**Supplementary table legends**

**Supplementary Table 1**.Complete list of lines used for different analyses in mapping and cloning of leaf shape locus in *G. hirsutum*. A symbol in front of variety name indicates analysis they were used for. Lines with \* symbol were part of association panel, lines with πsymbol were used for Principal Component Analysis (PCA), lines with + symbol were included for Genome wide association analysis (GWAS), and lines with θ symbol were included in morphometric analyses.

**Supplementary Table 2**. Chi-square analysis of a large F2 population:Chi-square analysis of individual plant phenotype confirmsthe assumption that leaf shape in cotton is controlled by a single, incompletely dominant gene. \*Three plants were classified as ambiguous owing to stunted growth.

**Supplementary Table 3.** Fine mapping of the leaf shape locus (*L-D1*) using association mapping of putative candidate gene-based markers. Marker details are presented in **Supplementary Fig. 1**. 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-195was completely associated with *normal* leaf shape. Individual marker data may be missing so numbers may not sum completely.

**Supplementary Table 4.** Statistics for association tests of each marker with *okra* vs *normal* leaf shape using logistic regression. An initial scan was performed fitting each marker individually along with the first three principal components of the marker data set to control for population structure. The most significant marker was then added to the model and a second scan was conducted, fitting each other marker individually along with the first three principal components and GhLS-STS1, the most significant marker from the first scan. Since the first marker fit in the model was GhLS-STS1, the other markers in the candidate gene region were not retested in the second scan because of their high correlation with GhLS-STS1.

**Supplementary Table 5.** Fine mapping of the leaf shape locus (*L-D1*) using isogenic lines and putative candidate gene-based markers. Marker details are presented in **Supplementary Fig. 1**. 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-195remained associated with *normal* leaf shape. Note: GhLS-SNP1 data for LA213 *okra* is missing.

**Supplementary Table 6.** 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 **Figures 3c** and **Supplementary Fig. 5**. 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.

**Supplementary Table 7.** Differentially expressed genes of plastochron 2 (P2) stage primordia between *okra* and *normal* genotypes. Illumina RNA-sequence reads were aligned to the *G. hirsutum* AD1\_NBI genome; Gene IDs, associated *Arabidopsis thaliana* homologues, *A. thaliana* gene descriptions, and GO IDs are listed in columns 1-4. EdgeR was used to test for significantly differentially expressed genes between the two genotypes. EdgeR calculations for log concentration, log fold-change, p-value, and adjusted p-value are listed in columns 5-8. Raw read counts were normalized to reads per million (rpm) for the heatmap visualization in Fig. 2, and are listed in columns 9-14.

**Supplementary Table 8.** GO enrichment categories for differentially expressed genes that are down-regulated in *okra* relative to *normal* P2 samples. Significantly enriched GO terms were identified using the Fisher exact test function in topGO. GO ID, associated GO term, GO category (BP = Biological Process, CC = Cellular Component, and MF = Molecular Function), number of genes annotated by the GO term, number of genes present in the tested dataset, number of genes expected by random chance, and the Fisher exact test statistic are listed.

**Supplementary Table 9.** GO enrichment categories for differentially expressed genes that are up-regulated in *okra* relative to *normal* P2 samples. Significantly enriched GO terms were identified using the Fisher exact test function in topGO. GO ID, associated GO term, GO category (BP = Biological Process, CC = Cellular Component, and MF = Molecular Function), number of genes annotated by the GO term, number of genes present in the tested dataset, number of genes expected by random chance, and the Fisher exact test statistic are listed.

**Supplementary Table 10.**  Primer Sequences of the markers used in association mapping in **Supplementary Tables 3** and **5**.

**Supplementary Table 11.** 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.

**Supplementary Table 12.** 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.*

**Supplementary Table 13.** 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.

**Supplementary Table 14.** *Gossypium* diploid accessions used in the genome specificity checks.These diploid cotton lines were used to confirm the genome specificity of primers detailed in  **Supplementary Table 12** and **Supplementary Table 13**. Primers were considered genome specific if they amplified specifically in the D genome diploids but failed to do so in *G. arboreum.*

**Supplementary Table 15.** 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 **Supplementary Tables 12** and **13.**