Theor Appl Genet (2014) 127:167–177
DOI 10.1007/s00122-013-2208-4

ORIGINAL PAPER

Mapping and genomic targeting of the major leaf shape gene (*L*) in Upland cotton (*Gossypium hirsutum* L.)

Ryan J. Andres · Daryl T. Bowman · Baljinder Kaur · Vasu Kuraparthy

Received: 7 May 2013 / Accepted: 3 October 2013 / Published online: 25 October 2013

Abstract
Key message **A major leaf shape locus (L) was mapped with molecular markers and genomically targeted to a small region in the D-genome of cotton. By using expression analysis and candidate gene mapping, two *LMI1*-like genes are identified as possible candidates for leaf shape trait in cotton.**
Abstract Leaf shape in cotton is an important trait that influences yield, flowering rates, disease resistance, lint trash, and the efficacy of foliar chemical application. The leaves of okra leaf cotton display a significantly enhanced lobing pattern, as well as ectopic outgrowths along the lobe margins when compared with normal leaf cotton. These phenotypes are the hallmark characteristics of mutations in various known modifiers of leaf shape that culminate in the mis/over-expression of Class I *KNOX* genes. To better understand the molecular and genetic processes underlying leaf shape in cotton, a normal leaf accession (PI607650) was crossed to an okra leaf breeding line (NC05AZ21). An F_2 population of 236 individuals confirmed the incompletely dominant single gene nature of the okra leaf shape trait in *Gossypium hirsutum* L. Molecular mapping with simple sequence repeat markers localized the leaf shape gene to 5.4 cM interval in the distal region of the short arm of chromosome 15. Orthologous mapping of the closely linked markers with the sequenced diploid D-genome (*Gossypium raimondii*) tentatively resolved the leaf shape locus to a small genomic region. RT-PCR-based expression analysis and candidate gene mapping indicated that the okra leaf shape gene (L^o) in cotton might be an upstream regulator of Class I *KNOX* genes. The linked molecular markers and delineated genomic region in the sequenced diploid D-genome will assist in the future high-resolution mapping and map-based cloning of the leaf shape gene in cotton.

Communicated by G. J. Bryan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2208-4) contains supplementary material, which is available to authorized users.

R. J. Andres · D. T. Bowman · B. Kaur · V. Kuraparthy (✉)
Crop Science Department, North Carolina State University, Raleigh, NC 27695, USA
e-mail: vasu_kuraparhty@ncsu.edu

Introduction

Cotton (*Gossypium* spp.) is the most important source of natural fiber in the world. The majority of cultivated cottons (*Gossypium hirsutum* and *G. barbadense*) are allotetraploid species ($2n = 4x = 52$, AADD) formed by the chance hybridization of two diploid progenitor species (Wendel and Cronn 2003; Chen et al. 2007).

Nearly all Upland cotton (*G. hirsutum*) cultivars possess the normal or broad leaf type. Along with normal (*l*), numerous mutant leaf types make up an allelic series at the leaf shape locus including okra (L^o), sub-okra (L^U), and super-okra (L^S) (Jones 1982). Sea-island (L^e) is a fourth leaf shape mutant found commonly in cultivars of extra-long staple cotton (*G. barbadense*). However, it has not been definitively proven that Sea-island is a separate allele from sub-okra (Meredith 1984). Instead, the relatively minor differences between these two shapes could be the result of different alleles being present at numerous loci of minor effect throughout the genetic background of the two species.

The normal leaf shape in Upland cotton is broad and palmate with five readily observable, yet insipid lobes (Fig. 1). The okra leaf shape is characterized by reduced

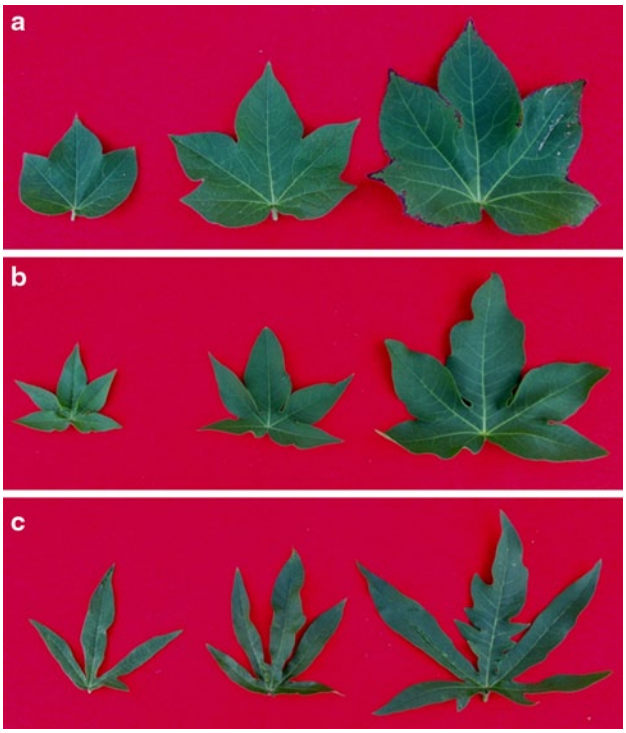

Fig. 1 Leaf shape phenotypes of cotton from *upper*, *middle*, and *lower* nodes (*left* to *right*). **a** Normal-shaped leaves of the landrace accession NC11-2100. **b** Leaf shape of F_1 heterozygote from NC11-2100 × NC05AZ21 that is intermediate between that of its parents. **c** Okra leaves of the breeding lines NC05AZ21. Note the increase in leaf lobing and abundance of ectopic outgrowths

photosynthetic area per leaf, a more pronounced lobing pattern with deeper sinuses, and ectopic outgrowths leading to abnormal leaf margins (Fig. 1). The appearance of ectopic outgrowths along the lobe margins is common, but position, shape, and number of these outgrowths appear random. Heterozygotes are intermediate between the two parental types and all three types can be easily categorized with brief visual observation (Fig. 1).

The advantages of okra leaf cultivars include reduced incidence of boll rot, accelerated flowering rates, early maturity, and increased resistance to whitefly and pink bollworm (Jones 1982). Okra leaf cultivars benefit from reduced lint trash, evapotranspiration, and chemical application rates due to their diminished leaf area. However, reduced leaf area leads to sub-optimal light capture and reduced photosynthetic rates, causing higher rates of boll shedding and a lower yield potential under optimal conditions (Jones 1982; Wells and Meredith 1986).

Currently, normal leaf cotton varieties dominate the market throughout the USA. Concerns over yield, as well as increased weed pressure, have prevented the widespread adoption of okra leaf cultivars despite their stated benefits. A better understanding of the genes and regulatory networks associated with leaf shape in cotton may allow the manipulation of leaf shape through breeding and/or biotechnology to combine the favorable aspects of both leaf shapes.

Despite its importance, the molecular and genetic control of leaf shape in cotton is poorly characterized. The factor underlying okra leaf shape is known to act during the initiation of the leaf primordia (Dolan and Poethig 1991). Therefore, the leaf shape in cotton is determined very early in development and remains constant throughout the subsequent growth and maturation of the leaf (Dolan and Poethig 1991). The increase in leaf complexity seen in okra leaf cotton is phenotypically similar to the ectopic expression of Class I *KNOTTED1-LIKE HOMEOBOX* (*KNOX1*) genes seen in other species (Lincoln et al. 1994; Janssen et al. 1998; Tanaka-Ueguchi et al. 1998; Rosin et al. 2003; Uchida et al. 2010). Briefly, re-establishment of *KNOX1* genes in developing leaf primordia cause a prolonged period of cell indeterminacy that results in increased leaf complexity. The presence of ectopic outgrowths is another hallmark of irregular *KNOX1* expression and the inability to probably regulate cell differentiation. Mutations to the upstream regulators of *KNOX1* can also result in similar phenotypes (Hay and Tsiantis 2010).

Currently, there is no genome sequence available for allotetraploid cotton. However, a draft sequence of the presumed D-genome donor diploid, *G. raimondii*, has recently been released (DOE Joint Genome Institute: Cotton D V2.0; www.phytozome.net). The availability of this draft genome sequence should help facilitate the fine mapping and cloning of genes in cotton by providing a physical map to integrate with constructed genetic maps. The gene annotation of this draft genome, as well as the available protein homology and gene ancestry tools, may allow the combination of physical and genetic maps to identify highly plausible candidate genes underlying phenotypes of interest.

Classical genetic analyses have placed the leaf shape locus on the short arm of chromosome 15 of the D subgenome of Upland cotton (Endrizzi and Brown 1964; Endrizzi and Kohel 1966). A leaf morphology quantitative trait loci (QTL) mapping study using restriction fragment length polymorphism (RFLP) markers and an interspecific *G. barbadense* × *G. hirsutum* population supports the location of the leaf shape locus on chromosome 15 (Jiang et al. 2000). However, a second *G. barbadense* × *G. hirsutum* interspecific leaf morphology QTL mapping study using simple sequence repeat (SSR) markers casts some doubt on these findings where 15 QTLs, 10 for lobe length, 13 for lobe width and 6 for lobe angle, were detected on 15 chromosomes or linkage groups (Song et al. 2005). The objective of the present study was to use an intraspecific *G. hirsutum* population to identify SSR or RFLP-based sequence-tagged site (STS) markers linked to the leaf shape

locus. These markers will serve as the basis for expanded efforts currently underway to fine map and clone the leaf shape gene in cotton.

Materials and methods

Plant material

The okra leaf breeding line NC05AZ21 was crossed to the normal leaf accession NC11-2100 in the summer of 2011 at the Central Crops Research Station in Clayton, NC. NC05AZ21 is a recently released okra leaf line with *Fusarium* wilt resistance and superior yield and lint percent (Kuraparthy et al. 2013). NC11-2100 (TX-2324; PI607650) was obtained from the USDA Cotton Germplasm collection in College Station, TX. NC11-2100 was selected as a normal leaf parent because this line was a photoperiod insensitive landrace that was expected to show better polymorphism for the markers used than an elite cotton line. A single F_1 plant was self-fertilized to obtain F_2 seeds during the winter of 2011–2012 utilizing an off-season nursery in Mexico. The F_2 population along with parental lines was grown in Clayton, NC, over the summer of 2012. Large-scale leaf samples were collected individually from all 236 F_2 plants for DNA extraction. Leaf shape phenotype was scored on all F_2 individuals at multiple points throughout the growing season. Plants were characterized as either normal, okra, or heterozygous based on their similarity to the images in Fig. 1.

SSR and STS marker analyses

Genomic DNA was isolated using a CTAB approach originally developed by Tel-Zur et al. (1999) and modified by Wendel (http://www.eeob.iastate.edu/faculty/WendelJ/dnaextraction.htm). Forty SSR markers mapped to chromosome 15 (Chr15) in a high-density consensus (HDC) genetic map between *G. hirsutum* and *G. barbadense* (Blenda et al. 2012) were selected and sequences were obtained from the Cotton Marker Database (http://www.cottondb.org/). Additionally, 23 RFLP markers placed on Chr15 by the HDC genetic map were converted into 40 STS markers. To develop STS markers from the RFLP markers, sequences of the restriction site region were obtained from the Cotton Genome Database (http://www.cottondb.org/) or from the Cotton Diversity Database (http://cotton.pgml.uga.edu/Cotton/index.aspx). Two candidate leaf shape genes present in the orthologous region of *G. raimondii* genome delineated by the flanking SSR markers were also used to develop STS markers. Primers were designed using the Primer 3 software (http://frodo.wi.mit.edu/). Multiple primers were designed from the same RFLP or candidate gene sequence when possible. The M13 tail sequence 5′-CACGACGTTGTAAAACGAC-3′ was added to the 5′ end of all forward primers to facilitate capillary-based gel electrophoresis (Schuelke 2000). All primer sequences were supplied by Integrated DNA Technologies (Coralville, IA, USA). These markers were evaluated on DNA isolated from the NC05AZ21 and NC11-2100 parents to select polymorphic markers to screen the F_2 population.

PCR reactions included 40 ng of genomic DNA, 1× reaction buffer with 15 mM $MgCl_2$, 0.48 mM dNTPs, 1 unit of Taq DNA polymerase, 0.96 μM forward primer, 7.2 μM reverse primer, and 7.2 μM M13 primer in a final volume of 12 μL. M13 primer was labeled with either HEX (hexachlorofluorescein) or 6-FAM (6-carboxyfluorescein) fluorescent tags. All primers were amplified using a Touchdown PCR protocol consisting of 5 min at 95 °C, followed by 15 cycles of 94 °C for 45 s, 65 °C for 45 s, and 72 °C for 1 min, then 25 cycles of 94 °C for 45 s, 50 °C for 45 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.

PCR products were run on a high-resolution 3 % agarose gel (GenePure HiRes, ISC BioExpress, Kaysville, UT, USA) as well as an ABI 3730XL capillary-based gel electrophoresis sequencer (Applied Biosystems, Carlsbad, CA, USA) to compare PCR product size. Size standard for all capillary-based gel electrophoresis was GeneScan–500 LIZ (Applied Biosystems, Carlsbad, CA, USA). Polymorphic markers were run on all 236 F_2 individuals and analyzed in a similar fashion utilizing 3 % agarose gels and the ABI 3730XL capillary sequencer. Data obtained from the ABI 3730XL were analyzed using GeneMarker V1.91 (SoftGenetics LLC, State College, PA, USA).

Linkage analysis

The mapping software JoinMap 4.1 (Kyazma BV., Wageningen, Netherlands) and an LOD score ≥ 6.0 were used to develop the linkage map for the leaf shape gene. Chi-square (χ^2) tests were performed to check for segregation distortion of markers and phenotypic data from expected Mendelian ratios.

Orthologous mapping of the leaf shape gene in the diploid D-genome

Linked SSR and STS markers from the present genetic map and markers in the genomic region of the L locus from a consensus map (Blenda et al. 2012) were used to establish the orthologous genomic region in the sequenced *G. raimondii* genome. SSR sequences were obtained from the Cotton Marker Database (http://www.cottonmarker.org/, while RFLP sequences were obtained from the Cotton Genome Database (http://www.cottondb.org/) and the Cotton Diversity Database (http://cotton.pgml.uga.edu/

Cotton/index.aspx). Both sequence types were BLAST searched against the draft sequence of the *G. raimondii* genome (DOE Joint Genome Institute: Cotton D V2.0; www.phytozome.net). Base pair positions of the highest scoring matches on *G. raimondii* Chr02, the homeolog of *G. hirsutum* Chr15, were used to order relevant SSR and RFLP markers on the genome sequence. A comparative map correlating Chr02 of the *G. raimondii* draft genome, Chr15 of the *G. hirsutum* HDC map, and Chr15 of the *G. hirsutum* leaf shape mapping population was assembled using the Strudel software (JHI Plant Bioinformatics, Dundee, Scotland). The output map was further redrawn to scale using MS PowerPoint to improve the resolution. The region encompassed by the closest proximal and distal markers, NAU2343 and Gh565 respectively, was annotated using the transcript, protein homolog, and gene ancestry features of the Phytozome's *G. raimondii* genome.

Gene expression analysis

The re-establishment of expression of Class I *KNOX* genes in developing leaf primordia is known to cause increases in leaf complexity and lobing across a wide range of species (Uchida et al. 2010). Therefore, the role of *KNOX1* genes in leaf shape variation in cotton was studied by semi-quantitative RT-PCR analysis. The four known *Arabidopsis* Class I *KNOX* genes (*STM*, *KNAT1*, *KNAT2*, and *KNAT6*) were used as queries in BLAST searches of the Phytozome *G. raimondii* genome. High-scoring matches for *KNOX1* genes in the *G. raimondii* genome were then BLAST searched against the DFCI cotton gene index to identify expressed sequence tags (ESTs) with high similarity to putative Class I *KNOX* genes in *G. hirsutum* (Table 4). Sequences of high-scoring ESTs and tentative contigs (TCs) were selected and primers were designed using the Primer 3 software. Primer sequences were supplied by Integrated DNA Technologies (Coralville, IA, USA).

RNA samples were collected from ~1 m tall cotton plants grown under greenhouse conditions. Varieties were the normal leaf line Texas Marker-1 (TM-1) and the okra leaf line NC05AZ21. TM-1 is used as a normal leaf sample instead of NC11-2100 because this line is a genetic standard with the most genomic resources developed in cotton. Also, TM-1 showed a similar growth pattern as NC05AZ21 for collecting the correct stage-specific tissue for analysis. NC11-2100 showed delayed flowering with extended juvenile/adult phases. Furthermore, at the time of expression analysis TM-1 was highly inbred, while NC11-2100 was subjected to selfing to reduce possible residual heterozygosity. Four different tissue types were collected from each variety: shoot apical meristem (SAM)-enriched stem, leaf primordia, young leaf, and mature leaf. SAM-enriched stem tissue consisted of the top 1–2 cm of the main stem with any developing leaves removed. Primordial tissues were developing leaves no greater than 2 cm in length harvested off the main stem. Young leaf tissue was taken from the center of an expanding leaf, while mature leaf tissue was taken from a central section of a fully expanded leaf. Samples were collected, immediately frozen in liquid nitrogen, and stored at −80 °C until grinding. RNA was isolated from 100 mg samples using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) and quantified using a NanoDrop 2000 UV–Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Quantification results were used to dilute the RNA concentration to approximately 125 ng/μL and the dilutions were confirmed using spectrophotometry.

RNA samples were then converted into cDNA using the ImProm-II™ Reverse Transcription System (Promega Corporation, Fitchburg, WI, USA). The cDNA samples were then used as templates in PCR reactions in a semi-quantitative approach to measure mRNA copy number and, by extension, transcription levels of *KNOX1* genes. As the leaf shape gene is mapped to the D-genome, its A-genome homeoloci are assumed to be non-polymorphic for the two samples used and no attempts were made to see the sub-genome specific expression patterns. A volume of 1.0 μL of the reverse transcription reaction was combined with one unit of standard 10× reaction buffer with 15 mM $MgCl_2$ (Apex™ Bioresearch Products, San Diego, CA, USA), dNTP mix (0.2 mM), forward and reverse primers (1 μM each), and 1 unit of *Taq* DNA polymerase (Apex™ Bioresearch Products, San Diego, CA, USA) in 30 μL reactions. The PCR protocol consisted of 2 min at 94 °C, 40 cycles of 1 min at 94 °C, 1 min at 60 °C, and 2 min at 72 °C followed by 5 min at 72 °C. PCR products were then separated on a 2 % agarose gel and visualized. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a positive control.

Candidate gene mapping

Single genes known to have a pronounced and significant impact on leaf shape and complexity in Arabidopsis were identified in the literature. Full-length cDNA sequences of these genes were obtained from The Arabidopsis Information Resource (TAIR; www.arabidopsis.org). The cDNA sequences were then BLAST searched against the *G. raimondii* genome (http://www.phytozome.net/) to analyze the gene ancestry of the corresponding *Arabidopsis* transcript for orthologous putative genes in the *G. raimondii* genome. Transcript sequences of putative *G. raimondii* genes whose predicted protein sequence was most similar to a known leaf shape modification protein were also BLAST searched against the DFCI cotton gene index (http://compbio.dfci.harvard.edu/cgi-bin/tgi/Blast/index.cgi) to identify high-scoring expressed sequence tags (ESTs) and tentative contigs (TCs) in tetraploid cotton.

Table 1 Primer sequence and allele sizes for SSR and STS markers associated with the leaf shape gene in Upland cotton

Marker	Forward sequence 5′→3′	Reverse sequence 5′→3′	NC05AZ21 Allele size (bp)	NC11-2100 Allele size (bp)
BNL2440	TGTTAAGCATACATTAGTTTCACTCG	CCGGCACCACAAAAGTAAAT	245	219
Gh565	AAAGACTCGGGTACCACCTAATC	GTCCTTCTCATTTATCTGAATTCACC	154	129
NAU2343	GCTTTGCTTTGGAATGAGAT	ATACTGCAACCCCTCACACT	266	280
DPL0402	TTACAAGCGAATTTAGGATGCC	ACTTGAGGTGCAATTGACGAG	260–263	272–275
TMB1664	AAATACCGGAACTTGATTGGG	AATTTGGTTGGGTTTCCACA	190–198	198–204
13-LS-195	ACCTTTTACGCAGGTGATGG	TCGGATATAGTCGTTTCCTGCT	180	172, 180

Results

Genetics of the leaf shape trait

The phenotypic ratio of individuals within the population fit the expected 1:2:1 segregation ratio ($\chi^2 = 0.31$, $p = 0.86$). This confirms the single gene nature of okra leaf shape. Furthermore, the appearance of the intermediate leaf shape phenotype in the F_1 hybrid relative to the parental types indicates that alleles of the okra and normal leaf shape trait show incompletely dominant phenotypic expression in the heterzygote.

Molecular mapping of the leaf shape locus (*L*)

Sixty-three SSR and RFLP markers genetically mapped on Chr15 in *G. hirsutum* were used to survey the polymorphism between the two parents, NC05AZ21 and NC11-2100. From the 31 SSR markers that amplified specifically in both parents, 5 were polymorphic (Table 1). Of the polymorphic SSRs, four (BNL2440, Gh565, DPL0402, and TMB1664) were co-dominant and one (NAU2343) was dominant for the NC11-2100 parent. From the 40 PCR-based sequence-tagged site (STS) markers developed from the 23 RFLP sequences, only 24 amplified specifically and none were polymorphic. Of the 27 STS markers developed based on candidate genes *Gorai.002G244000* and *Gorai.002G244200*, showing protein similarity to Arabidopsis *LATE MERISTEM IDENTITY1* (*LMI1*) gene, 1 was polymorphic. STS marker 13-LS-195 was dominant for NC11-2100 parent. The five polymorphic SSR markers and one STS marker were mapped in the F_2 population of 236 individuals (Figs. 2, 3).

Two of the polymorphic SSR markers, Gh565 and NAU2343, showed tight linkage with the leaf shape locus with Gh565 mapping 2.6 cM distally and NAU2343 mapping 2.8 cM proximally to the leaf shape locus. In the consensus genetic map the flanking markers are localized toward the telomeric region on chromosome 15 of Upland cotton (Blenda et al. 2012). Thus, the leaf shape gene (*L*) was mapped to a 5.4 cM region near the telomere on the short arm of Chromosome 15 in *G. hirsutum*.

There were only two publicly available markers between Gh565 and NAU2343 in the high-density consensus (HDC) map, the RFLP marker A1485 and the SSR cluster CLU1520 (Blenda et al. 2012). No polymorphism between the two parents was detected in the single STS marker designed from the A1485 sequence. Similarly, no polymorphisms were detected in any of the five SSR markers that comprise CLU1520 (MGHES32, HAU2398, NAU3938, MON_CGR5326, and MON_CGR5902).

Genomic targeting of leaf shape locus in the diploid D-genome

Simple sequence repeat and RFLP marker sequences of interest with regard to the leaf shape locus were blasted against the *G. raimondii* genome to establish this orthologous region (Table 2). All markers that were used in the BLAST search showed significant homology to and a high-scoring match within a ~2 Mb region near the telomere of *G. raimondii* Chr02. The order of the markers and relative distances between them in the *G. raimondii* physical map was fairly consistent with that of the available genetic map, although minor discrepancies did exist (Fig. 3).

In the *G. raimondii* genome, the orthologous region between the most closely linked SSR markers, Gh565 and NAU2343, spanned a physical distance of 337 kb. Annotation of this region identified 34 putative genes, of which 3 carry functional domains associated with genes known to affect leaf shape in other species (Table 3). These genes include one containing an ankyrin repeat (Ha et al. 2004) and two containing homeodomain leucine-zipper (HD-Zip) motifs (Saddic et al. 2006). Of the remaining 31 genes, 4 carry no known functional domain while the rest carry domains associated with functions that do not include leaf shape in model plants. The highest scoring protein homologs in *Arabidopsis* were obtained for all three genes that possessed domains associated with variation in leaf shape and only two showed substantial homology to genes implicated in leaf shape (Table 3). These two genes, *Gorai.002G244000* and *Gorai.002G244200*, are only 31.5 kb apart and separated by a single gene, a serine/

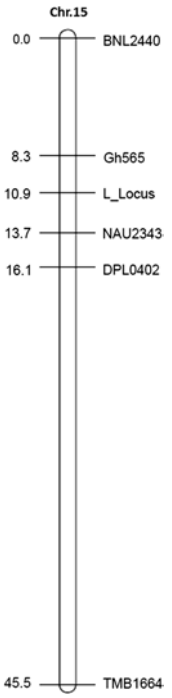

Fig. 2 Molecular markers associated with the leaf shape (L) locus on *G. hirsutum* chromosome 15. Genetic distance in centiMorgan is on the *left* with marker name on the *right*. Orientation of the map is in agreement with Blenda et al. (2012) with the *top* of the map toward the telomere

threonine protein kinase. The predicted proteins from these two genes were highly similar (71.2 % amino acid identity) and both showed high sequence similarity to the *Arabidopsis* meristem identity regulator *LATE MERISTEM IDENTITY 1* (*LMI1*), also known as *ATHB51*.

Candidate gene mapping

Analysis of 26 different gene families known to influence leaf morphogenesis in *Arabidopsis* identified 90 homologous members in the *G. raimondii* genome (Table S1). Of these 90 candidate genes, only 9 mapped to Chr02 of the annotated physical sequence of *G. raimondii*. Four of these nine, *Gorai.002G067300.1*, *Gorai.002G113300.1*, *Gorai.002G038900.1*, and *Gorai.002G064500.1*, mapped to the long arm of Chr02 at physical positions 7.85, 16.00, 3.11, and 7.55 Mb, respectively, which was the opposite arm of *G. raimondii* Chr02 proposed as carrying the leaf shape locus. These genes were most similar, based on percent amino acid similarity, to the *Arabidopsis* leaf shape modification pathway genes *CUC2*, *CUC3*, *GA2ox1*, and *TCP4*, respectively (Table S1).

Of the remaining five, only *Gorai.002G244000.1* and *Gorai.002G244200.1* were between the 337 kb physical region delineated by the flanking markers Gh565 and NAU2343 on Chr02 of *G. raimondii* (physical position 60.69-61.03 Mb). As stated above, *Gorai.002G244000.1* and *Gorai.002G244200.1* were very similar to each other and in close physical proximity and were most similar to the known leaf shape modification gene *LMI1/ATHB51*. The remaining three candidate genes on *G. raimondii* Chr02 lie relatively close to, but outside of, the leaf shape locus region established here. These genes include *Gorai.002G226700.1* (physical position 58.35 Mb), *Gorai.002G255800.1* (61.80 Mb), and *Gorai.002G222400.1* (57.71 Mb) which were most similar to the leaf shape modification pathway genes *BOP2*, *BLR*, and *COMT1*, respectively (Table S1). None of the candidate Class I *KNOX* genes were mapped in the genomic region delineated by the flanking markers of the leaf shape gene (Table S1).

Based on both the annotation of genes in the delineated orthologous genomic region and in silico mapping of the Arabidopsis leaf shape candidate genes in the diploid D-genome, only two genes *Gorai.002G244000* and *Gorai.002G244200* were found to be the likely candidates for leaf shape locus. The STS marker 13-LS-195 developed from candidate gene *Gorai.002G244000* showed co-segregation with leaf shape phenotype in the F_2 population (Fig. 3), suggesting marker 13-LS-195 was tightly linked to the leaf shape locus or that the *Gorai.002G244000* was a possible candidate gene for leaf shape trait in cotton.

Gene expression analysis

The four known *Arabidopsis* Class I *KNOX* genes (*STM*, *KNAT1*, *KNAT2*, and *KNAT6*) when used as queries in the BLAST searches of *G. raimondii* genome and cotton gene index resulted in the identification of ESTs with high similarity to putative Class I *KNOX* genes in *G. hirsutum* (Table 4). Figure 4 shows the semi-quantitative RT-PCR-based gene expression pattern of some of these putative *G. hirsutum* Class I *KNOX* orthologs across four different tissue types (SAM-enriched stem, developing primordia, young leaf, and mature leaf) and two different lines: the okra leaf NC05AZ21 and TM-1, a genetic standard with normal leaf shape. All five ESTs showed strong and uniform expression in the SAM-enriched stem of both lines (Fig. 4). This is consistent with the conserved and required function of Class I *KNOX* genes in the maintenance of the SAM and the pluripotent nature of its stem cells (Uchida et al. 2010). However, all five ESTs also showed strong expression in the developing leaf primordia of the okra leaf NC05AZ21. This expression was generally only slightly weaker than that seen in the corresponding NC05AZ21 SAM-enriched stem samples. Conversely, the expression of the putative *KNOX1* orthologs were conspicuously absent in the developing leaf primordia of the normal leaf TM-1 with the exception of DW236483. While the expression of the EST DW236483 in

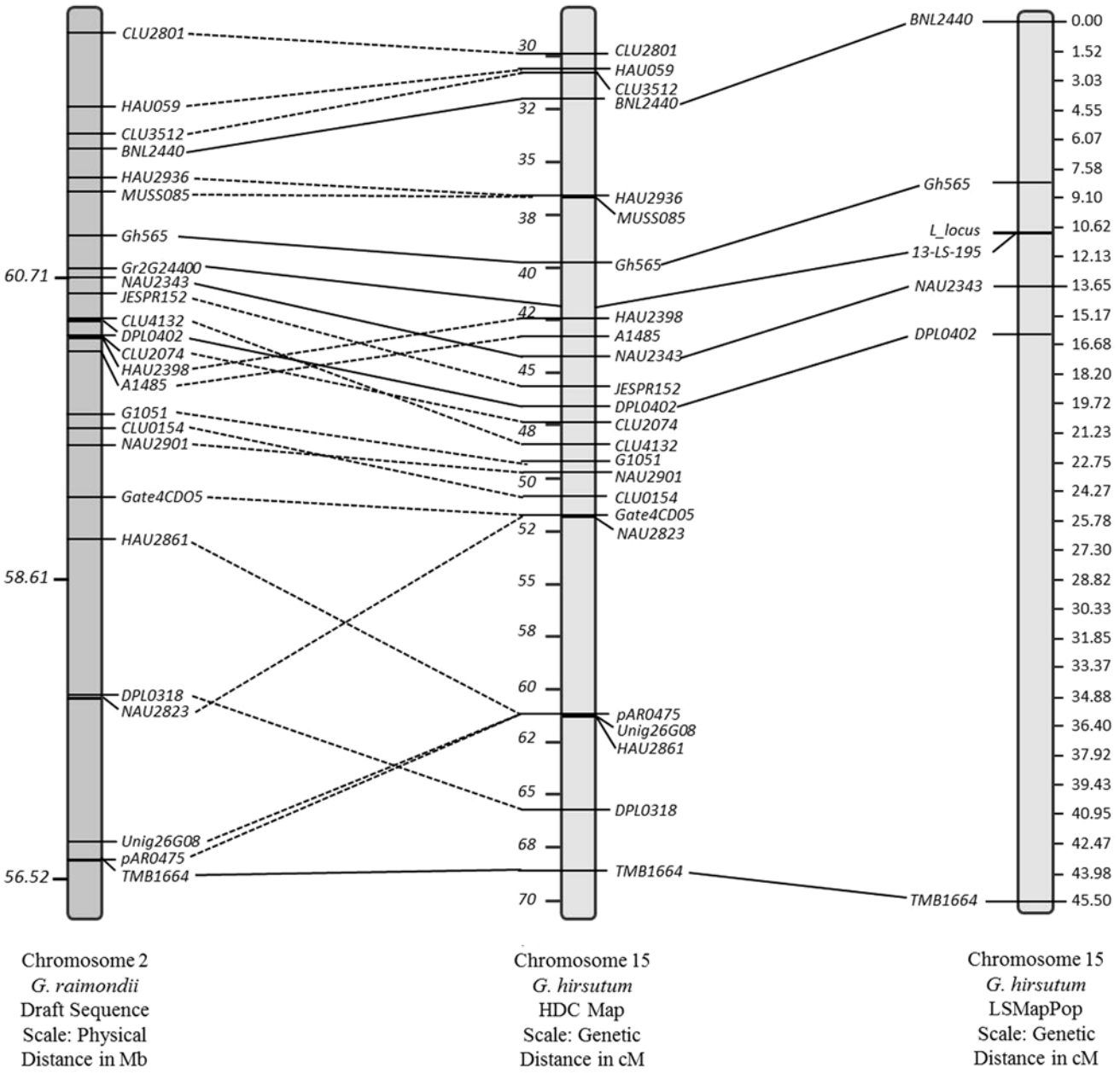

Fig. 3 Orthologous relationship between Chr02 of the *G. raimondii* draft sequence (DOE Joint Genome Institute: Cotton D V2.0, www.phytozome.net), Chr15 of the *G. hirsutum* high-density consensus (HDC) map (Blenda et al. 2012), and Chr15 of the *G. hirsutum* map of the present study (LSMapPop) at the region encompassing leaf shape locus of Upland cotton. The leaf shape gene (*L*) maps between SSR markers NAU2343 and Gh565 in the LSMapPop. The *LMI1* candidate gene (*Gorai.002G244200.1*) based STS marker 13-LS-195 co-segregates with leaf shape gene in the LSMapPop. Markers that showed conserved order and relative distances across all three maps are indicated with *solid lines*

NC05AZ21 persisted into both the young and mature leaf samples, expression was not seen past the developing primordia stage in TM-1. Expression of both TC273462 and TC254362 was detected readily in the mature leaf stage of NC05AZ21. Expression of a corresponding strength was not seen in the normal leaf TM-1 at this stage, although very faint bands were visible. Therefore, with the five putative *KNOX1* EST orthologs displayed here, all show expression that persists later into leaf development in an okra leaf cotton line when compared with a normal leaf line.

Discussion

Okra leaf shape and its allelic variants in cotton have been widely assumed to be under single gene control that

Table 2 Establishing the orthologous relationship between Chr15 of *G. hirsutum* and Chr02 of *G. raimondii* at the region encompassing the leaf shape locus. SSR and RFLP markers linked to the leaf shape locus from the current study and Blenda et al. (2012) were BLAST searched against draft *G. raimondii* genome sequence

Markers from Chr15	Type	Orthologous relationship with *G. raimondii* genome (BLASTn)		
		Score	E-value	Physical position on Chr02 (bp)
BNL2440	SSR	257.4	9.4e-67	61,637,363–61,637,509
Gh565	SSR	342.1	3.9e-92	61,031,975–61,032,199
NAU2343	SSR	205.1	7.3e-51	60,694,487–60,699,332
DPL0402	SSR	291.6	2.8e-77	60,397,282–60,397,471
NAU2814 (CLU4132)	SSR	994.9	0	60,385,657–60,386,318
HAU2398 (CLU1520)	SSR	625.3	5.9e-177	60,297,166–60,301,417
A1485	RFLP	998.5	0	60,213,673–60,214,255
G1051	RFLP	742.5	0	59,761,625–59,761,830
TMB1664	SSR	224.9	1e-57	59,715,822–59,715,970

is considered incompletely dominant in nature (Rahman et al. 2005). An F_2 population derived from the cross of an okra leaf breeding line to a normal leaf accession confirms previous reports that leaf shape is controlled by a single nuclear gene and alleles of the okra and normal leaf shape trait show incompletely dominant phenotypic expression in the heterzygote. SSR and STS marker-based molecular genetic mapping not only confirms the previous efforts on mapping major leaf shape loci (Endrizzi and Brown 1964; Endrizzi and Kohel 1966) and QTLs (Jiang et al. 2000) in tetraploid cotton, but also localizes the leaf shape locus to a small genomic region toward the telomeric end on Chr15. A previous mapping report showed that leaf morphology displayed a quantitative inheritance pattern in an interspecific mapping population of okra leaf *G. hirsutum* × normal leaf *G. barbadense* with the major effect QTL mapping to Chr15 (Jiang et al. 2000).

Five polymorphic markers showed varying degrees of linkage to the leaf shape locus on Chr15. The two closest markers, Gh565 and NAU2343, mapped 2.6 cM distally and 2.8 cM proximally, respectively. Only two other publicly available markers mapped between Gh565 and NAU2343, neither of which were polymorphic. Comparison of the *G. hirsutum* marker sequences to the *G. raimondii* draft genome indicates that the physical distance of the 5.4 cM region between Gh565 and NAU2343 is only ~337 kb and contains 34 putative genes. Thus, the leaf shape gene is localized to a 337 kb region on chromosome 15 of tetraploid cotton. Co-segregation of candidate gene *Gorai.002G244000* based STS marker 13-LS-195 with the leaf shape phenotype confirmed the orthologous region of leaf shape gene between *G hirsutum* and *G. raimondii*. Further, the genetic order of the co-segregating STS marker and flanking SSR markers of the leaf shape gene in *G. hirsutum* was consistent with the physical order of their orthologous sequences in *G. raimondii* genome, suggesting that major rearrangements might not exist in the region delineated by the flanking SSR markers.

The leaf shape locus maps to between 60.70 and 61.03 Mb on *G. raimondii*'s Chr02, the homeolog of *G. hirsutum* Chr15 (Blenda et al. 2012). The end of *G. raimondii* Chr02 is projected at 62.8 Mb, indicating that the leaf shape locus lies near the telomere and that its encompassing region is likely subject to the associated high rates of recombination. Recombination was found to be exceptionally high in cotton (Rong et al. 2004) and most of the recombination was also reported to be localized to the gene-rich regions toward the distal or telomeric ends in plants (Werner et al. 1992; Gill et al. 1996). Therefore, the large genetic distance, but small physical distance, may be a result of the region falling in the recombination-rich telomeric region of Chr15. As recombination is a prerequisite for the map-based gene cloning of traits in plants (Martin et al. 1993; Faris et al. 2003; Yan et al. 2004; Lin et al. 2012), the mapping of the leaf shape locus to a gene and recombination-rich region of the cotton genome should make the genetic dissection of the leaf shape trait in tetraploid cotton amenable to a map-based cloning approach.

Preliminary investigations have shown that okra leaf cotton has prolonged and elevated expression levels of Class I *KNOTTED1-LIKE HOMEOBOX* (*KNOX1*) orthologs during the early stages of leaf development compared to normal leaf cotton. The okra leaf NC05AZ21 showed expression of multiple putative *G. hirsutum KNOX1* genes later in leaf development than the normal leaf genetic standard TM-1. This differential expression was most noticeable during the leaf primordia stage of development. The ectopic expression of *KNOX1* genes at this stage has been shown to result in increased leaf complexity in numerous species, including the increased leaf lobing and ectopic outgrowth characteristics of okra leaf cotton (Lincoln et al. 1994; Janssen et al. 1998; Tanaka-Ueguchi et al. 1998; Rosin et al. 2003; Uchida et al. 2010). Combined with previous findings that the mechanism underlying okra leaf shape in cotton acts early on in development (Dolan an Poethig 1991), it appears that modifications in the expression pattern of *KNOX1* genes are involved in the formation of the various leaf shapes comprising the classical Chr15 leaf shape locus. However, no Class I *KNOX* genes map to the region of the L locus in the orthologous D-genome sequence. Therefore, based on the combined results of expression analysis as well as candidate gene mapping, leaf shape variation at the L

Table 3 Annotated gene sequences in the genomic region of *G. raimondii* Chr02 delineated by markers closely linked to leaf shape gene and their homology with *Arabidopsis* genes. Highest scoring EST/TC from the Cotton Gene Index is also included

G. raimondii		*G. hirsutum* gene index		Homology with *Arabidopsis*		
Locus name	Physical location	EST/TC	E-value	Protein homolog	Similarity (%)	Putative function
Gorai.002G246200	61027658–61032210	TC229763	4.4E-90	AT3G53190	66.4	Pectate lyase
Gorai.002G246100	61015799–61018395	TC239893	7.0E-86	ACA7	63.4	Carbonic anhydrase
Gorai.002G246000	61009256–61011927	TC239893	1.1E-200	ACA4	53.1	Carbonic anhydrase
Gorai.002G245900	61004759–61006420	TC233148	2.1E-104	TIP1;1	88.4	Aquaporin transporter
Gorai.002G245800	60992779–60993648	TC253115	1.8E-30	AT2G28200	55.0	Unknown
Gorai.002G245700	60987773–60990298	TC262237	0.56	AT3G09890	59.3	Ankyrin repeat
Gorai.002G245600	60979680–60986732	TC253684	1.9E-220	AT3G09920	83.3	Kinase
Gorai.002G245500	60950960–60959459	TC274111	5.1E-179	AT2G36720	55.9	PH-D finger, protein binding
Gorai.002G245400	60934232–60942403	TC274111	6.0E-162	AT2G36720	57.1	PH-D finger, protein binding
Gorai.002G245300	60918534–60922736	ES851936	3.2E-184	AT2G36670	75.4	Aspartyl protease
Gorai.002G245200	60915361–60918548	CO494244	2.8E-54	AT3G10060	68.2	Isomerase, protein folding
Gorai.002G245100	60909449–60912281	TC265610	5.8E-155	AT5G03810	72.7	Hydrolase, lipid metabolism
Gorai.002G245000	60902520–60910968	DR462456	6.7E-99	UPL7	73.4	Ubiquitin transferase
Gorai.002G244900	60889583–60893220	TC232237	5.9E-174	UGE5	89.9	Epimerase, steroid biosynthesis
Gorai.002G244800	60882963–60886378	TC237927	0	ADSS	83.6	Synthetase, nucleotide synthesis
Gorai.002G244700	60878611–60881268	CO089290	5.7E-192	EMB166	68.0	Pentatricopeptide repeat
Gorai.002G244600	60867000–60871561	TC244038	1.4E-177	AT5G03795	62.8	Exostosin
Gorai.002G244500	60863949–60866523	TC269303	6.1E-221	AT5G03795	43.7	Exostosin
Gorai.002G244400	60856168–60863202	CO090499	1.1E-173	SRPK4	64.0	Ser/Thr protein kinase
Gorai.002G244300	60852082–60854077	TC239774	3.6E-167	STV1	58.5	Ribosomal protein L24e
Gorai.002G244200	60848207–60849955	TC252073	9.8E-166	LMI1	55.5	HD-Zip Transcription Factor*
Gorai.002G244100	60818675–60821101	TC277482	1.5E-39	RLK1	44.6	Ser/Thr protein kinase
Gorai.002G244000	60816695–60817565	TC252073	2.0E-66	LMI1	42.4	HD-Zip Transcription Factor*
Gorai.002G243900	60802515–60804472	TC252850	1.7E-120	AT1G76870	47.8	Unknown
Gorai.002G243800	60793342–60798916	TC239482	1.2E-275	AT3G52990	95.4	Pyruvate kinase
Gorai.002G243700	60784258–60789349	TC244120	0	AT3G52990	89.4	Pyruvate kinase
Gorai.002G243600	60775425–60777957	ES804942	3.2E-215	AT2G36570	75.2	Ser/Thr protein kinase
Gorai.002G243500	60763514–60767171	TC253760	5.1E-131	DUF668	63.9	Unknown
Gorai.002G243400	60739751–60745494	TC236797	1.2E-273	RAT4	89.6	Glycosyl transferase
Gorai.002G243300	60734924–60737571	ES803238	4.5E-220	AT5G55840	43.1	Pentatricopeptide repeat
Gorai.002G243200	60733892–60734923	TC238398	1.5E-188	AT3G52960	74.0	Redoxin
Gorai.002G243100	60725536–60729527	TC259411	0	LOS2	94.4	Enolase
Gorai.002G243000	60721265–60724366	TC274400	3.5E-243	HDT3	20.8	Unknown
Gorai.002G242900	60700190–60718341	DR452531	3.8E-111	FK	78.1	Ergosterol biosynthesis

Candidate leaf shape genes are indicated with an asterisk

locus is likely not the result of a Class I *KNOX* gene itself, but may be due to an upstream regulator of *KNOX1*. Two genes, *Gorai.002G244000.1* and *Gorai.002G244200.1,* whose coding sequence predicts proteins with high amino acid similarity to the *Arabidopsis* meristem identity regulator *LMI1* are present in a 5.4 cM region shown to contain the Chr15 leaf shape locus. An STS marker developed based on *Gorai.002G244000.1* showed co-segregation with leaf shape phenotype in the F_2 population used in the current study. Further, loss-of-function mutations to *LMI1* in *Arabidopsis* have been shown to lead to elevated expression levels of a *KNOX1* gene and leaf morphogenesis defects characteristic of ectopic *KNOX1* expression (Saddic et al. 2006). Hence, *LMI1* likely serves as an upstream negative regulator of *KNOX1* gene expression (Saddic et al. 2006). Furthermore, no other known leaf shape modification genes are likely present in the 5.4 cM region delineated here. Taken together, the above findings

Table 4 Select *G. hirsutum* ESTs most similar to putative *G. raimondii* Class I *KNOX* genes used in the semi-quantitative RT-PCR-based gene expression analysis

Arabidopsis KNOX gene(s)	Homolog in the *G. raimondii* genome	EST/TC in *G. hirsutum* gene index	Score	E-value
KNAT1	Gorai.009G223200	DW224477	2,778	3.3E-171
KNAT1	Gorai.009G223200	TC273538	2,626	1.7E-113
KNAT6	Gorai.005G180500	DW236483	2,444	9.2E-261
KNAT1	Gorai.005G098100	TC254202	3,555	1.4E-155
STM	Gorai.009G181500	TC273462	4,895	3.0E-216
KNAT1	Gorai.010G029000	TC254362	6,874	0

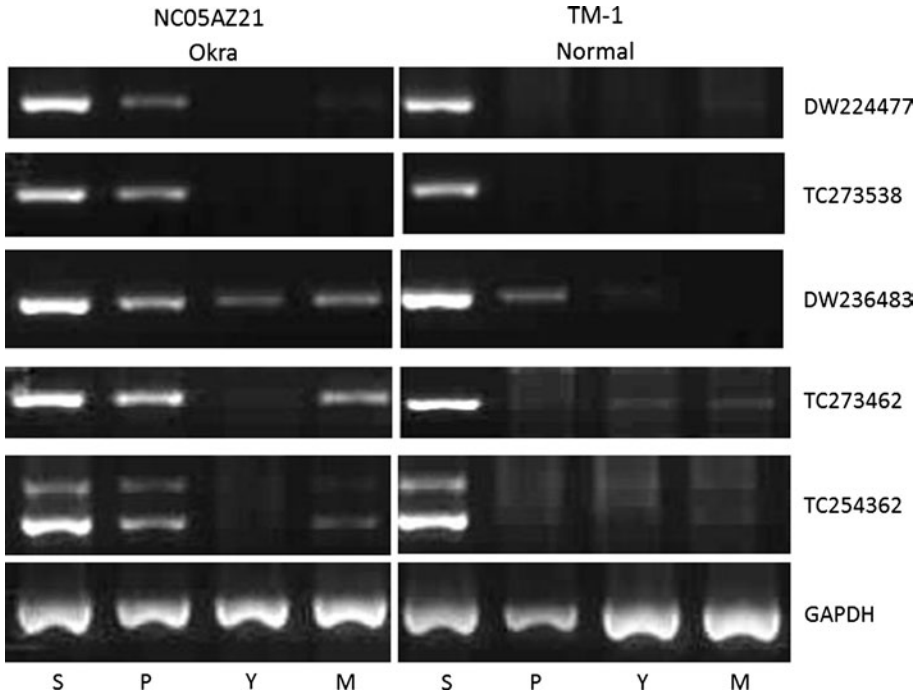

Fig. 4 Comparison of *KNOX1* gene expression between okra and normal leaf cotton. Okra leaf cotton shows later and stronger expression of multiple *KNOX1* genes during the development of leaves when compared with normal leaf cotton. *S* shoot apical meristem (SAM)-enriched stem, *P* leaf primordia, *Y* young leaf, *M* mature leaf

and inferences indicate that modifications to an *LMI1*-like ortholog could be possible candidates for the molecular mechanism underlying the Chr15 leaf shape locus in cotton.

Identification and molecular characterization of the genes involved in leaf shape are essential prerequisites to elucidate the molecular mechanism controlling leaf shape and permit its manipulation for cotton cultivar improvement. Efforts are currently underway to fine map the leaf shape locus, definitively determine the underlying gene, and identify the specific nucleotide differences between the alleles at this locus.

Acknowledgments We thank Cotton Incorporated for supporting this work through its core research and doctoral fellowship programs. We thank Kim Howell, Jared Smith, and Sharon Williamson at the USDA Eastern Regional Small Grains Genotyping Laboratory for assistance in capillary-based analysis of molecular markers. We also thank Rich Tuttle for supplying primers for GAPDH-positive control used in the expression analysis.

Conflict of interest The authors declare that there are no conflicts of interest in the reported research.

Ethical standards The authors note that this research was performed and reported in accordance with ethical standards of the scientific conduct.

References

Blenda A, Fang DD, Rami JF, Garsmeur O, Luo F, Lacape JM (2012) A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. PLoS One 7:1–16

Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell RG, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Deynze AV, Zhu Y, Yu S, Abdurakhmonov I, Katageri I, Kumar PA, ur-Rhaman M, Zafar Y, Yu JZ, Kohel RJ, Wendel JF, Paterson AH (2007) Toward sequencing cotton (Gossypium) genomes. Plant Physiol 145:1303–1310

Dolan L, Poethig RS (1991) Genetic analysis of leaf development in cotton. Development 1:39–46

Endrizzi JE, Brown MS (1964) Identification of monosomes for six chromosomes in *Gossypium hirsutum*. Am J Bot 51:117–120

Endrizzi JE, Kohel RJ (1966) Use of telosomes in mapping three chromosomes in cotton. Genetics 54:535–550

Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. Genetics 164:311–321

Gill KS, Gill BS, Endo TR, Boyko EV (1996) Identification and high-density mapping of gene rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Ha CM, Jun JH, Nam HG, Fletcher JC (2004) *BLADE-ON-PETIOLE1* encodes BTB/POZ domain protein required for leaf morphogenesis in *Arabidopsis thaliana*. Plant Cell Physiol 45(10):1361–1370

Hay A, Tsiantis M (2010) KNOX genes: versatile regulators of plant development and diversity. Development 137:3153–3165

Janssen BJ, Lund L, Sinha N (1998) Overexpression of a homeobox gene, *LeT6*, reveals indeterminate features in the tomato compound leaf. Plant Physiol 117:771–786

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418

Jones JE (1982) The present state of the art and science of cotton breeding for leaf-morphological types. Proceedings of the Belt-wide Cotton Production Research Conferences, National Cotton Council of America, Memphis, TN pp 93–99

Kuraparthy V, Bowman DT, Jones DC (2013) Registration of NC05AZ21 okra-leaf and NC05-11 normal-leaf Upland cotton germplasm. J Plant Registrat 7:334–338

Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F, Yu J (2012) Parallel domestication of the Shattering1 genes in cereals. Nat Genet 44:720–724

Lincoln C, Long J, Yamaguchi J, Serikawa K, Hake S (1994) A knotted1-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell 6:1859–1876

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Meredith WR (1984) Influence of leaf morphology on lint yield of cotton-enhancement by the sub okra trait. Crop Sci 24:855–857

Rahman HU, Bibi A, Latif M (2005) Okra-leaf accessions of the upland cotton (*Gossypium hirsutum* L.): genetic variability in agronomic and fibre traits. J Appl Genet 46:149–155

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (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

Rosin FM, Hark JK, Horner HT, Davies PJ, Hannapel DJ (2003) Overexpression of a *Knotted*-like homeobox gene of potato alters vegetative development by decreasing gibberellin accumulation. Plant Physiol 132:1106–1117

Saddic LA, Huvermann B, Bezhani S, Su Y, Winter CM, Kwon CS, Collum RP, Wagner D (2006) The *LEAFY* target *LMI1* is a meristem identity regulator and acts together with *LEAFY* to regulate expression of *CAULIFLOWER*. Development 133:1673–1682

Schuelke M (2000) An economic method of the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234

Song XL, Guo WZ, Han ZG, Zhang TZ (2005) Quantitative trait loci mapping of leaf morphological traits and chlorophyll content in cultivated tetraploid cotton. J Integr Plant Biol 47:1382–1390

Tanaka-Ueguchi M, Itoh H, Oyama N, Koshioka M, Matsuoka M (1998) Over-expression of a tobacco homeobox gene, *NTH15*, decreases the expression of a gibberellin biosynthetic gene encoding GA-20 oxidase. Plant J 15:391–400

Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y (1999) Modified CTAB procedure for DNA isolation from epiphytic cacti of the genera *Hylocereus* and *Selenicereus* (Cactaceae). Plant Mol Biol Rep 17:249–254

Uchida N, Kimura S, Koenig D, Sinha N (2010) Coordination of leaf development via regulation of *KNOX1* genes. J Plant Res 123:7–14

Wells R, Meredith WR Jr (1986) Normal vs. okra leaf yield interactions in cotton. II: analysis of vegetative and reproductive growth. Crop Sci 26:223–228

Wendel JF, Cronn RC (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Werner JE, Endo TR, Gill BS (1992) Toward a cytogenetically based physical map of the wheat genome. Proc Natl Acad Sci USA 89:11307–11311

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, San-Miguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644