Supporting information for

Modifications to a *LATE MERISTEM IDENTITY-1* gene are responsible for the major leaf shapes of Upland cotton (*Gossypium hirsutum* L.)

Ryan Andres[a,1], Viktoriya Coneva[b], Margaret H. Frank[b], John R. Tuttle[a], Luis Fernando Samayoa[a, c], Sang-Won Han[d], Baljinder Kaur[a], Linglong Zhu[a], Hui Fang[a], Daryl T. Bowman[a], Marcela Rojas-Pierce[c], Candace H. Haigler[a, c], Don C. Jones[e], James B. Holland[a,d], Daniel H. Chitwood[b,1], Vasu Kuraparthy[a,1]

To whom correspondence should be addressed. E-mail: vasu_kuraparthy@ncsu.edu (V.K.) or dchitwood@danforthcenter.org (D.H.C.).

This file includes:

Tables and Figures

Supplementary Tables

Table S1. Chi-square analysis of a large F_2 population: Chi-square analysis of individual plant phenotype confirms the assumption that leaf shape in cotton is controlled by a single, incompletely dominant gene. *Three plants were classified as ambiguous owing to stunted growth.

Phenotype	**Observed**	**Expected (1:2:1)**	**Chi-square**
Okra	262	256	0.141
Heterozygous	509	512	0.018
Normal	253	256	0.035
Total	**1,024***	**1,024**	**0.194 (p =0.91)**

Table S2. Fine mapping of the leaf shape locus (*L-D*$_1$) using association mapping of putative candidate gene-based markers. Marker details are presented in *SI Appendix*, Fig. S1. There was no association between leaf shape and GhLS-STS2 allowing the candidate region to be reduced to the 52 kb, four gene interval between *Gorai.002G243900* and *Gorai.002G244200* shown in Fig. 2D. The near complete association between leaf shape and GhLS-SNP2 was unable to further reduce the candidate region. The large promoter variant of GhLS-STS1 was completely associated with *okra* and *super-okra* leaf shape while the smaller genic polymorphism of 13-LS-195 was completely associated with *normal* leaf shape. Individual marker data may be missing so numbers may not sum completely.

Leaf shape	**Putative candidate gene markers**									**Total**
	GhLS-STS1 *Gorai.002G244000* Promoter Size		**13-LS-195 *Gorai.002G244000* Genic InDel**		**GhLS-SNP2 *Gorai.002G244100* Intronic SNP**		**GhLS-STS2 *Gorai.002G244200* 3' Intergenic InDel**			
	Large	**Small**	**172 bp**	**180 bp**	**C**	**A**	**471 bp**	**458 bp**	**450 bp**	
Normal	0	393	391	0	363	0	171	214	4	393
Okra	109	0	0	109	0	101	34	66	8	109
Sub-Okra	0	23	0	24	5	17	21	1	1	24
Super-Okra	12	0	0	12	0	6	0	12	0	12
Total	121	416	391	145	368	124	226	293	13	538

Table S3. Fine mapping of the leaf shape locus (*L-D*$_1$) using isogenic lines and putative candidate gene-based markers. Marker details are presented in *SI Appendix*, Fig. S1. GhLS-STS2 failed to show association with the *normal* leaf shape, confirming that the candidate region can be reduced to the 52 kb, four gene interval between *Gorai.002G243900* and *Gorai.002G244200* shown in Fig. 2D. GhLS-SNP2 remained associated with leaf shape, rendering the maker unable to further reduce the candidate region. The large promoter variant of GhLS-STS1 remained associated with *okra* and *super-okra* leaf shape while the smaller genic polymorphism of 13-LS-195 remained associated with *normal* leaf shape. GhLS-SNP1 data for LA213 *okra* is missing.

Isoline	**Putative candidate gene markers**							
	GhLS-STS1 *Gorai.002G244000* Promoter Size		**13-LS-195 *Gorai.002G244000* Genic InDel**		**GhLS-SNP2 *Gorai.002G244100* Intronic SNP**		**GhLS-STS2 *Gorai.002G244200* 3' Intergenic InDel**	
	Large	**Small**	**172 bp**	**180 bp**	**C**	**A**	**471 bp**	**458 bp**
LA213-*sub-okra*		X		X		X	X	
LA213-*super-okra*	X			X		X		X
LA213-*okra*	X			X			X	
Stoneville 7A-*okra*	X			X		X	X	
LA213-*normal*		X	X		X			X
Stoneville 7A-*normal*		X	X		X		X	

Table S4. The 20 *Gossypium* accessions used to sequence *GhLMI1*-like genes. Sanger sequences of both *GhLMI1-D1a* and *GhLMI1-D1b* were collected from these 20 tetraploid cotton varieties in order to construct Fig. 3C and *SI Appendix*, Fig. S6. There are five varieties of each of the four major leaf shapes; *normal*, *okra*, *super-okra*, and *sub-okra*. A consensus sequences was determined from each group of five in order to generate a single sequence for each leaf shape at both loci.

Entry No.	Accession/Variety	Leaf shape	Species
1	NC11-2100	Normal	*G. hirsutum*
2	NC11-2091	Normal	*G. hirsutum*
3	Coker 312	Normal	*G. hirsutum*
4	TM-1	Normal	*G. hirsutum*
5	LA 213 Frego Bract	Normal	*G. hirsutum*
6	NC05AZ21	Okra	*G. hirsutum*
7	Stoneville 7A Okra	Okra	*G. hirsutum*
8	Acala Red Okra Leaf	Okra	*G. hirsutum*
9	LA213 Okra	Okra	*G. hirsutum*
10	Aub Okra-16	Okra	*G. hirsutum*
11	Super Okra UA2-5	Super Okra	*G. hirsutum*
12	Acala 6010-15-U Okra	Super Okra	*G. hirsutum*
13	LA213 Super Okra	Super Okra	*G. hirsutum*
14	P62Lo	Super Okra	*G. barbadense*
15	Acala 6010-27-10 Okra	Super Okra	*G. hirsutum*
16	TAMCOT CAMD-ES	Sub-Okra	*G. hirsutum*
17	DPL 5540-85	Sub-Okra	*G. hirsutum*
18	MD 65-11	Sub-Okra	*G. hirsutum*
19	DES 422	Sub-Okra	*G. hirsutum*
20	NC05AZ06	Sub-Okra	*G. hirsutum*

Table S5. Primer Sequences of the markers used in association mapping in *SI Appendix*, Table S2 and S3.

Marker Name	Type of Polymorphism	*G. raimondii* Gene	Forward Sequence 5' → 3'	Reverse Sequence 5' → 3'
GhLS-STS1	133bp Duplication	*Gorai.002G244000*	TACAAAGCCTACCCCATCGT	TGGAGAGAGGGTGGACTTGT
13-LS-195	Exonic 8bp InDel	*Gorai.002G244000*	ACCTTTTACGCAGGTGATGG	TCGGATATAGTCGTTTCCTGCT
STK4100-1	C-A SNP-KASP	*Gorai.002G244100*	GAAGGTGACCAAGTTCATGCTAC AAACTCCTTCTGTCAGTC	GGGCTGTATACAAAGGGGTCAACAA
STK4100-2	C-A SNP-KASP	*Gorai.002G244100*	GAAGGTCGGAGTCAACGGATTCT GCTACAAACTCCTTCTGTCAGTA	GGGCTGTATACAAAGGGGTCAACAA
GhLS-STS2	3' Intergenic InDel	*Gorai.002G244200*	TTGAAAGCCATTGTTGAAGG	CGGATCGTATGGTAAGTTTGC

Table S6. Names and sequences of primers used in RT-PCR expression analysis. These primers were used in semi-quantitative and/or qRT-PCR studies. Primers were used to determine the expression of the specified gene in the candidate interval determined in association mapping.

Primer	Forward Sequence 5' → 3'	Reverse Sequence 5' → 3'
GhLMI1-D1a **Expression 1**	ACAAATCGCTGTCTGGTTCC	TGCTTTCGACAGTCTCTTCG
GhLMI1-D1a **Expression 2**	GGGAACTTGGACTTCAACCA	ACCCCAATAAGGGGATGAAA
GhLMI1-D1b **Expression 1**	GGCACCATTCGACCCTTTAT	TGCCAGCTTCTTCAAATCCT
GhLMI1-D1b **Expression 2**	ACAAATCGCTGTCTGGTTCC	GTCGCCTGTTCCACTAGCAT
GAPDH **- Positive Control**	ATCAAGGGCACCATGACTACCACT	ACCAGTTGAAGTCGGGACGATGTT
UBI14 **Positive Control for** ***GhLMI1-D1a***	GCTTACGGGAAAGACGATCA	GAAGATCTGCATCCCACCTC
UBI14 **Positive Control for** ***GhLMI1-D1b***	GCTGGGAAACAACTGGAAGA	CCATTACAGGGCACAGTCCT
Gorai.002G44100 **Expression 1**	ACTTGGAAGAGGTGCTTTCG	TTCCGGTGCTACATAACCTCTAGT
Gorai.002G44100 **Expression 2**	GCCATTAGAGGAACTAGAGGTTATGT	AAATGCAAAAGGGGAAGGAC
Gorai.002G43900 **Expression 1**	GATGATGAGTCTGATGGTGATGAT	AAGCTTGATAGGTCTCCCCTTT
Gorai.002G43900 **Expression 2**	AAAGGTGGTGGTGGTAGTGG	ATCATCACCATCAGACTCATCATC
GhLMI1-D1b **VIGS 1**	GAACACTCATATAATACGCTTAAACAT	CTTGGTTCCTTGTCGCC
GhLMI1-D1b **VIGS 2**	CGAATCGAGCTACTTTCGATG	CTTGGTTCCTTGTCGCC

Table S7. Genome specificity of *GhLMI1-D1a* amplification for sequencing primers. Sequence differences resulting in a genome-specific primer are highlighted and underlined in red. *G. raimondii* and *G. arboreum* sequences were retrieved from https://www.cottongen.org/tools/gbrowse. *193R is the second reverse primer used in the nested amplification of the 5' region of *GhLMI1-D1a.*

Primer Combination	Primer	*G. raimondii* Sequence 5'→3'	*G. arboreum* Sequence 5'→3'	Genome Specific?
14-LS-5F + 13-LS-145R	14-LS-5F	GGTATTTCTACGGATCAATGTGC	GGTATTTCTACGGATCAATGTGC	No
	13-LS-145R	TGCTTTCGACAGTCTCTTCG	TGCTTTCAACGGTCTCTTCG	Yes
	13-LS-193R*	TGGTTCCTTGTCACTTGTTCC	TGGTTCCTTGTCACTTGTTCC	No
13-LS-145F +13-LS-	13-LS-145F	GGGAACTTGGACTTCAACCA	GGGAACTTGGACTTCAACCA	No
	13-LS-198R	ATGCCAGCACGAGGTACG	N/A	Yes

Table S8. Genome specificity of *GhLMI1-D1b* amplification for sequencing primers. Sequence differences resulting in a genome-specific primer are highlighted and underlined in red. *G. raimondii* and *G. arboreum* sequences were retrieved from https://www.cottongen.org/tools/gbrowse. *****Single base pair insertion may be too far from the 3' end of the primer to impact amplification.

Primer Combination	Primer	*G. raimondii* Sequence 5'→3'	*G. arboreum* Sequence 5'→3'	Genome Specific?
14-LS-12F + 14-LS-9R	14-LS-12F	TACAAAGCCTACCCCATCGT	TAC AAGCCTACCCCATCCGTTACGCTCATCCGT	Yes
	14-LS-9R	TGGAGAGAGGGTGGACTTGT	TGGAGAGAGGGTGGACTTGT	No
13-LS-200F + 13-LS-195R	13-LS-200F	ATTCCCTTCTCTCGCTCTCT	ATTCCCTTCTCTCACTCTCT	Yes
	13-LS-195R	TCGGATATAGTCGTTTCCTGCT	TCGGATCATAGTCGTTTCCTGCT	No*
14-LS-9F + 13-LS-199R	14-LS-9F	TGGTTGTTGGGATCAACCTT	TGGTTCTTGGGATCAACCTA	Yes
	13-LS-199R	AGTAGTGAGTGGAAACTGGGT	AGTAGTGAGTGGAAACTGGGT	No

Table S9. *Gossypium* diploid accessions used in the genome specificity checks. These diploid cotton lines were used to confirm the genome specificity of primers detailed in *SI Appendix*, Table S7 and S8. Primers were considered genome specific if they amplified specifically in the D genome diploids but failed to do so in *G. arboreum.*

NC Accession No.	Species	Genome	PI No.	Name
174	*G. raimondii*	DD	530920	$D_{5\text{-}1}$
175	*G. raimondii*	DD	530921	$D_{5\text{-}2}$
176	*G. raimondii*	DD	530928	$D_{5\text{-}4}$
235	*G. trilobum*	DD	530967	$D_{8\text{-}5}$
222	*G. thurberi*	DD	530767	$D_{1\text{-}3}$
223	*G. thurberi*	DD	530768	$D_{1\text{-}4}$
260	*G. thurberi*	DD	530766	$D_{1\text{-}2}$
262	*G. thurberi*	DD	530782	$D_{1\text{-}18}$
265	*G. thurberi*	DD	530789	$D_{1\text{-}25}$
501	*G. arboreum*	AA	167906	N/A
503	*G. arboreum*	AA	529740	SMA4
505	*G. arboreum*	AA	615700	Chinese Narrow Leaf
506	*G. arboreum*	AA	615701	Soudanense
507	*G. arboreum*	AA	615750	Punjabi 39
508	*G. arboreum*	AA	615759	231R

Table S10. Primers used for sequencing *GhLMI1*-like genes. These primers were used to complete the Sanger sequencing of both *LMI1*-like candidate leaf shape genes. Sequencing template was purified PCR product from amplification using the primer pairs listed in *SI Appendix*, Table S7 and S8.

Marker Name	Orientation	Sequence 5' → 3'
***GhLMI1-D1a* Seq. 1**	Forward	TTGAAAGCCATTGTTGAAGG
***GhLMI1-D1a* Seq. 2**	Reverse	CAAAAATTCAGCCTTATTCATGC
***GhLMI1-D1a* Seq. 3**	Reverse	ATGCCAGCACGAGGTACG
***GhLMI1-D1a* Seq. 4**	Forward	GGAACAAGTGACAAGGAACCA
***GhLMI1-D1a* Seq. 5**	Reverse	TCCAACAATGGCTTTCAACA
***GhLMI1-D1a* Seq. 6**	Forward	GGGAACTTGGACTTCAACCA
***GhLMI1-D1a* Seq. 7**	Forward	TGGACTTGGGTAATGAGATGG
***GhLMI1-D1a* Seq. 8**	Reverse	GGAACCAGACAGCGATTTGT
***GhLMI1-D1a* Seq. 9**	Reverse	GGAGATCTCAAAATTCATGAACAA
***GhLMI1-D1a* Seq. 10**	Reverse	CCCAAGACAGGGTTTAACGA
***GhLMI1-D1a* Seq. 11**	Forward	ACCCTTTGTTTCACGACCAG
***GhLMI1-D1a* Seq. 12**	Forward	CCCAAAGCCAATGTTTTA
***GhLMI1-D1a* Seq. 13**	Forward	GGTATTTCTACGGATCAATGTGC
***GhLMI1-D1a* Seq. 14**	Reverse	TGGTTCCTTGTCACTTGTTCC
***GhLMI1-D1a* Seq. 15**	Forward	AACCGGTTAGAGAGAGCTAAAGA
***GhLMI1-D1a* Seq. 16**	Reverse	GGAAAGGTGCAATCACAGGT
***GhLMI1-D1a* Seq. 17**	Forward	GGGAACTTGGACTTCAACCA
***GhLMI1-D1a* Seq. 18**	Reverse	TCCATAAACCCTTGATGTTGC
***GhLMI1-D1b* Seq. 1**	Forward	TACAAAGCCTACCCCATCGT
***GhLMI1-D1b* Seq. 2**	Reverse	TCATTCATTACATCAATCGTCAAA
***GhLMI1-D1b* Seq. 3**	Forward	TGGTTGTTGGGATCAACCTT
***GhLMI1-D1b* Seq. 4**	Reverse	TGGAGAGAGGGTGGACTTGT
***GhLMI1-D1b* Seq. 5**	Forward	AAGTGATGATTTGACGATTGATG
***GhLMI1-D1b* Seq. 6**	Reverse	GGAACCAGACAGCGATTTGT
***GhLMI1-D1b* Seq. 7**	Forward	AACTTTTGTTTGATGAGGGATTG
***GhLMI1-D1b* Seq. 8**	Reverse	TCAGCCGTAGTAAGCAAGCA
***GhLMI1-D1b* Seq. 9**	Forward	ACCTTTTACGCAGGTGATGG
***GhLMI1-D1b* Seq. 10**	Reverse	TCGGATATAGTCGTTTCCTGCT
***GhLMI1-D1b* Seq. 11**	Forward	GGCACCATTCGACCCTTTAT
***GhLMI1-D1b* Seq. 12**	Forward	TGGAACTTGGTGCACTGTTT

Supplementary Figures

Fig. S1. Markers used in association mapping and isogenic line studies. Four novel markers: GhLS-STS1, GhLS-STS2, GhLS-SNP1 and GhLS SNP2 were developed and run in both studies along with 13-LS-195 that had previously co-segregated with leaf shape phenotype (1). Four markers: GhLS-STS1, 13-LS-195, GhLS-SNP1 and GhLS-SNP2 showed complete association with leaf shape phenotype. GhLS-STS2 showed no association with leaf shape.

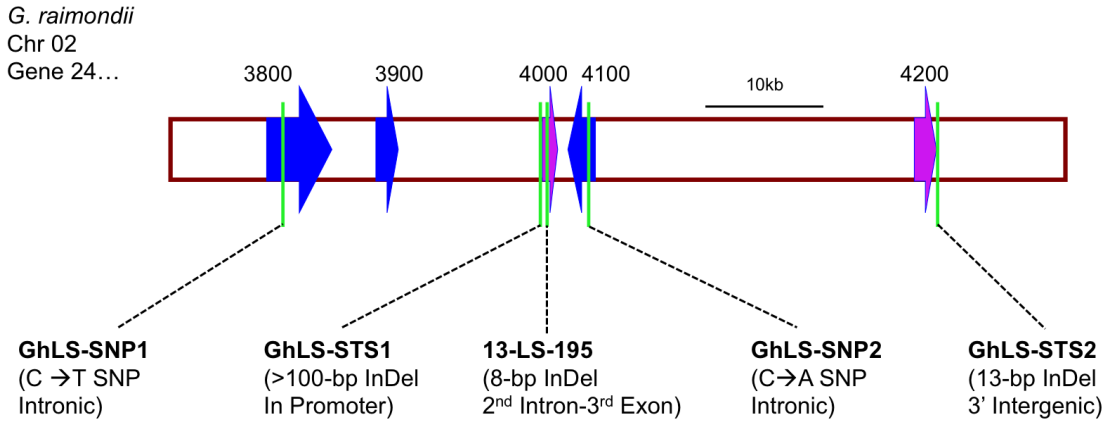
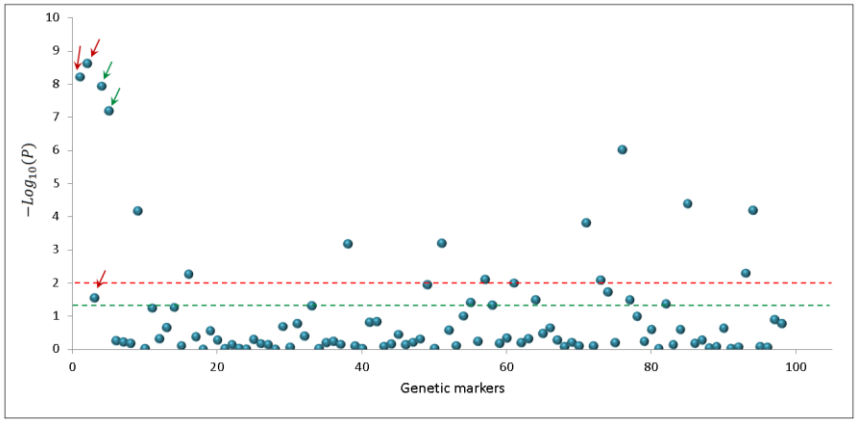


Fig. S2. Results of genome-wide association scan for leaf shape in diverse set of 406 cotton inbred lines (*A*). Blue circles represent the *P*-value of each marker tested in a logistic regression model that also included the first three principal components of the population structure analysis. Three size-based markers from candidate genes are indicated by red arrows while two SNP markers are indicated by green arrows. The green and red dashed lines represent the 0.05 and 0.01 significance levels. Results of second genome-wide association scan for leaf shape in diverse set of 406 cotton inbred lines, adjusting for the effects of the candidate gene (*B*). Blue circles represent the *P*-value of each marker tested in a logistic regression model that also included the most significant candidate gene marker from the initial scan plus the first three principal components of the population structure analysis. The green and red dashed lines represent the 0.05 and 0.01 significance levels.

A



B

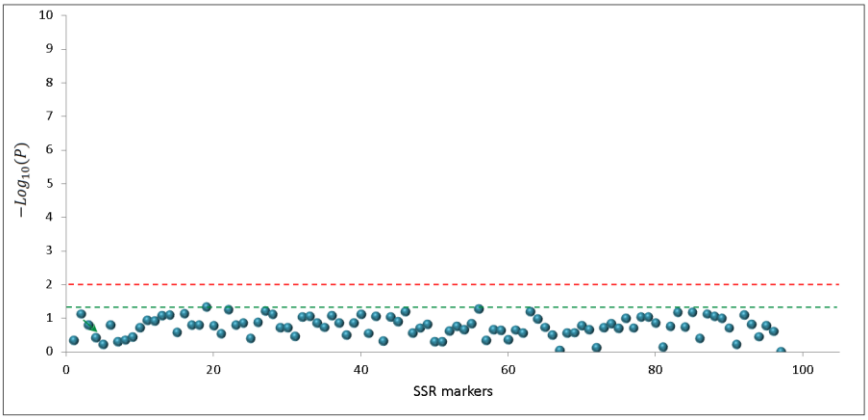


Fig. S3. Allelism between *okra* leaf shapes of parental accession (NC05AZ21) and isoline (LA213-*okra*). *Okra* leaf shape genes in isoline LA213-*okra* and NC05AZ21 are allelic. Leaf shape phenotypes of greenhouse grown parents and their F_1 hybrid at approximately 40 days after planting.

Fig. S4. Semi-quantitative expression analysis of leaf shape candidate genes. Neither *GhHRA1* (*Gorai.002G243900*) nor *GhRLK1* (*Gorai.002G244100*) were expressed in critical young leaf tissue, eliminating these two genes from consideration. *GhLMI1-D1a* (*Gorai.002G244200*) appeared equally expressed across leaf shapes. *GhLMI1-D1b* (*Gorai.002G244000*) was expressed only in leaf shapes (*okra* and *super-okra*) with the larger promoter indicating differential expression of this gene could play a major role in leaf shape. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was the reference gene.

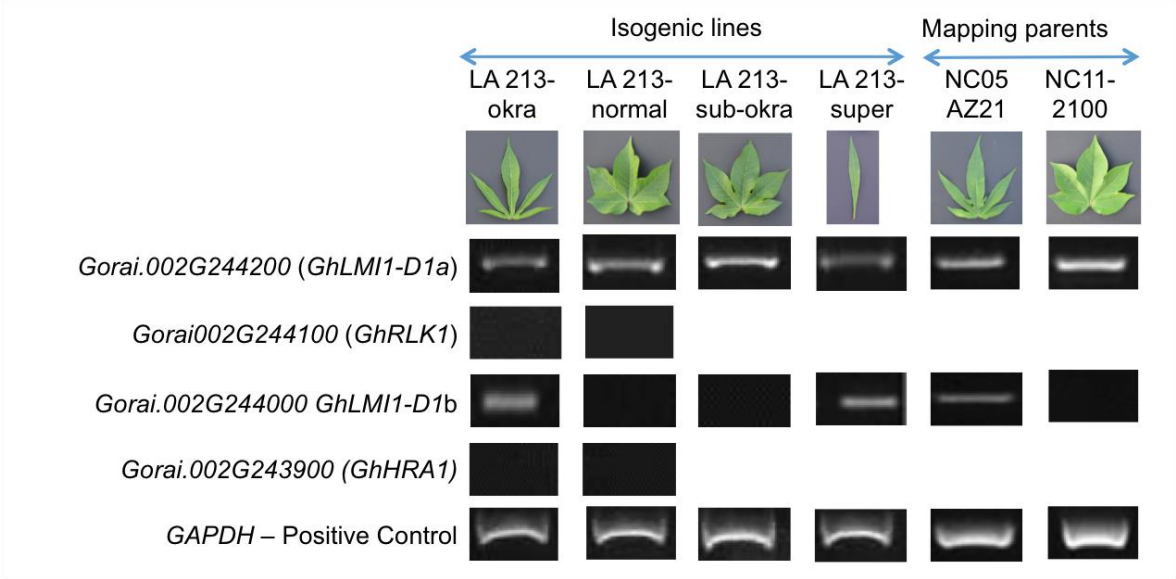


Fig. S5. RNA-Seq indicates that *GhLMI1-D1b* (Cotton_D_gene_10005014) is significantly upregulated in three biological replicates of *okra* relative to *normal* plastochron 2 samples, and is the only differentially expressed gene within the 52kb candidate interval. Illumina sequenced reads were mapped to the *G. raimondii* D5 BGI-CGP v1.0 genome, and significantly differentially expressed genes (false discovery rate ≤ 0.05) were identified using edgeR. Expression levels are plotted in reads per million (RPM) with false discovery rate values listed above the bars. Gene models that are marked with N/A did not pass the edgeR filter of having at least 1 RPM mapped across at least 3 samples. BGI gene identifiers are listed below the bars with the corresponding JGI gene identifiers in parentheses. *Normal* samples are plotted in blue and *okra* samples in orange.

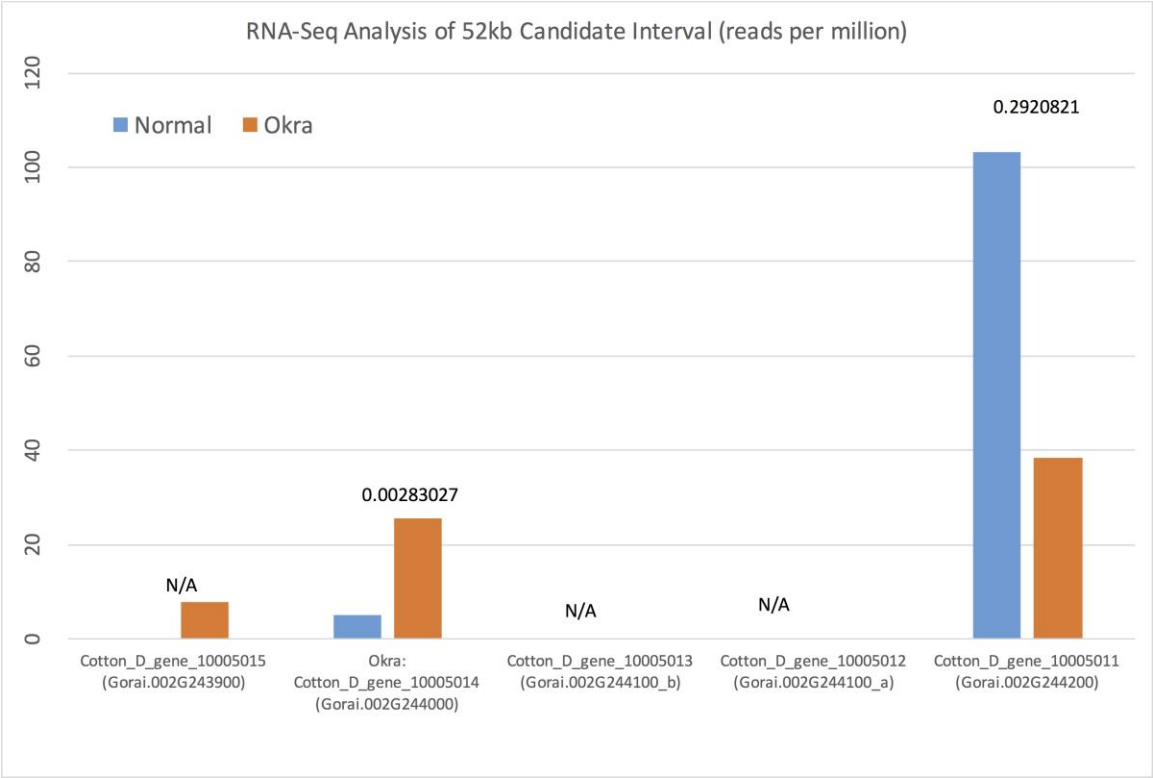


Fig. S6. Nucleotide polymorphisms of the *GhLMI1-D1a* gene among different leaf shapes. Two variants of *GhLMI1-D1a* were found, neither of which perfectly associated with leaf shape. Both variants encode full-length proteins and have no obvious polymorphisms that would alter function.

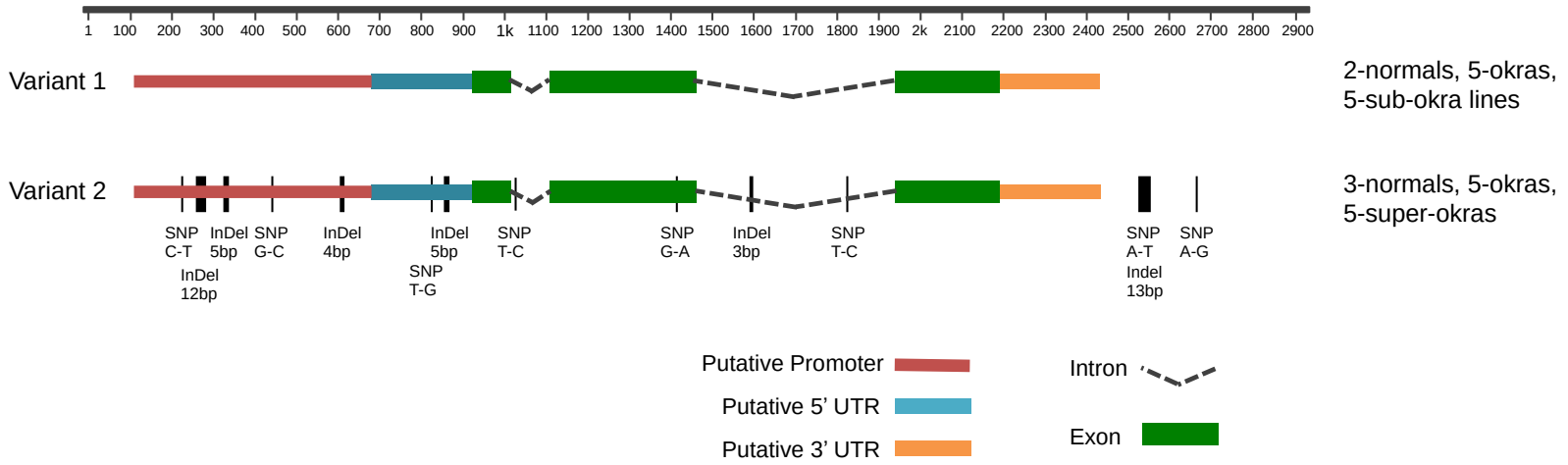


Fig. S7. Leaves from representative VIGS plants. True leaves five through ten from representative plants of all VIGS treatments are shown. Severe reductions in leaf lobing and sinus depth was seen in the LA213-*okra* TRV: *LMI1-D1b* that briefly produced normal leaves. Abolishment of viral silencing proceeded similar to that seen in the TRV: CHLI positive control treatment.

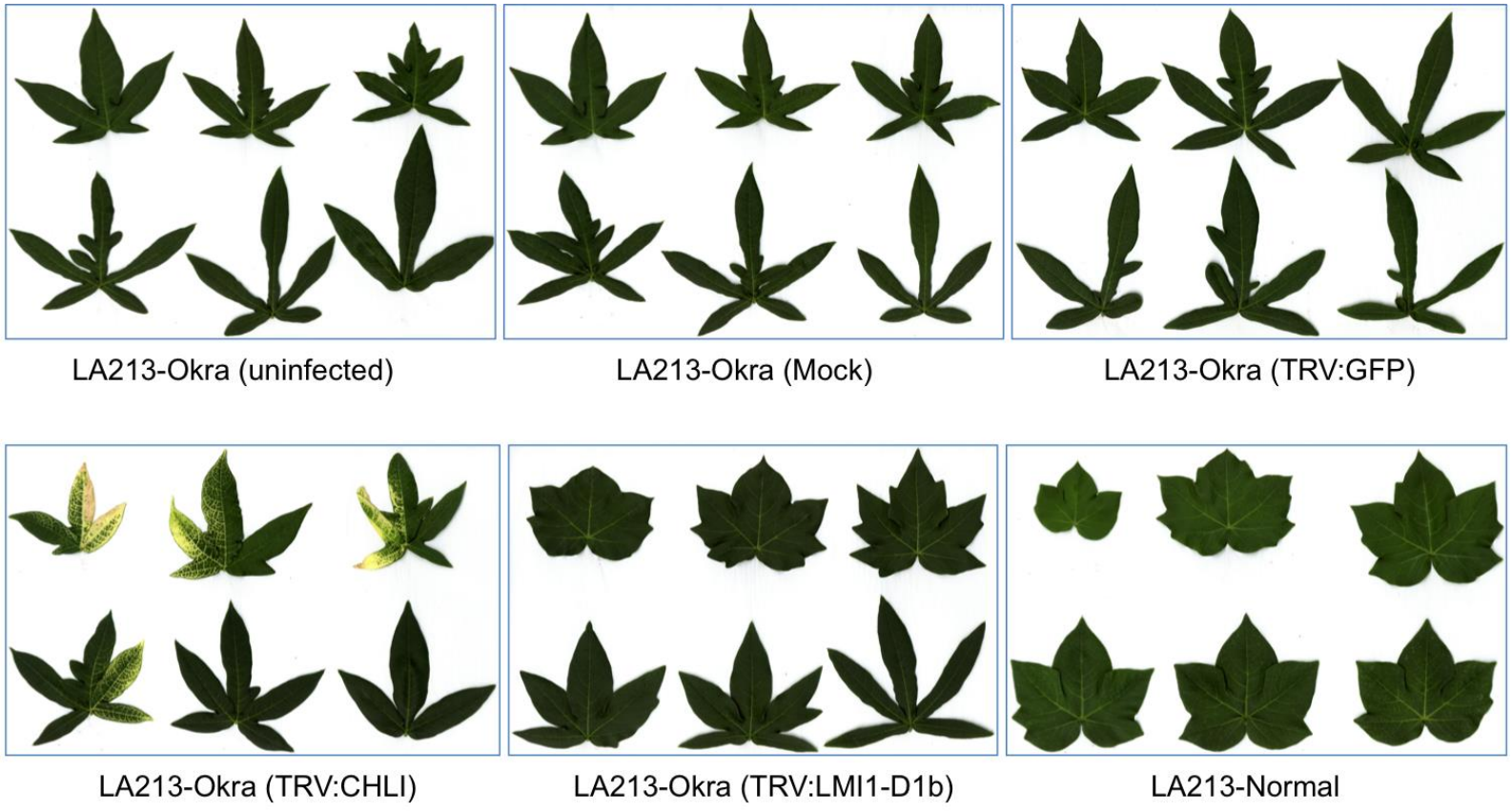


Fig. S8. Enriched Biological Process (BP) GO terms for differentially expressed genes up-regulated in *okra* (A) or *normal* (B) P2 samples (Datasets S5 and S6). GO terms are plotted in order of significance (Fisher's exact test statistic), with the most-significantly enriched term at the top. Purple bars represent the number of occurrences of each term in the respective dataset, while the frequency of each term that would be expected by chance is plotted in green.

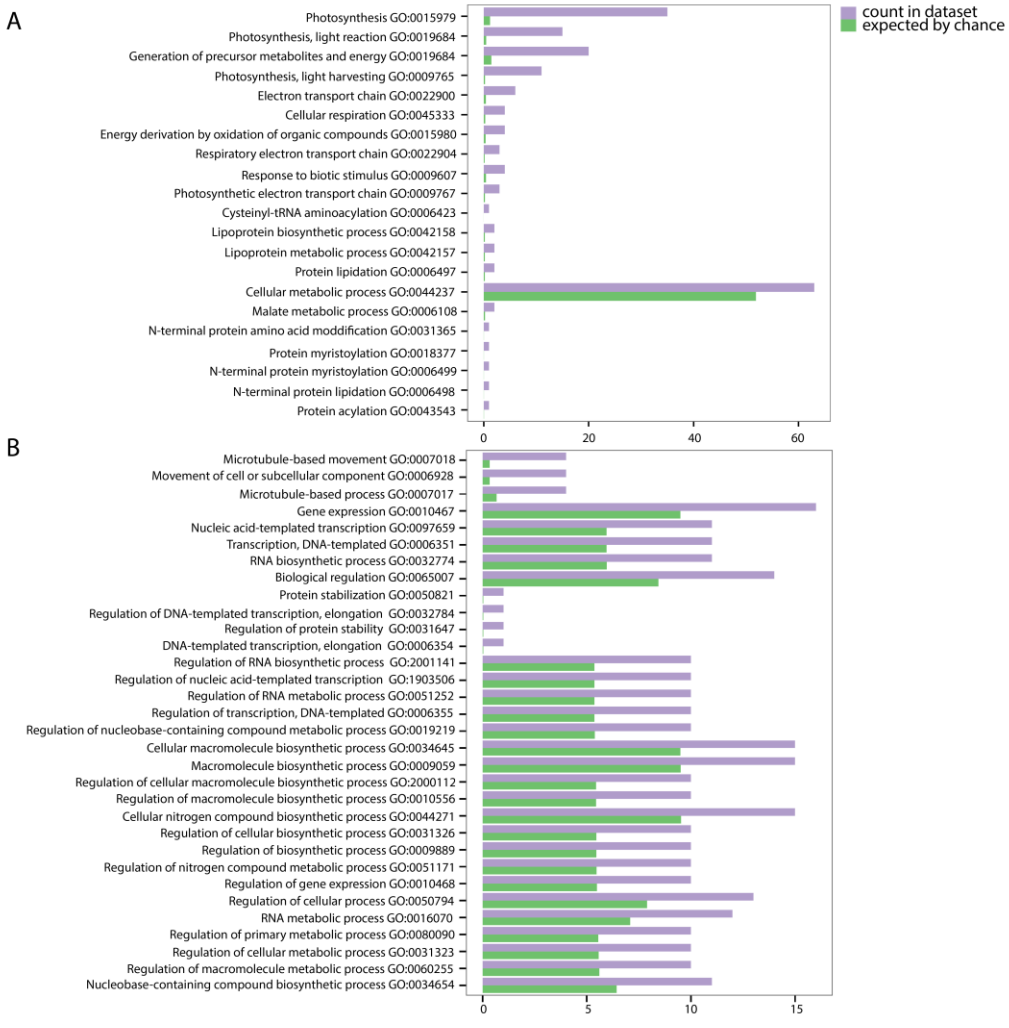


Fig. S9. GhLMI1-D1b^{Okra} localizes to the nucleus. A GhLMI1-D1b^{Okra}-GFP fusion construct driven by the 35S (*A*-*C*) or pUBQ10 (*D*-*F*) promoter was transiently expressed in *Nicotiana benthamiana* leaves. Signal from GhLMI1-D1b-^{Okra}-GFP (*A*, *D*), the nuclear marker RFP-Histone2B (*B*, *E*) or the merged images including brightfield (*C*, *F*) are shown. Scale bar = 20µm.

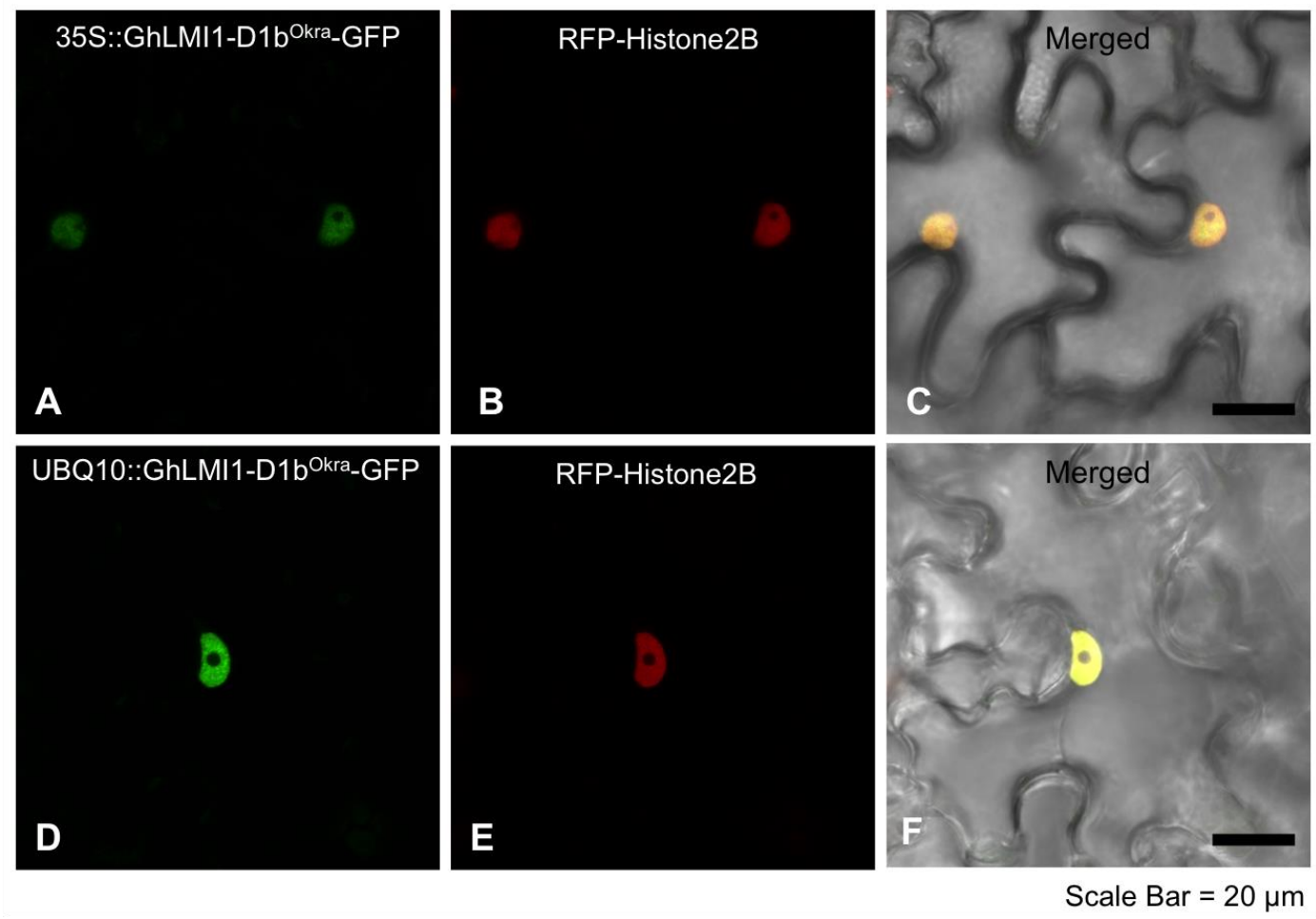
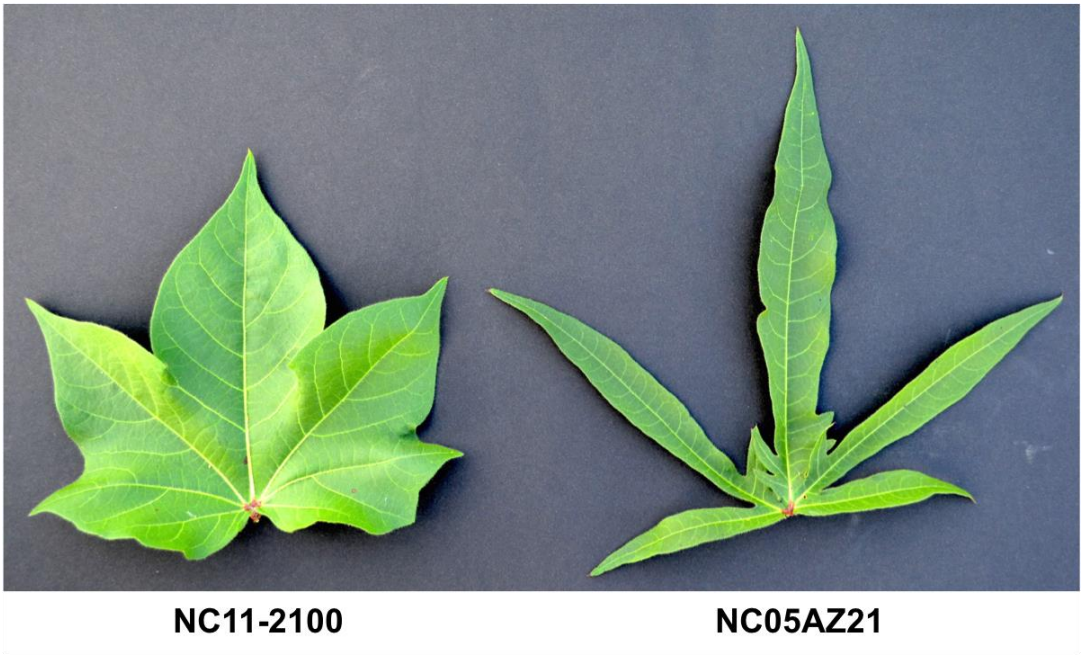


Fig. S10. Leaf shape phenotypes of the field grown parental accessions used for fine mapping the *L-D*$_1$ locus of cotton. Left- *normal* shaped leaves of the landrace accession NC11-2100. Right- *okra* leaf shape of the germplasm line NC05AZ21.



SI References

1. Andres RJ, Bowman DT, Kaur B, Kuraparthy V (2014) Mapping and genomic targeting of the major leaf shape gene (*L*) in Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet* 127(1): 167-177.