At-HD1* gene in these primitive varieties. PCR products from 7 of 14 primitive cultivated forms were sequenced and digested by restriction enzymes. As found in modern varieties, the trichome phenotype in the primitive cultivated forms was closely association with the amplification of *At-HD1* (Table S1). Enzyme digestion and sequencing of PCR products with primers of HDZIP1F/HDZIP2R and UP2/Down2 indicated that 6 of 14 (43%) primitive cultivated forms have both the At and Dt copies of *HD1*, which represented a much higher abundance of *At-HD1* than that

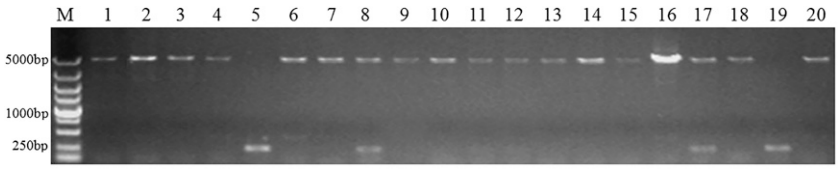

Figure 6 PCR amplification of fragments with (*e.g.*, lanes 1–4) and without DNA insertion (*e.g.*, lanes 5 and 19), as well as both (heterozygous, lanes 8 and 17) in *G. barbadense* varieties collected from Anyang, Henan, China, using AGF/AGR primers.

observed in modern *G. barbadense* varieties (~5%). The promoters of the two accessions (GB217 with insertion and GB656 without insertion) also were sequenced, and the sequences of both At and Dt subgenomes were the same as those in modern varieties (Table S3). When primer AP/Down2 was used, a fragment of ~8 kb also was amplified in nine accessions that did not show normal amplification of *At-HD1* (Table S1), indicating that a TE insertion probably also occurred in the primitive cultivated forms of *G. barbadense* (Figure 7). However, the sizes of the larger fragments were slightly different among different accessions, indicating that they were probably different insertions. Using primers AGF/AGR, which are specific to identify the *Ty1* LTR-retrotransposon insertion that is ubiquitous among modern *G. barbadense* cultivars, only three accessions (GB333, GB372, and GB1596) had the same retrotransposon at the same site as modern varieties. Thus, primitive forms appear to have experienced multiple independent insertions into *At-HD1*, only a subset of which has been carried over to modern varieties. When AP/CK-fan primers were used, a 6663-bp fragment was amplified from the six accessions lacking *At-HD1* but that do not have the *Ty1* LTR-retrotransposon insertion that is ubiquitous among modern *G. barbadense* cultivars (Figure 7). Finally, a TE insertion event was confirmed to have occurred in the sixth exon, 1609 bp downstream from the start codon. Sequencing of this DNA insertion indicated that the transposon is almost the same as that isolated from the other cotton varieties, except that it lacked 516 bp, including the right LTR, and had an additional 34 bp on the right side of the TE (Figure S3). Based on the sequence divergence between the two LTRs, we deduced that this TE insertion occurred earlier than the one at the ninth exon in the modern varieties. Consequently, we named the earlier event TE1 and the later event TE2.

Defective HD1 gene expression is caused by the retrotransposon insertion

Two *G. barbadense* accessions with pubescent stems and four with glabrous stems were chosen for RT-PCR analysis. Accessions that have stem trichomes (without TE insertion) showed three bands after digestion, and accessions with no stem trichomes (with TE insertion) showed only one band, which was the same size as the undigested Dt genome, or this band plus two very faint bands (Figure 8). These results suggested that both the *At-HD1* and *Dt-HD1* genes are expressed at similar levels in plants with pubescent stems. In plants with glabrous stems, only the *Dt-HD1* gene was expressed, *At-HD1* being faint or absent. These results indicated that TE insertion into *At-HD1* caused defective gene expression, which may have resulted in the inhibition of trichome growth.

In addition, *HD1* gene expression in the stem also was analyzed by qualitative PCR (qPCR). The results indicated that in modern domesticated cottons, NLD26 (with the retrotransposon insertions at the *At-HD1*) displayed significantly lower (fivefold) gene expression than NLD52 (containing the intact *At-HD1* gene) (Figure 8). Similar lower expression patterns were detected in the *At-HD1*-disrupted primitive cottons, such as GB217 and GB1596, compared with the *At-HD1*-intact varieties, such as GB656. To further evaluate the effects of blocked gene expression of *At-HD1* on the function of downstream members of the signal pathway in regulating trichome initiation in these *G. barbadense* accessions, the expressions of *GbMYB25* and *GbHOX3* also were analyzed. As seen in Figure 9, *GbMYB25* and *GbHOX3* showed parallel gene expression trends to *GbHD1* in stems of these *G. barbadense* accessions; *i.e.*, they were expressed at a lower level in TE-inserted varieties/accessions than in those with intact *HD1* genes, no matter whether they were modern domesticated or primitive cottons.

Discussion

HD1 is critical to stem trichome growth in Gossypium

Trichome initiation in *Arabidopsis thaliana* is an important model for understanding cell fate and patterning (Marks 1997; Larkin *et al.* 2003; Marks and Esch 2003; Hulskamp 2004). By contrast, very little is known about the molecular mechanism of trichome growth in cotton. Although some cotton homologs of *Arabidopsis* trichome genes have been cloned and found to be related to cotton fiber and trichome growth (Yang and Ye 2013), there is no direct evidence that confirms which gene(s) is(are) critical to cotton trichome growth and how the genes regulate trichome growth and development collectively. Classic genetic studies have revealed that cotton stem and leaf trichomes are mainly controlled by a dominant gene (*T1*) located on chromosome 6 (A) in tetraploid cotton (Wright *et al.* 1999) and its homoeologous chromosome LG A03 of diploid A genome species (Rong *et al.* 2005; Desai *et al.* 2008; He *et al.* 2013). The same mapping result was found in this study using two F_2 populations from the crosses between T586 and two other *G. barbadense* varieties (Pima S6 and Tu 75-37). In *Arabidopsis*, *ATML1*, encoding a putative HD-ZIP transcription factor, was confirmed to affect trichome growth together with *AtPDF2* (Abe *et al.* 2003). The cotton *HD1* gene, which is the homolog of *ATML1*, cosegregated with *T1* in this study, suggesting that

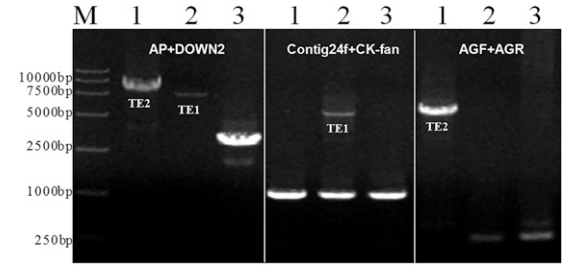

Figure 7 Detection of two retrotransposon insertions (TE1 and TE2) in primitive cultivated *G. barbadense* forms by PCR amplification using different primer combinations. Lanes 1, 2, and 3 represent different cotton varieties. 1: GB333; 2: GB217; and 3: GB102.

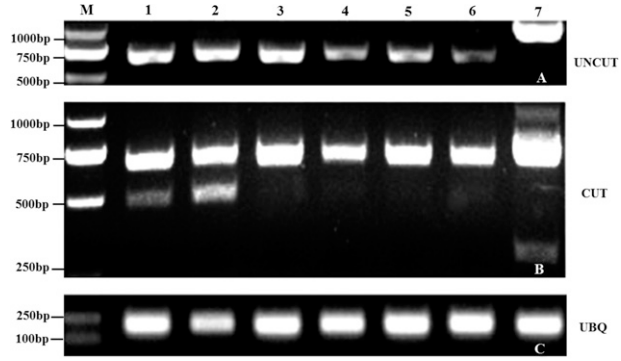

Figure 8 Allelic gene expression of *At-HD1* and *Dt-HD1* genes before (A) and after (B) digestion with *Xmn*I of the fragments amplified from stems of primitive cultivated forms by RT-PCR using SMF7/HDZIP-3R primers. Lines 1–6 indicate RT-PCR results from samples from GB102, GB656, GB825, GB1063, GB217, and GB154. Lane 7 is the control sample from the genomic DNA of NLD52-1. The housekeeping gene *UBQ* was used as the internal control to monitor RNA quality and PCR efficiency.

HD1 is a candidate gene for the T1 phenotype. Previously, *GhHD1* was postulated to be related to trichome formation and fiber initiation by gene silencing and gene expression (Walford *et al.* 2012). In the present research, a *Ty1-copia* retrotransposon was found inserted into *At-GbHD1* that blocks its expression. Sequence analysis and trichome phenotyping of a large population of *G. barbadense* genotypes showed that the trichome phenotype in each *G. barbadense* genotype was closely correlated with this TE insertion event. Mapping of the glabrous trichome phenotype and the TE insertions in *At-HD1* also provided evidence of a close association between TE insertion and trichome-less stems. It is extremely unlikely that different varieties of *G. barbadense* containing this retrotransposon in the same gene and at the same position also would lack trichomes by chance. Thus, we inferred that in *G. barbadense* the formation of trichomes largely depended on the presence of an intact *GbHD1* gene.

Different LTR retrotransposons in At-HD1 clarify the dispersal phases of G. barbadense

LTR retrotransposons, characterized by the presence of LTRs at the ends of elements such as *Ty1-copia* and *Gypsy* groups, are widely distributed in plant genomes. In cotton, *copia* retrotransposon elements have long been known to be widespread in tetraploid and diploid species (VanderWiel *et al.* 1993). Research in *G. barbadense* indicated that *Ty1-copia* retrotransposons are quite heterogeneous and could be grouped into 11 distinct families (Zaki 2005). The activation of retrotransposons, particularly the *Gypsy*-like retrotransposons, is suggested to be one of the major forces driving genome size variation of diploid cotton species (Hawkins *et al.* 2009). Also, it is thought that the activity of retrotransposons could drive cotton speciation by inserting into important genes that would result in dramatic changes in important evolutionary or agronomic traits (This *et al.* 2007; Tsuchiya and Eulgem 2013). Until now, direct evidence of cotton retrotransposons affecting phenotypic evolution had not been reported.

When we checked the geographic locations where these randomly selected primitive cultivated forms originated, we found that the accessions with different TE insertions were not randomly distributed (Figure 10). First, among 13 accessions whose places of origin could be traced, 5 originated from Peru, including 2 accessions with intact *At-HD1* and *Dt-HD1* genes (named as AD), 2 with TE2 (2565 bp downstream from ATG), and 1 with TE1 (1609 bp downstream from ATG). Two were from Brazil, and the other 6 were from six different countries (Table S1). Thus, Peru has more genetic diversity than other places and contains all kinds of primitive cultivated forms. This result strongly supports the hypothesis that Peru is the origin of modern *G. barbadense*, which was derived from early domesticated forms (Hutchinson *et al.* 1947). Second, the geographic distribution is much more dispersed for TE1, with six accessions scattered in Cuba, Guyana, Brazil, Argentina, and Peru, while accessions with the TE2 insertion originated from a neighboring region of Peru and Colombia. Using allozyme analysis, Percy and Wendel (1990) suggested a dispersal route of *G. barbadense* from west of the Andes, a primary domestication site, to northern South America (primary dispersal), Central America (secondary-stage dispersal), and a region including Argentina and Paraguay (a post-Columbian dispersal) (Hutchinson 1962; Brubaker *et al.* 1999; Wendel *et al.* 2010). Although our results could not further distinguish the three dispersal stages or reveal a detailed dispersal route, as described earlier, the distribution of accessions with different TE insertions suggests that the dispersal processes proposed by Percy and Wendel (1990) happened between the TE1 and TE2 insertion events. In addition, the diffusion of accessions with TE1 was limited to an area including South America, Central America, and northern North America but did not spread worldwide. Meanwhile, the TE2 insertion occurred quite recently and spread along the trans-Andean route in northern South America (through Colombia) into Central America, the Caribbean, and the Pacific (Percy and Wendel 1990) and then was further diffused to Egypt, forming "Egyptian cotton," and to the southern United States, forming "Pima cotton" (Figure 10). Therefore, the two TE insertions found in this research

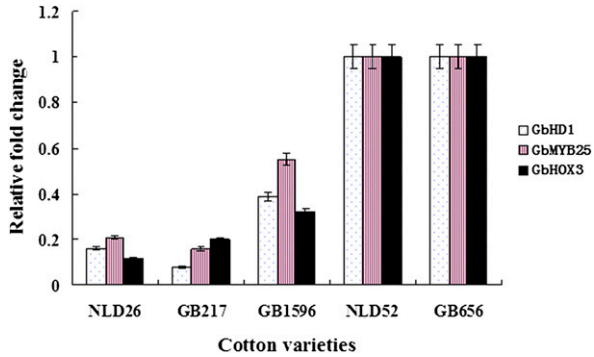

Figure 9 Quantitative qRT-PCR analysis of the gene expression of *GbHD1*, *GbMYB25*, and *GbHOX3* in five *G. barbadense* varieties/accessions.

represent important time points dividing the two dispersal phases in diffusion of *G. barbadense*.

One notable difference between the two major cultivated cotton species is their stem trichome phenotypes: *G. hirsutum* are almost all pubescent, while *G. barbadense* are almost all glabrous. We now know that this difference is closely associated with the disruption of the *At-HD1* gene in *G. barbadense*. Since *HD1* is also involved in fiber development (Walford *et al.* 2012), we speculated that this insertion could be one of the main reasons why most *G. barbadense* varieties have little or no fuzz fiber. The association between the TE insertion, the glabrous stem, and fuzz-less seeds is being investigated in our laboratory. The initial results suggest that the growth of fuzz fiber is also affected by the *GbHD1* gene. This might explain why more modern varieties have more TE insertions than primitive cultivated forms, because linter fiber is more easily detached from smooth-seeded varieties, making it favored by hand ginning in artificial selection (Hutchinson 1962; Percy and Wendel 1990; Brubaker *et al.* 1999; Wendel *et al.* 2010).

G. barbadense varieties carrying a TE insertion in At-HD1 are a valuable resource to study molecular mechanisms related to cotton trichome and fiber development

Forward genetics has been explored extensively to study molecular mechanisms controlling plant phenotypes, growth habits, and development. Some natural retroelement insertions in genes have been identified and used to understand the function of related genes (Hori *et al.* 2007). One of them, a retroelement insertion in the Medicago *FLOWERING LOCUS T* (*MtFTa1*) gene helped to determine its function and decipher possible related pathways that regulate flowering time (Jaudal *et al.* 2013). In cotton, such progress remains limited. This might be because fewer useful mutant genes have been cloned compared with model plants such as *Arabidopsis* and rice. The varieties with LTR-retrotransposon insertion in the *At-GbHD1* gene identified in this study will be very powerful resources to study molecular mechanisms related to epidermal cell growth, similar to *Arabidopsis* mutants.

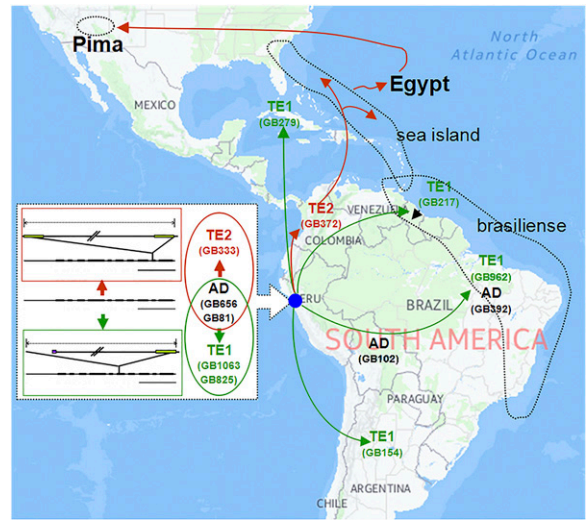

Figure 10 Dispersal route of primitive cultivated *G. barbadense* forms inferred by retrotransposon insertion events. Green arrows indicated the probable dispersal route of *G. barbadense* accessions during the stage between the TE1 and TE2 insertion events. Red arrows indicate the diffusion pathway of *G. barbadense* accessions after the TE2 insertion. A blue dot represents the point of origin of *G. barbadense* in Peru.

Real-time PCR results revealed that the retrotransposon insertion in *At-HD1* caused significant down-regulation in the expression of both *At-HD1* and *Dt-HD1* genes. In *G. barbadense*, *GbML1* (*GbHD1*) could bind to its own promoter region, leading to positive-feedback regulation of its own gene expression (Zhang *et al.* 2010). Our gene expression analysis partially confirmed this observation: the insertion in *At-HD1* caused significant reduction in the expression of both alleles of *HD1*. *GhMYB25* belongs to a *MIXTA* type MYB transcription factor family and controls cotton fiber initiation. *GhHOX3*, a homologous gene of *Arabidopsis HDG11/12*, regulates plant epidermal cell differentiation (Nakamura *et al.* 2006; Machado *et al.* 2009). Significantly reduced gene expression of *GbMYB25* and *GbHOX3* in the TE-inserted cottons provided further evidence that *GhHD1* is a critical upstream gene of *GhMYB25* (Bedon *et al.* 2014) and *GbHOX3*. Taken together, these results suggest that *At-HD1* is an important molecular switch that can activate *GhMYB25* and *GhHOX3* or that these three proteins may collaborate to regulate epidermal cell development into trichomes (Bedon *et al.* 2014).

Acknowledgments

We thank Zhixin Xie at Texas Tech University for his thorough revision of the manuscript and constructive suggestions. This work was supported jointly by the National Natural Science Foundation of China (grant nos. 31200909 and 31301372) and the Genetically Modified Organisms Breeding Major Projects (2009ZX05009-028B). Support also was received from the State Key Laboratory of Cotton Biology Open Fund (CB2014B04, CB2014A06) to J.R. and Y.J.

Literature Cited

Abe, M., H. Katsumata, Y. Komeda, and T. Takahashi, 2003 Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. Development 130: 635–643.

Bedon, F., L. Ziolkowski, S. A. Walford, E. S. Dennis, and D. J. Llewellyn, 2014 Members of the MYBMIXTA-like transcription factors may orchestrate the initiation of fiber development in cotton seeds. Front. Plant Sci. 5: 179.

Brubaker, C. L., A. H. Paterson, and J. F. Wendel, 1999 Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42: 184–203.

Cao, Y., Y. Jiang, M. Ding, S. He, H. Zhang *et al.*, 2015 Molecular characterization of a transcriptionally active Ty1/copia-like retrotransposon in Gossypium. Plant Cell Rep. 34: 1037–1047.

Cui, Y., S. Barampuram, M. G. Stacey, C. N. Hancock, S. Findley *et al.*, 2013 Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. Plant Physiol. 161: 36–47.

Desai, A., P. W. Chee, O. L. May, and A. H. Paterson, 2008 Correspondence of trichome mutations in diploid and tetraploid cottons. J. Hered. 99: 182–186.

Esch, J. J., M. Chen, M. Sanders, M. Hillestad, S. Ndkium *et al.*, 2003 A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in Arabidopsis. Development 130: 5885–5894.

Guan, X., J. J. Lee, M. Pang, X. Shi, D. M. Stelly *et al.*, 2011 Activation of *Arabidopsis* seed hair development by cotton fiber-related genes. PLoS One 6: e21301.

Hawkins, J. S., S. R. Proulx, R. A. Rapp, and J. F. Wendel, 2009 Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc. Natl. Acad. Sci. USA 106: 17811–17816.

He, S., Y. Zheng, A. Chen, M. Ding, L. Lin *et al.*, 2013 Converting restriction fragment length polymorphism to single-strand conformation polymorphism markers and its application in the fine mapping of a trichome gene in cotton. Plant Breeding **132:** 337–343.

Hori, Y., R. Fujimoto, Y. Sato, and T. Nishio, 2007 A novel wx mutation caused by insertion of a retrotransposon-like sequence in a glutinous cultivar of rice (Oryza sativa). Theor. Appl. Genet. 115: 217–224.

Hulskamp, M., 2004 Plant trichomes: a model for cell differentiation. Nat. Rev. Mol. Cell Biol. 5: 471–480.

Humphries, J. A., A. R. Walker, J. N. Timmis, and S. J. Orford, 2005 Two WD-repeat genes from cotton are functional homologues of the Arabidopsis thaliana TRANSPARENT TESTA GLABRA1 (TTG1) gene. Plant Mol. Biol. 57: 67–81.

Hutchinson, J. B., 1962 The history and relationships of the world's cotton. Endeavour 21: 5–15.

Hutchinson, J. B., R. A. Silow, and S. G. Stephens, 1947 *The Evolution of Gossypium and the Differentiation of the Cultivated Cottons*. Oxford University Press, Oxford, UK.

Jaudal, M., C. C. Yeoh, L. Zhang, C. Stockum, K. S. Mysore *et al.*, 2013 Retroelement insertions at the Medicago FTa1 locus in spring mutants eliminate vernalisation but not long-day requirements for early flowering. Plant J. 76: 580–591.

Lacape, J. M., and T. B. Nguyen, 2005 Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. J. Hered. 96: 441–444.

Larkin, J. C., M. L. Brown, and J. Schiefelbein, 2003 How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis. Annu. Rev. Plant Biol. 54: 403–430.

Larkin, J. C., D. G. Oppenheimer, A. M. Lloyd, E. T. Paparozzi, and M. D. Marks, 1994 Roles of the GLABROUS1 and TRANSPARENT TESTA GLABRA genes in Arabidopsis trichome development. Plant Cell 6: 1065–1076.

Lee, H. H., W. J. Lo, and K. B. Choo, 1992 Mutational analysis by a combined application of the multiple restriction fragment-single strand conformation polymorphism and the direct linear amplification DNA sequencing protocols. Anal Biochem 205: 289–293.

Lee, J. A., 1985 Revision of the genetics of the hairiness-smoothness system of Gossypium. J. Hered. 76: 123–126.

Machado, A., Y. Wu, Y. Yang, D. J. Llewellyn, and E. S. Dennis, 2009 The MYB transcription factor GhMYB25 regulates early fibre and trichome development. Plant J. 59: 52–62.

Marks, M. D., 1997 Molecular genetic analysis of trichome development in Arabidopsis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48: 137–163.

Marks, M. D., and J. J. Esch, 2003 Initiating inhibition: control of epidermal cell patterning in plants. EMBO Rep. 4: 24–25.

Nakamura, M., H. Katsumata, M. Abe, N. Yabe, Y. Komeda *et al.*, 2006 Characterization of the class IV homeodomain-leucine zipper gene family in Arabidopsis. Plant Physiol. 141: 1363–1375.

Paterson, A., C. Brubaker, and J. Wendel, 1993 A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. Plant Mol. Biol. Rep. 11: 122–127.

Paterson, A., J. Wendel, H. Gundlach, H. Guo, J. Jenkins *et al.*, 2012 Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492: 423–427.

Payne, T., J. Clement, D. Arnold, and A. Lloyd, 1999 Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in Nicotiana tabacum. Development 126: 671–682.

Percy, R. G., and J. F. Wendel, 1990 Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor. Appl. Genet. 79: 529–542.

Rong, J., C. Abbey, J. Bowers, C. Brubaker, C. Chang *et al.*, 2004 A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166: 389–417.

Rong, J., F. A. Feltus, V. N. Waghmare, G. J. Pierce, P. W. Chee *et al.*, 2007 Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176: 2577–2588.

Rong, J., G. Pierce, V. Waghmare, C. Rogers, A. Desai *et al.*, 2005 Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. Theor. Appl. Genet. 111: 1137–1146.

This, P., T. Lacombe, M. Cadle-Davidson, and C. L. Owens, 2007 Wine grape (Vitis vinifera L.) color associates with allelic variation in the domestication gene VvmybA1. Theor. Appl. Genet. 114: 723–730.

Tsuchiya, T., and T. Eulgem, 2013 An alternative polyadenylation mechanism coopted to the Arabidopsis RPP7 gene through intronic retrotransposon domestication. Proc. Natl. Acad. Sci. USA 110: E3535–3543.

VanderWiel, P. L., D. F. Voytas, and J. F. Wendel, 1993 Copia-like retrotransposable element evolution in diploid and polyploid cotton (Gossypium L.). J. Mol. Evol. 36: 429–447.

Walford, S.-A., Y. Wu, D. J. Llewellyn, and E. S. Dennis, 2012 Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. Plant J. 71: 464–478.

Wan, Q., H. Zhang, W. Ye, H. Wu, and T. Zhang, 2014 Genome-wide transcriptome profiling revealed cotton fuzz fiber development having a similar molecular model as Arabidopsis trichome. PLoS One 9: e97313.

Wan, Q., Z. Zhang, M. Hu, L. Chen, D. Liu *et al.*, 2007 T-1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (Earias spp.) resistance. Euphytica 158: 241–247.

Wang, S., J. W. Wang, N. Yu, C. H. Li, B. Luo *et al.*, 2004 Control of plant trichome development by a cotton fiber MYB gene. Plant Cell 16: 2323–2334.

Wendel, J. F., F. M. Bourland, and T. Seelana, 2010 The origin and evolution of *Gossypium*, pp. 1–18 in *Physiology of Cotton*, edited by J. M. Stewart, Springer Science, New York.

Wright, R. J., P. M. Thaxton, K. H. El-Zik, and A. H. Paterson, 1999 Molecular mapping of genes affecting pubescence of cotton. J. Hered. 90: 215–219.

Yang, C., and Z. Ye, 2013 Trichomes as models for studying plant cell differentiation. Cell. Mol. Life Sci. 70: 1937–1948.

Zaki, E. A., 2005 Ty1-copia group retrotransposon families in cultivated cottons G. barbadense L. identified by reverse transcriptase domain analysis. DNA Seq. 16: 288–294.

Zhang, F., A. Gonzalez, M. Z. Zhao, C. T. Payne, and A. Lloyd, 2003 A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130: 4859–4869.

Zhang, F., K. Zuo, J. Zhang, X. Liu, L. Zhang *et al.*, 2010 An L1 box binding protein, GbML1, interacts with GbMYB25 to control cotton fibre development. J. Exp. Bot. 61: 3599–3613.

Zhao, M., K. Morohashi, G. Hatlestad, E. Grotewold, and A. Lloyd, 2008 The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. Development 135: 1991–1999.

Zhou, L., K. Zheng, X. Wang, H. Tian, and S. Wang, 2014 Control of trichome formation in Arabidopsis by poplar single-repeat R3 MYB transcription factors. Front. Plant Sci. 5: 262.

Communicating editor: J. A. Birchler

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178236/-/DC1

The Hairless Stem Phenotype of Cotton (*Gossypium barbadense*) Is Linked to a *Copia*-Like Retrotransposon Insertion in a *Homeodomain-Leucine Zipper* Gene (*HD1*)

Mingquan Ding, Wuwei Ye, Lifeng Lin, Shae He, Xiongming Du, Aiqun Chen, Yuefen Cao, Yuan Qin, Fen Yang, Yurong Jiang, Hua Zhang, Xiyin Wang, Andrew H. Paterson, and Junkang Rong

DOI: 10.1534/genetics.115.178236

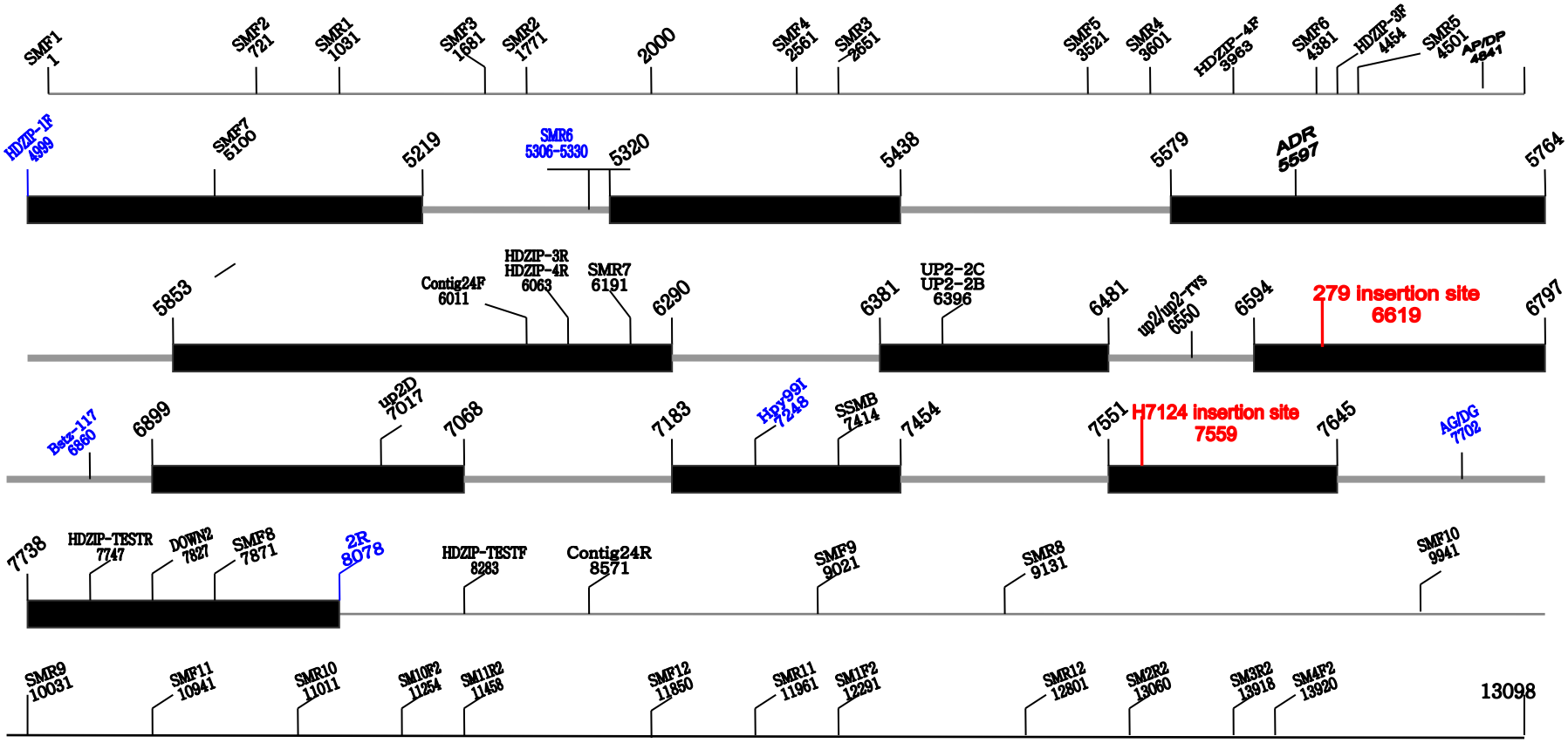

Figure S1. ***HD1*** **gene structure and positions of some important primers, TE insertions and sites recognized by enzymes**

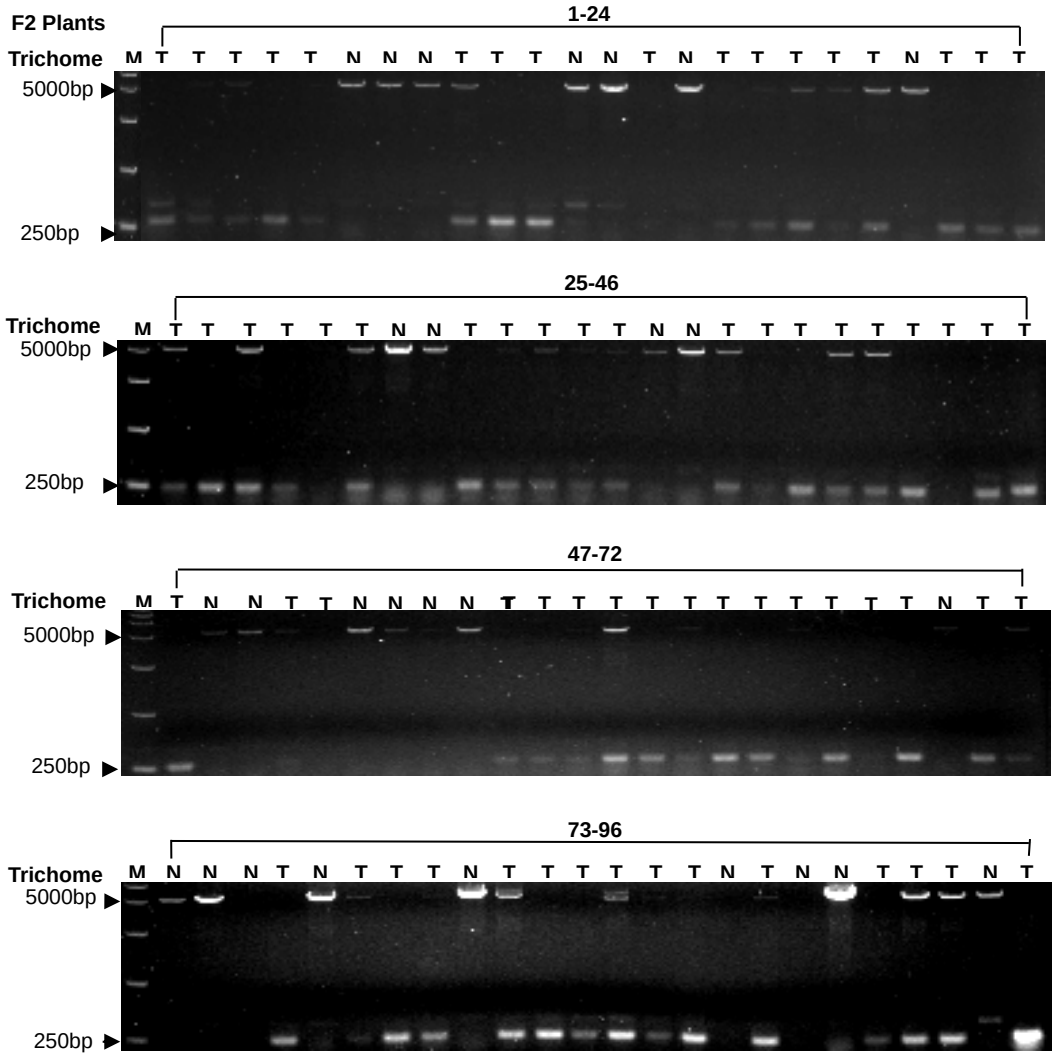

Figure S2. PCR amplification by using primers of AGF/AGR to illustrate the co-segregation of TE insertion and glabrous stems in a F_2 segregating population of T586 x Tu 75-37. T: the plants with trichome; N: the plants without trichome or few trichome.

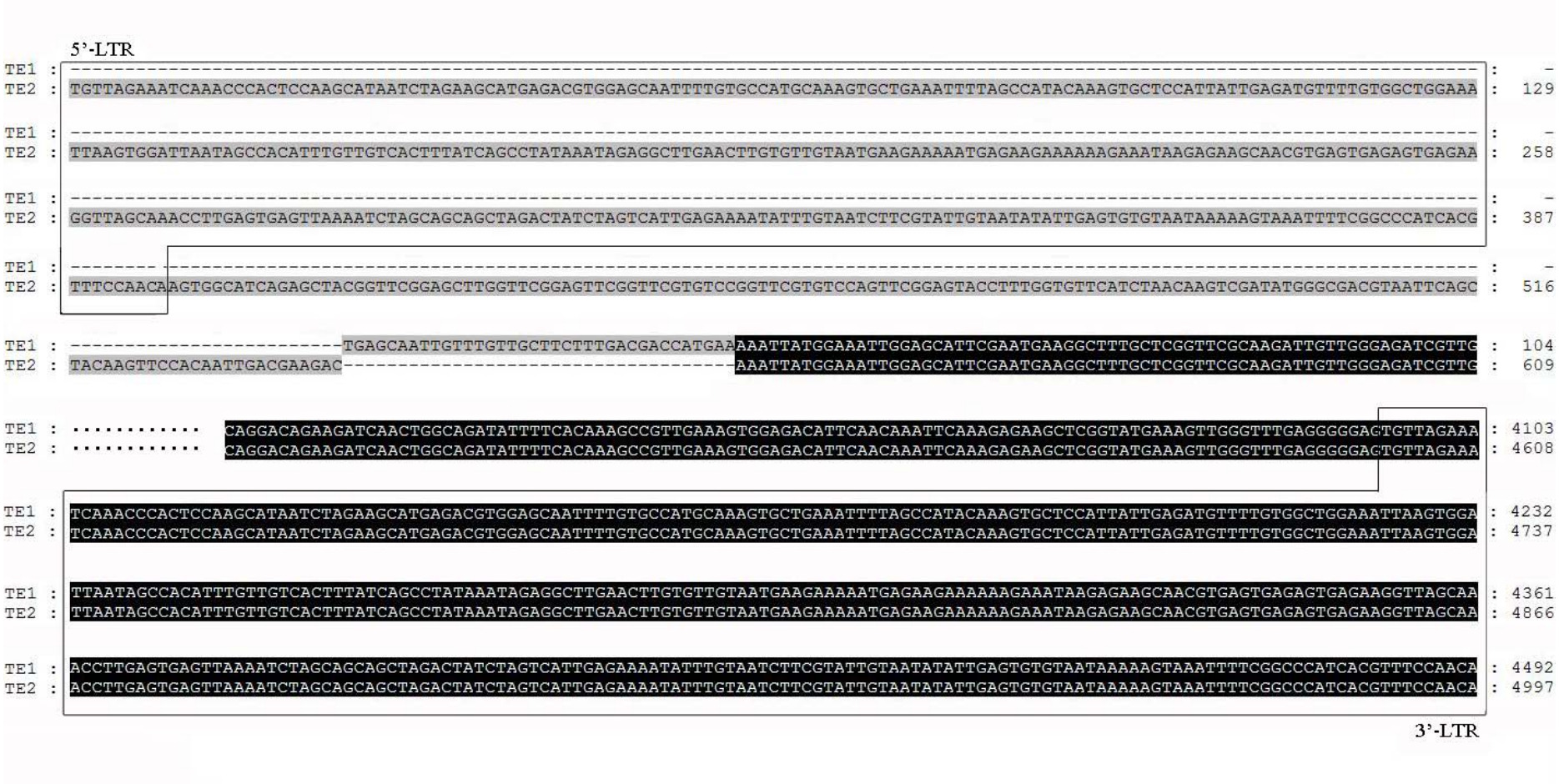

Figure S3. DNA sequence comparison between TE1 and TE2 retrotransposons. A 5'-LTR was missed and 34 additional bp were added in TE1 compared to TE2. Framed base pairs are LTR and the dots in the middle region of sequences represented the missed sequences.

Tables S1-S5

Available for download as Excel files at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178236/-/DC1

Table S1. Cotton materials, *HD1* amplification, TE insertion and trichome phenotype. Please see the attached supplemental file.

Table S2. Primers designed for different PCR amplification and sequencing. Please see the attached supplemental file.

Table S3. Genome identity of the *HD1* gene revealed by sequencing of PCR clones from different tetraploid cotton varieties. Please see the attached supplemental file.

Table S4. Genome identity of the *HD1* gene in *G. hirsutum* and *G. barbadense* varieties revealed by enzyme digestion and sequencing analysis. Please see the attached supplemental file.

Table S5. Genome identity of the *HD1* promoter revealed by sequencing PCR clones from different tetraploid cotton varieties. Please see the attached supplemental file.