

## **Supplementary Information**

### **Targeted mutagenesis in cotton (*Gossypium hirsutum* L.) using the CRISPR/Cas9 system**

Xiugui Chen<sup>1,2</sup>, Xuke Lu<sup>1</sup>, Na Shu<sup>1</sup>, Shuai Wang<sup>1</sup>, Junjuan Wang<sup>1</sup>, Delong Wang<sup>1</sup>, Lixue Guo<sup>1</sup> and Wuwei Ye<sup>1,\*</sup>

<sup>1</sup> State Key Laboratory of Cotton Biology/Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China

<sup>2</sup> National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China

\*Author for correspondence. WY: yew158@163.com.

Supplementary Information contains:

Supplementary Figures S1-S2

Supplementary Tables S1-S4

**A**

*GhCLA1*-F5: 5' - TTGGCCGGATTACAGAAA-3'  
 At\_chrl1 5' - TTGGCCGGATTACAGAAACGGTCGGAGAGTGAATACGATTGCTTCGGGACTGGTCACAGT  
 Dt\_chrl1 5' - TTGGCCGGATTACAGAAACGGTCGGAGAGTGAATACGATTGCTTCGGGACTGGTCACAGT

TCACCCACCAATCTCTGCAGGCTTGGGTATTAAATCCAGTCTCTCGCAITCTTTTTTGCAI  
 TCACCCACCAATCTCTGCAGGCTTGGGTATTAAATCCAGTCTCTCGCAITCTTTTTTGCAI

ATTATATGATGGTTAAAAATTGAAACCAAGATAACTCTTATCGAACCGTATCGAGTTC  
 ATTATATGATGGTTAAAAATTGAAACCAAGATAACTCTTATCGAACCGTATCGAGTTC

TTTGATGTTGCTCTTAGCTAACATGCGATTTTGTGTTTTTTTAAATGAAATTTTTGCTT  
 TTTGATGTTGCTCTTAGCTAACATGCGATTTTGTGTTTTTTTAAATGAAATTTTTGCTT

TTGATATGTTGAAATACTATTTTCAATCCGGTATGCAAACTAGTTTCACTTGAICTTC  
 TTGATATGTTCAAATACTATTTTCAATCCGGTATGCAAACTAGTTTCACTTGGTCTTC

AGAGATTTAACTGTCTTGAATGTTTCGGGTGGACAGGATGGCTGTGGGAAGGGATCTGAA  
 AGAGATTTAACTGTCTTGAATGTTTCGGGTGGACAGGATGGCTGTGGGAAGGGATCTGAA

AGGTGAAAGGAATCATGTTGTTGCCGTCATAAGTGTGATGCTATGACTGCAGCAAGC  
 AGGTGAAAGGAATCATGTTGTTGCTGTCATAAGTGTGATGCTATGACTGCAGCAAGC

TTACGAAGCAATGAACAATGCCGATATCTTGATCTGATATGATTGTTTACTTAAITGA  
 TTACGAAGCAATGAACAATGCCGATATCTTGATCTGATATGATTGTTTACTTAAITGA

CAATAAACCAAGTTTCTCTGCCAATGCCAATC-3'  
 CAATAAACCAAGTTTCTCTGCCAATGCCAATC-3'  
*GhCLA1*-R5: 5' - GATTGCCAGTTGGCAGAG-3'

**B**

*GhVP*-F4: 5' - CAGTTAACCTGTGGATGTATCG-3'  
 At\_chr6 5' - CAGTTAACCTGTGGATGTATCGGTGAACAGGTGCAACATCCCTTTCTTTTACTGAATAT  
 Dt\_chr6 5' - CAGTTAACCTGTGGATGTATCGGTGAACAGGTGCAACATCCCTTTCTTTTACTGAATAT

CAGTATGTTGGCATCTTCATGATTGCTTTTGCTATTTCTGATTTTCCTCTTCTGGGCTCT  
 CAGTATGTTGGCATCTTCATGATTGCTTTTGCTATTTCTGATTTTCCTCTTCTGGGCTCT

GTCGAGGGCTTCAGCATGAAGAGCCAGCCTTGACCTATGATAAGGAGAAGATGTCAA  
 GTTGAAGGGCTTCAGCATGAAGAGCCAGCCTTGACCTATGATAAGGAGAAGATGTCAA

CCAGCTCTTCGCACTGCTATCTTCAGTACTGTATCCTTCTTCTGAGGTGCCATCACTCA  
 CCAGCTCTTCGCACTGCTATCTTCAGTACTGTATCCTTCTTCTGAGGTGCCATCACTCA

**GTCTTTCTGGATTCTTGG**AATGAAAATTGCTACTATGCCAATGCAAGAACACCTTG  
**GTCTTTCTGGATTCTTCTTGG**AATGAAAATTGCTACTATGCCAATGCAAGAACACCTTG

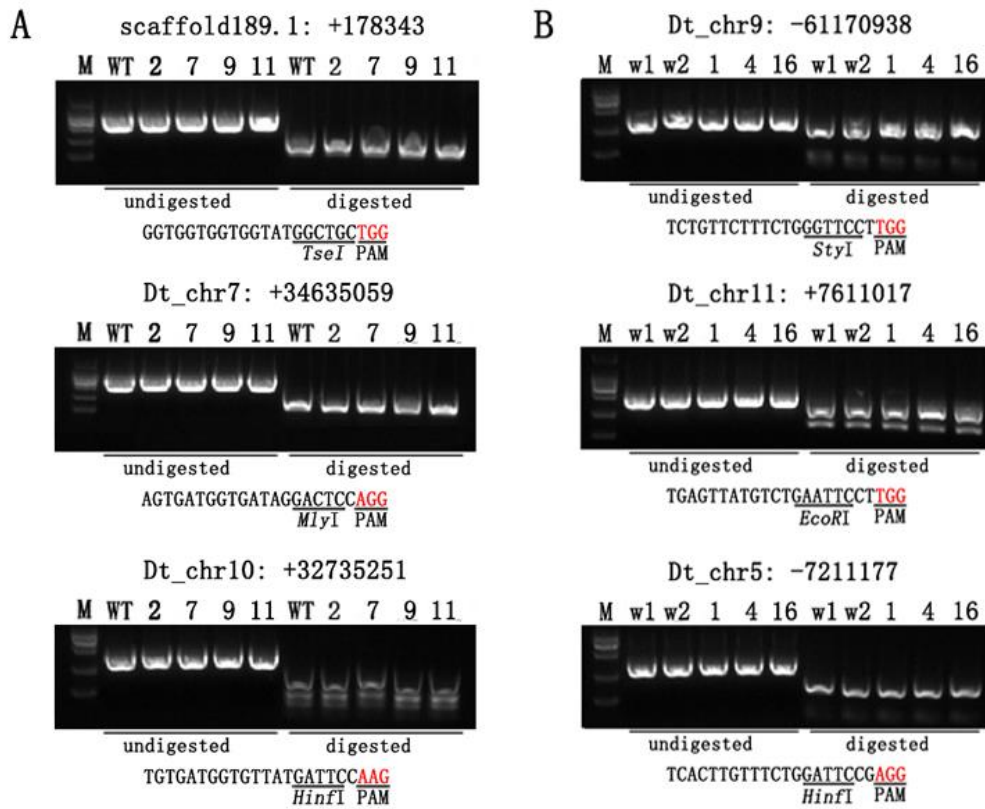
GAAGCAAGAAAAGGTGTGCGAAGGCTTTTATTGTTGCGTTTATGATCTGGTGCAGTAATG  
 GAAGCAAGAAAAGGCTGTGCGAAGGCTTTTATTGTTGCGTTTATGATCTGGTGCAGTAATG

GGTTTTCTCCTTGCAGCAATGGTTTATGGTACTTTACATTGCCAATCAATCTAATCAAG  
 GGTTTTCTCCTTGCAGCAATGGTTTATGGTACTTTACATTGCCAATCAATCTAATCAAG

CTGTACTATGGTATGACTGGGAAGCCCTTTTGGAGCTATTACTGGTTATGGTCTGGT  
 CTGTACTATGGTATGACTGGGAAGCCCTTTTGGAGCTATTACTGGTTATGGTCTGGT

GGATCCCTATGGCACTCTTTG-3'  
 GGATCCCTATGGCACTCTTTG-3'  
*GhVP*-R4: 5' - CAAAGAGTGCATAGAGGATC-3'

**Supplementary Fig. S1.** Identification of target sites and the primer sites for PCR amplification of DNA in the two different genomes. **(A)** Sequences of the PCR fragment of *GhCLA1*. **(B)** Sequences of the PCR fragment of *GhVP*. Red letters represent the target sites, blue letters represent the PAM sequence.



**Supplementary Fig. S2.** Examples of restriction-site loss assays with potential off-targets. **(A)** Analysis examples of potential off-targets from plants containing verified *GhCLA1* gene mutations. **(B)** Analysis examples of potential off-targets from plants containing verified *GhVP* gene mutations. The restriction enzyme sites are showed with an underscore. The PAMs are highlighted with red.

**Supplementary Table S1.** List of primers in this study.

Usage	Primer name	Sequence (5'-3')
Cloning	<i>Cas9</i> -F	TCCTGCAGACAGTGAAGGTG
	<i>Cas9</i> -R	GCCTTATCCAGTTCGCTCAG
Mutation detection	<i>GhCLA1</i> -F5	TTGGCCGGATTCACGAAA
	<i>GhCLA1</i> -R5	GATTGGCAGTTGGCAGAG
	<i>GhVP</i> -F4	CAGTTAACCTGTGGATGTATCG
	<i>GhVP</i> -R4	CAAAGAGTGCCATAGAGGATC

**Supplementary Table S2.** Target sequences of two genes in cotton and oligonucleotides used to express sgRNA in the vectors.

Target gene	Target site sequences	Oligo forward (5'-3')	Oligo reverse (5'-3')	Enzyme
<i>GhCLA1</i>	GGTGATGGTGCTATG <u>ACTGC</u> <b>AGG</b>	CAGGAAACAGCTATG ACCATATTCATTCGGA GTTTTTGTATC	gctattctagctctaaaacGCAGTC ATAGCACCATCACCaatcact acttcgactct	<i>Pst</i> I
<i>GhVP</i>	TCAGTTCTTTCTGG <u>ATTC</u> <b>CCTGG</b>	CAGGAAACAGCTATG ACCATATTCATTCGGA GTTTTTGTATC	gctattctagctctaaaacAGGAAT CCAGAAAGAACTGACaatc actacttcgactct	<i>Hinf</i> I

The restriction enzyme sites are showed with an underscore. The PAMs are highlighted with red.

**Supplementary Table S3.** List of potential off-target sites identified for the *GhCLA1* target sequence in the *G. hirsutum* genome.

Sequence of the putative off-target site	Putative off-target locus	Number of matching bases (include PAM)	Available restriction enzyme cutting site	Tested?
GGAGATGGAGCTATGACTGCTGG	At_chr10: -32691215	20	NO	YES
	Dt_chr8: -13606802	20	NO	YES
GGTGATGGAGCAATGACTGCTGG	Dt_chr13: +22584502	20	NO	YES
	Dt_chr13: +22620167	20	NO	YES
	scaffold72.1: -1097120	20	NO	YES
	scaffold72.1: -1130168	20	NO	NO
	scaffold72.1: -1164191	20	NO	NO
GGTGGTGGTGGTATGGCTGCTGG	scaffold189.1: +178343	19	TseI	YES
	At_chr9: -4372202	19	TseI	YES
AGTGATGGTGATAGGACTCCAGG	scaffold2376.1: -159740	19	MlyI	YES
	Dt_chr7: +34635059	19	MlyI	YES
GGTGATGGTGCTGGGCCTGCTGG	Dt_chr13: +16377109	19	Cac8I	YES
GGTGGTGGTGAATGACTGGGG	Dt_chr7: +14078964	18	BsrI	YES
TGTGATGGTGTATGATCCAAG	Dt_chr10: +32735251	18	Hinfi	YES
GGTGATGGTGGTGGCACTGCTGG	At_chr8: -16619447	18	BtsI	YES
GGTGATGGTGGTGGCAATGCTGG	Dt_chr13: -46985487	17	BsrDI	NO
GGTGATGGTGGATGACACCAGG	Dt_chr13: +20112017	19	NO	NO
	Dt_chr7: -19747713	19	NO	NO
GGTATGGTGTTCATGATTGCAGG	Dt_chr13: -15976702	19	NO	NO
GGTGATGGAGCAATGACGGCTGG	Dt_chr13: +22563464	19	NO	NO
GATGATGGTGCTAGGACTACAAG	Dt_chr4: -32253836	19	NO	NO
GGTGACTGTGCTAAGCCTGCTGG	Dt_chr7: -35412683	18	NO	NO
AGTGAGGTTGCTATGAGTGCAAG	Dt_chr1: -63423097	18	NO	NO
GGTGAGGTTGCTGTGGCTGCTGG	scaffold3130.1: +69411	18	NO	NO
GGTATGGTACTATGAGTGTAAAG	Dt_chr5: -59686925	18	NO	NO
GGTATGGCACTATGAGTGTAAAG	Dt_chr1: +11783798	18	NO	NO
	Dt_chr1: -80500846	18	NO	NO

The restriction sites are shown with an underscore; mismatching bases are shown in blue.

**Supplementary Table S3.** List of potential off-target sites identified for the *GhCLA1* target sequence in the *G. hirsutum* genome (continued).

Sequence of the putative off-target site	Putative off-target locus	Number of matching bases (include PAM)	Available restriction enzyme cutting site	Tested?
GGTGA <u>CT</u> GTGCTGAG <u>C</u> CTGCTGG	Dt_chr1: +65502164	17	NO	NO
	At_chr2: +58074573	17	NO	NO
	Dt_chr11: +25261498	17	NO	NO
	Dt_chr11: +54287243	17	NO	NO
	Dt_chr10: -39718103	17	NO	NO
	Dt_chr4: +25237184	17	NO	NO
GGTGA <u>CT</u> ATGCTAAG <u>C</u> CTGCTGG	At_chr2: +58103544	17	NO	NO
G <u>T</u> TG <u>C</u> TATGCTATGACTG <u>A</u> TGG	Dt_chr5: +1180457	17	NO	NO
G <u>T</u> TGATGGTGCCATAAGT <u>G</u> TGG	Dt_chr13: +21995540	17	NO	NO

The restriction sites are shown with an underscore; mismatching bases are shown in blue.

**Supplementary Table S4.** List of potential off-target sites identified for the *GhVP* target sequence in the *G. hirsutum* genome.

Sequence of the putative off-target site	Putative off-target locus	Number of matching bases (include PAM)	Available restriction enzyme cutting site	Tested?
TC <u>T</u> GTTCTTTCTGG <u>G</u> TTCC <u>T</u> TGG	Dt_chr9: -61170938	21	StyI	YES
TCAGTTCTTAT <u>T</u> GGATT <u>G</u> CTTGG	Dt_chr4: +27630888	20	Tth111II	YES
	Dt_chr4: +27631776	20	Tth111II	YES
	At_chr4: +87726589	20	Tth111II	YES
	At_chr4: +87727477	20	Tth111II	YES
	At_chr4: +87728332	20	Tth111II	NO
TCAG <u>T</u> CTTTGAGGATT <u>C</u> CTTGG	Dt_chr9: +42031687	20	Hinfl	YES
TGAGTTAT <u>G</u> TCTGA <u>A</u> TTCC <u>T</u> TGG	Dt_chr11: +7611017	19	EcoRI	YES
	At_chr4: +71425367	19	EcoRI	YES
TGAGTTAT <u>A</u> TCTGA <u>A</u> TTCC <u>T</u> TGG	At_chr11: -60320285	19	EcoRI	YES
TC <u>A</u> CTTGTTTCTGGATT <u>C</u> CGAGG	Dt_chr5: -7211177	19	Hinfl	YES
TCTGTT <u>G</u> TACTGGATT <u>C</u> CTTAG	At_chr6: +53150767	19	Hinfl	YES
TCAGATCATTCTT <u>G</u> ATTCC <u>A</u> CGG	At_chr12: -22967324	18	Hinfl	YES
	At_chr3: +27499491	18	Hinfl	YES
TCTGTT <u>G</u> TAACTGGATT <u>C</u> CTTAG	Dt_chr6: +11310880	18	Hinfl	YES
TCAGATCATTCTT <u>G</u> ATTCC <u>G</u> CGG	Dt_chr9: -42381100	18	Hinfl	YES
TC <u>C</u> TTTCTTTG <u>T</u> GGAT <u>G</u> CCT <u>C</u> GG	scaffold165.1: +486549	18	SfaNI	YES
TTAGTAGTTT <u>C</u> AGGATT <u>C</u> CTTAG	Dt_chr11: -41707429	18	Hinfl	YES
TCATTTCTT <u>C</u> GTAGAT <u>G</u> CCT <u>C</u> GG	At_chr2: -22997961	17	SfaNI	NO
TCATTTCTT <u>C</u> GCGGAT <u>G</u> CCT <u>C</u> GG	At_chr9: -47177313	17	SfaNI	NO
TC <u>C</u> TTTCTTTG <u>T</u> AGAT <u>G</u> CCT <u>C</u> GG	At_chr3: -41473876	17	SfaNI	NO
	Dt_chr1: +10912255	17	SfaNI	NO
	At_chr9: +49349713	17	SfaNI	NO
TCAGATCTTTCTT <u>G</u> TTCC <u>T</u> AGG	scaffold4724.1: -30530	19	NO	NO
TCATTTCTTTTGGATT <u>T</u> TAGG	Dt_chr8: +15208796	18	NO	NO
TAATTTGTTTCTGG <u>T</u> TCCTTAG	Dt_chr11: +12832761	18	NO	NO
TAATTTGTTTCTGG <u>T</u> TC <u>T</u> TAG	Dt_chr2: +13309509	17	NO	NO
	At_chr2: +60047189	17	NO	NO

The restriction sites are shown with an underscore; mismatching bases are shown in blue.