Supplemental information

Genome sequence of _Gossypium anomalum_ facilitates interspecific introgression breeding

Zhenzhen Xu, Jiedan Chen, Shan Meng, Peng Xu, Caijiao Zhai, Fang Huang, Qi Guo, Liang Zhao, Yonggang Quan, Yixin Shangguan, Zhuang Meng, Tian Wen, Ya Zhang, Xianggui Zhang, Jun Zhao, Jianwen Xu, Jianguang Liu, Jin Gao, Wanchao Ni, Xianglong Chen, Wei Ji, Nanyi Wang, Xiaoxi Lu, Shihong Wang, Kai Wang, Tianzhen Zhang, and Xinlian Shen

Supporting Information

Genome sequence of *Gossypium anomalum* facilitates interspecific introgression breeding

Zhenzhen Xu[1,5], Jiedan Chen[2,5], Shan Meng[1,5], Peng Xu[1], Caijiao Zhai[1], Fang Huang[1], Qi Guo[1], Liang Zhao[1], Yonggang Quan[3], Yixin Shangguan[1], Zhuang Meng[4], Tian Wen[3], Ya Zhang[4], Xianggui Zhang[1], Jun Zhao[1], Jianwen Xu[1], Jianguang Liu[1], Jin Gao[1], Wanchao Ni[1], Xianglong Chen[1], Wei Ji[1,4], Nanyi Wang[1], Xiaoxi Lu[1,4], Shihong Wang[3], Kai Wang[4,*], Tianzhen Zhang[2,*], Xinlian Shen[1,*]

[1]Key Laboratory of Cotton and Rapeseed (Nanjing), Ministry of Agriculture and Rural Affairs, the Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

[2]Institute of Crop Science, Plant Precision Breeding Academy, Zhejiang Provincial Key Laboratory of Crop Genetic Resources, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China

[3]JOIN HOPE SEEDS Co., Ltd., Changji, China

[4]Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops (MOE), Fujian Agriculture and Forestry University, Fuzhou, China

[5]These authors contributed equally to this article.

[*]Correspondence: Kai Wang (kwang5@126.com), Tianzhen Zhang (cotton@zju.edu.cn), Xinlian Shen (xlshen68@126.com)
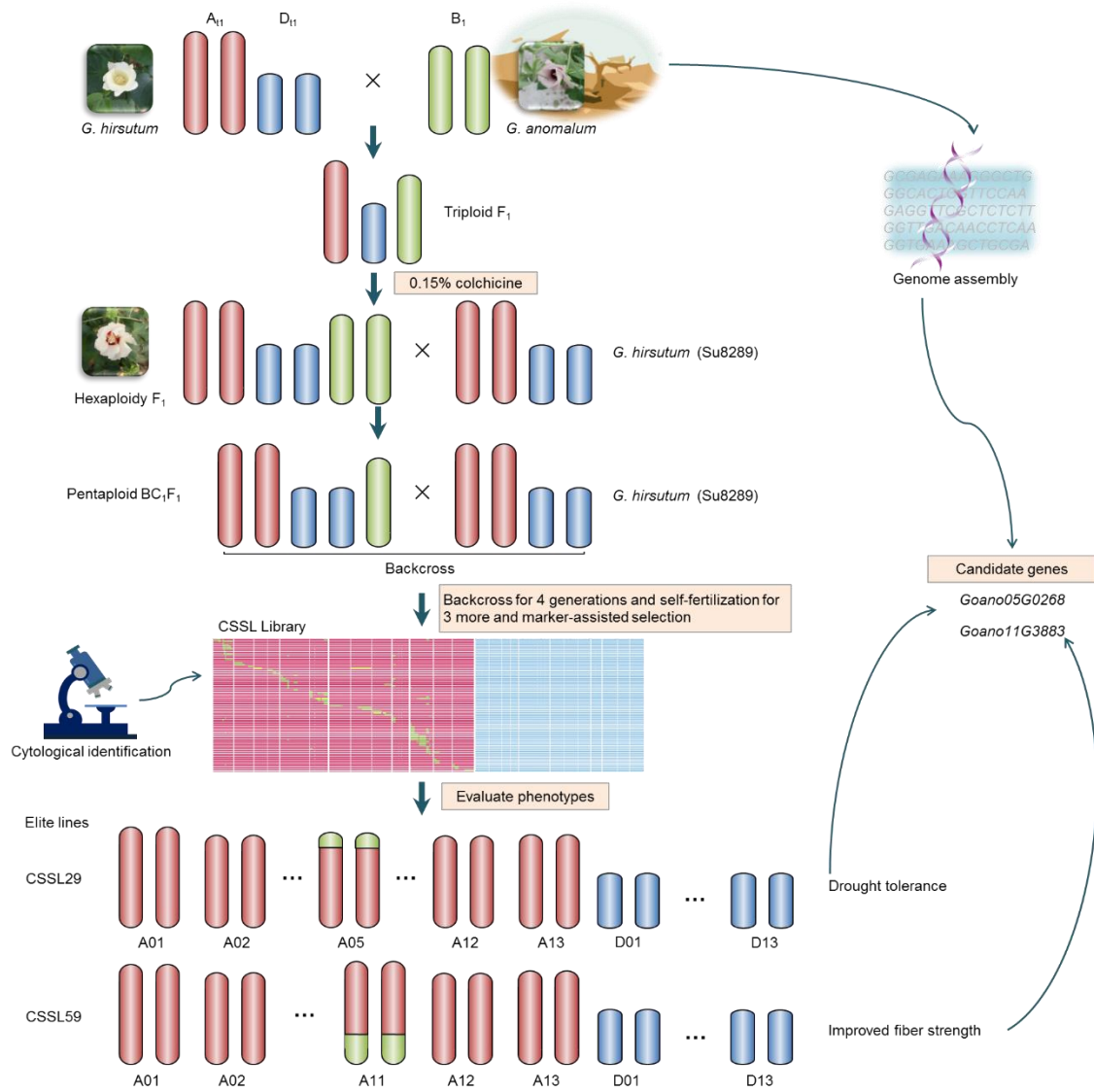
**Supplemental Figure 1.** Scheme for generating the CSSL population and rapidly discovering and cloning agronomic genes from *G. anomalum*.
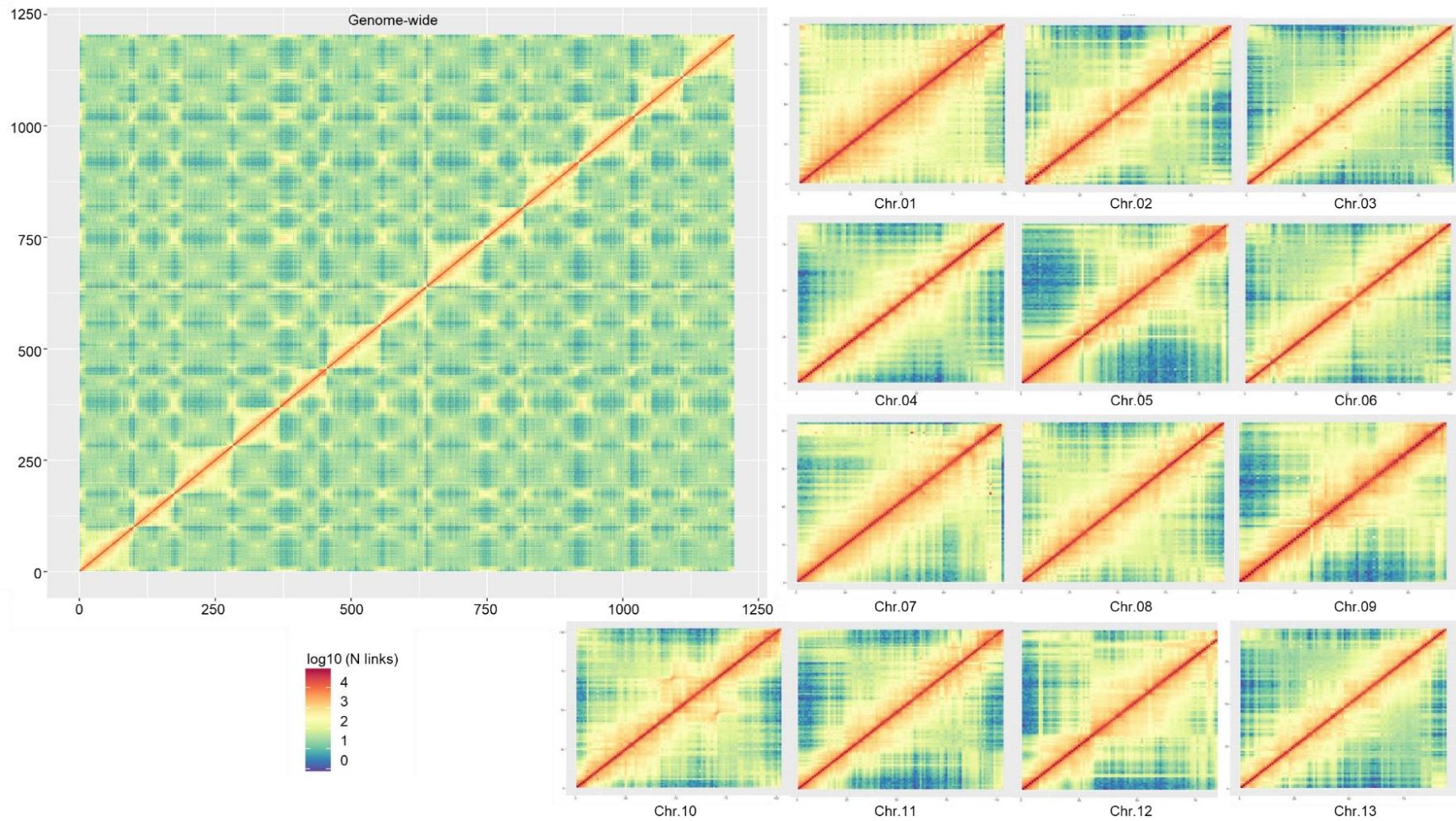
Supplemental Figure 2. Chromatin interaction frequency distribution of Hi-C links among chromosomes of *G. anomalum*. Strong signals (dark red dots) were distributed in diagonal lines, and this indicated that the scaffolds were accurately grouped and oriented on the pseudo-chromosomes.
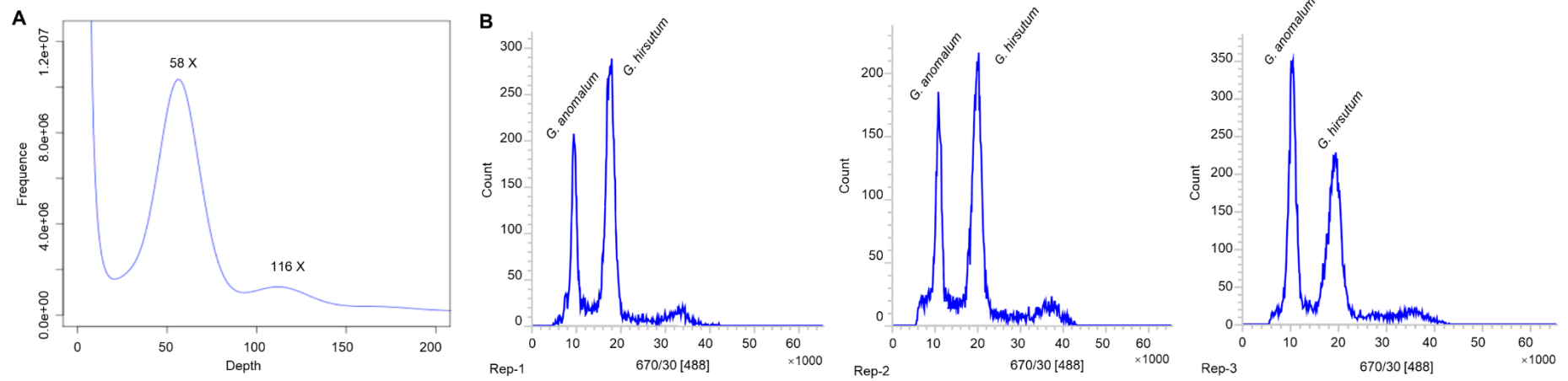
Supplemental Figure 3. (A) Distribution of sequencing depth and K-mer number frequency (K-mer = 17). **(B)** The schematic diagram of genome size determined by flow cytometry.
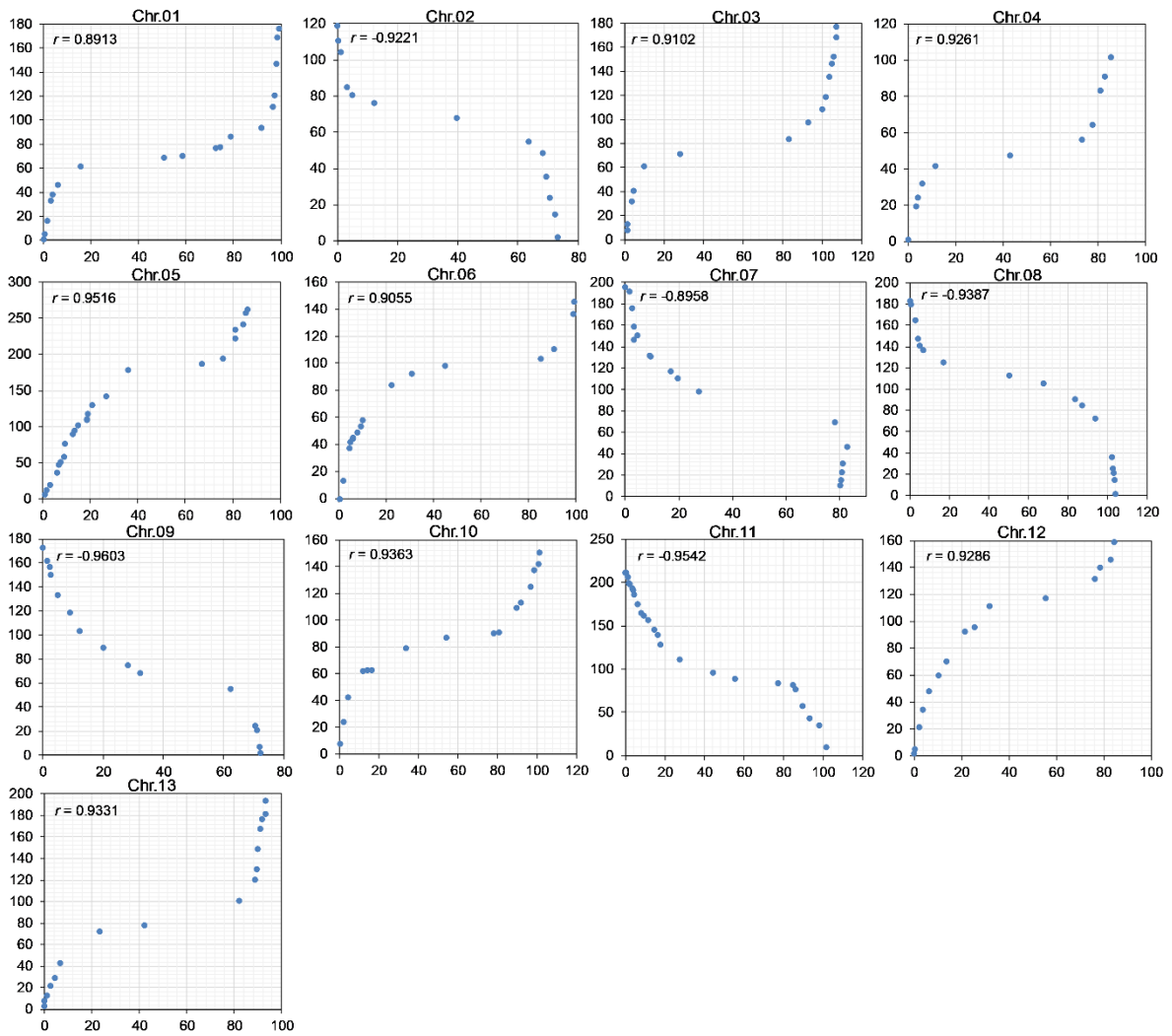
Supplemental Figure 4. Comparison of Hi-C directed chromosome assembly with genetic map in *G. anomalum*. The x-axes represent the physical positions in the assembled chromosomes (Mb) and the y-axes represent the genetic distance on the genetic map (cM). The *r* represents the correlation coefficient between the genetic and physical map distance.

|		Start (Mb)	End (Mb)	Size (Mb)
Chr.01		56.49	58.20	1.71
Chr.02		35.47	36.76	1.30
Chr.03		54.56	56.00	1.43
Chr.04		45.57	47.46	1.89
Chr.05		52.71	54.10	1.39
Chr.06		52.55	54.01	1.46
Chr.07		59.00	60.14	1.14
Chr.08		43.40	44.93	1.53
Chr.09		45.46	47.04	1.57
Chr.10		60.54	62.03	1.49
Chr.11		57.68	59.04	1.36
Chr.12		38.91	40.04	1.13
Chr.13		47.98	48.94	0.96
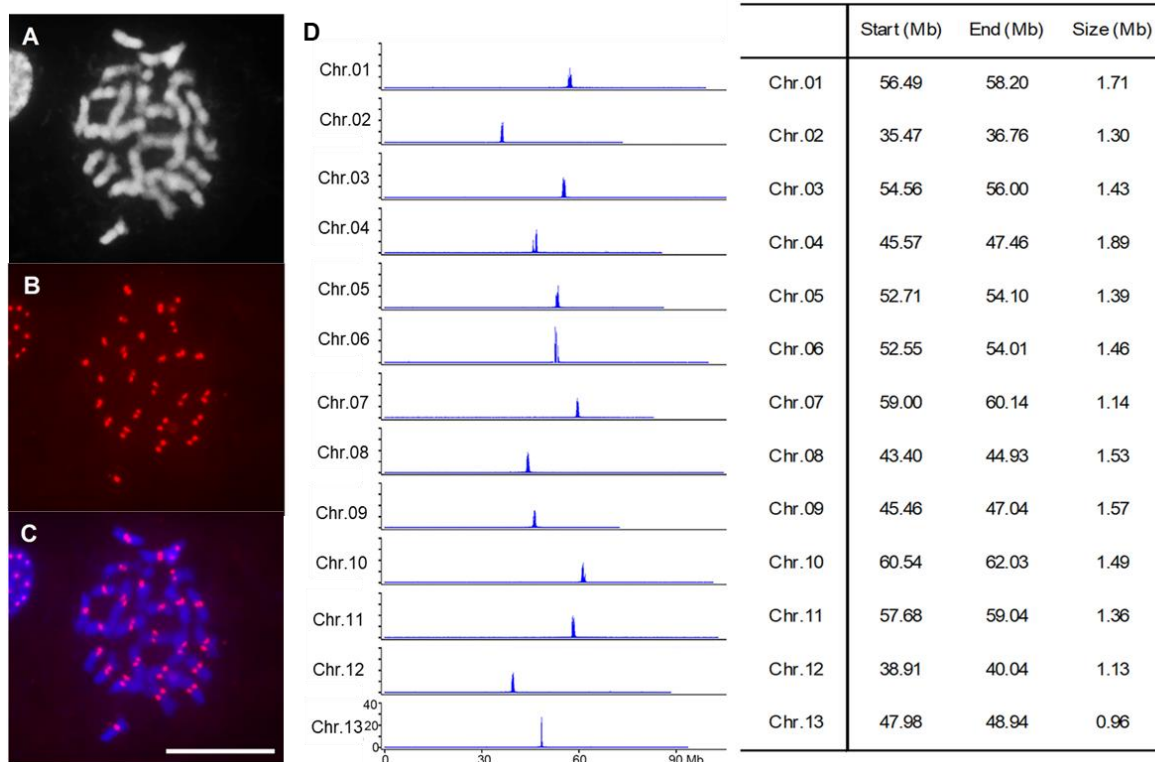
Supplemental Figure 5. Immunostaining using cotton anti-CenH3 antibodies in *G. anomalum*. Somatic metaphase chromosomes were stained with 4', 6-diamidino-2-phenylindole (DAPI) and showed in gray and blue in **(A)** and **(C)**, respectively. Immunofluorescence signals **(B)** are visible at the primary constrictions of the chromosomes **(C)**. Scale bar, 10 µm. **(D)** The centromere distribution in each chromosome. Left panel, CenH3 binding region i.e. centromere on each chromosome by CenH3 ChIP-seq mapping. The x- and y-axes indicate the position on the chromosome and the read density, respectively. Read density was represented by the percentage of mapped unique reads of respective 10-kb window. Right panel, the list of the size and spanning regions of each centromere.
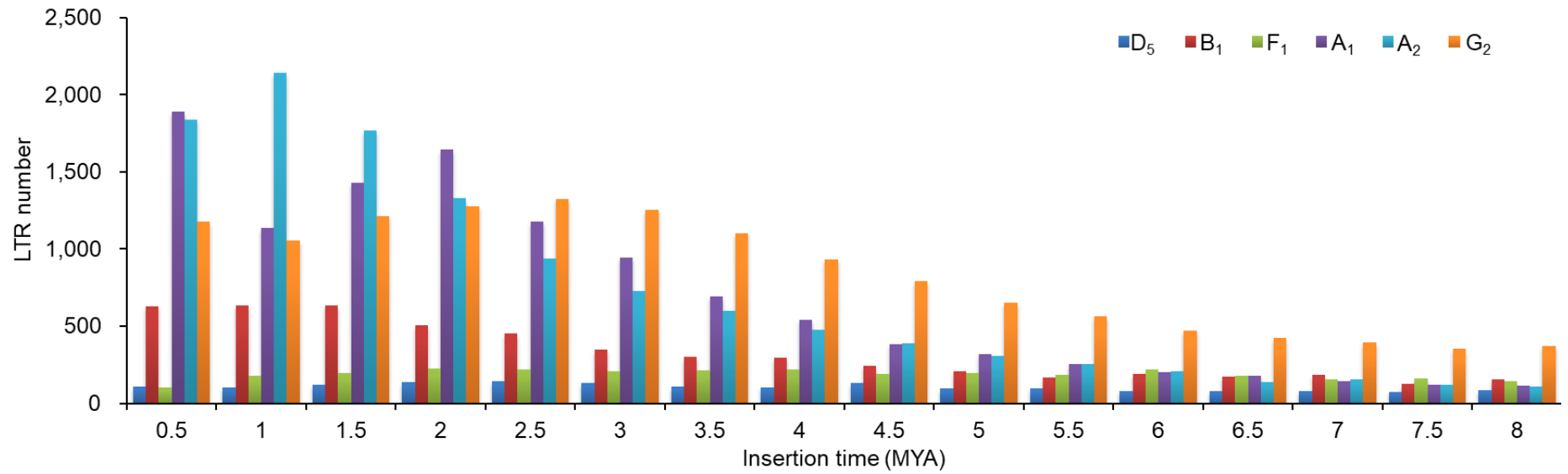
Supplemental Figure 6. Analysis of the LTR number and insertion time in D_5, B_1, F_1, A_1, A_2, and G_2 genome.
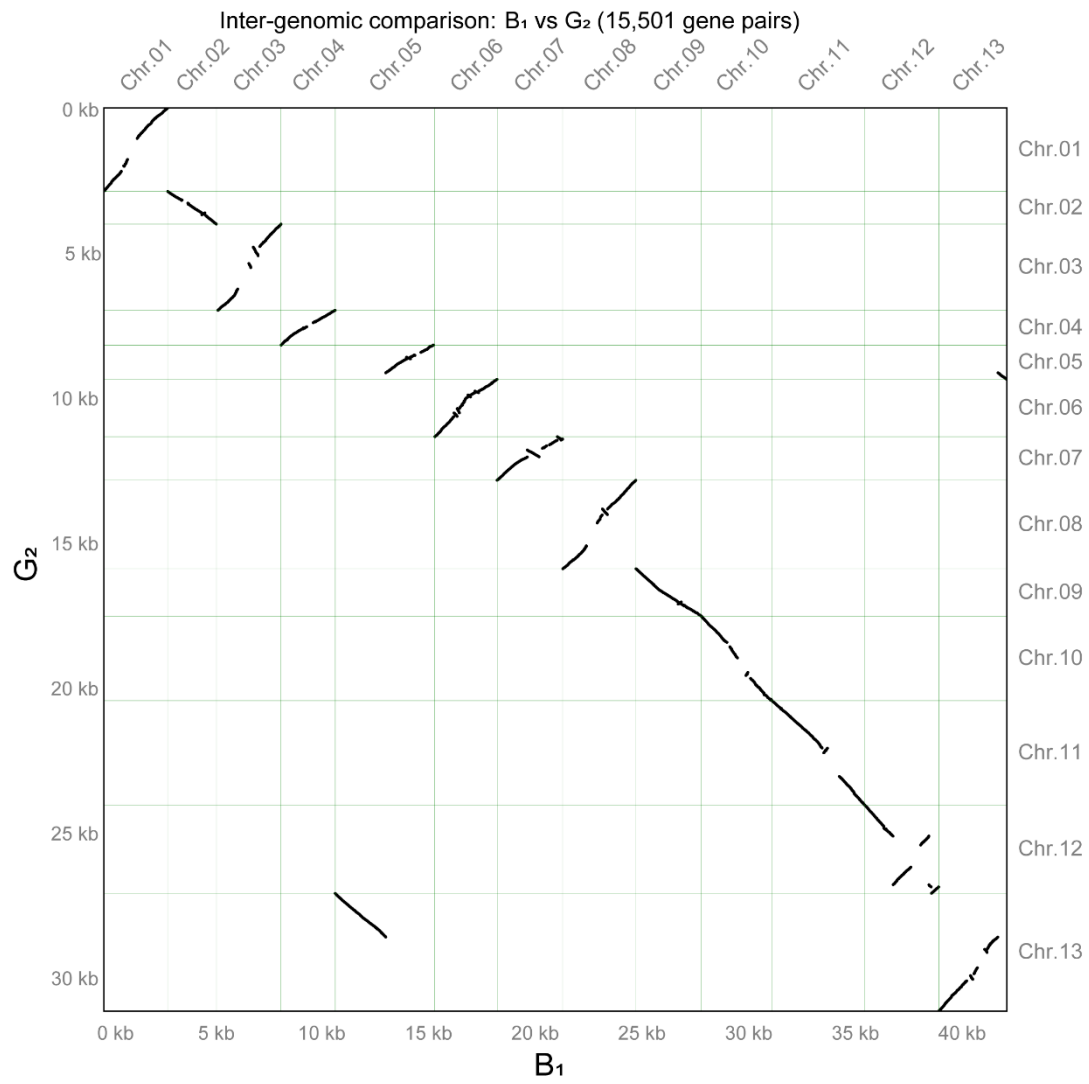
Supplemental Figure 7. Analysis of the synteny between *G. anomalum* and *G. austral* genome.
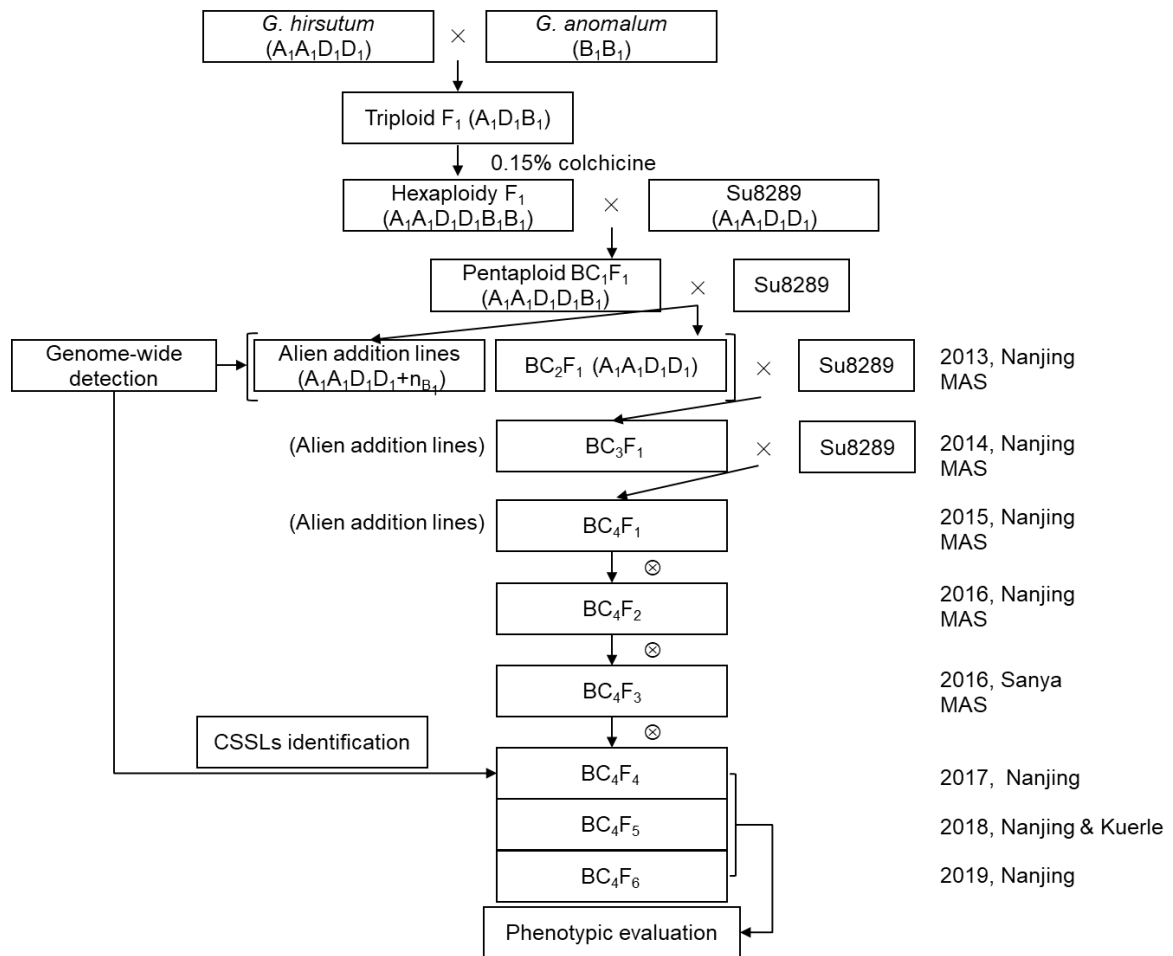
Supplemental Figure 8. Development of the CSSL population.
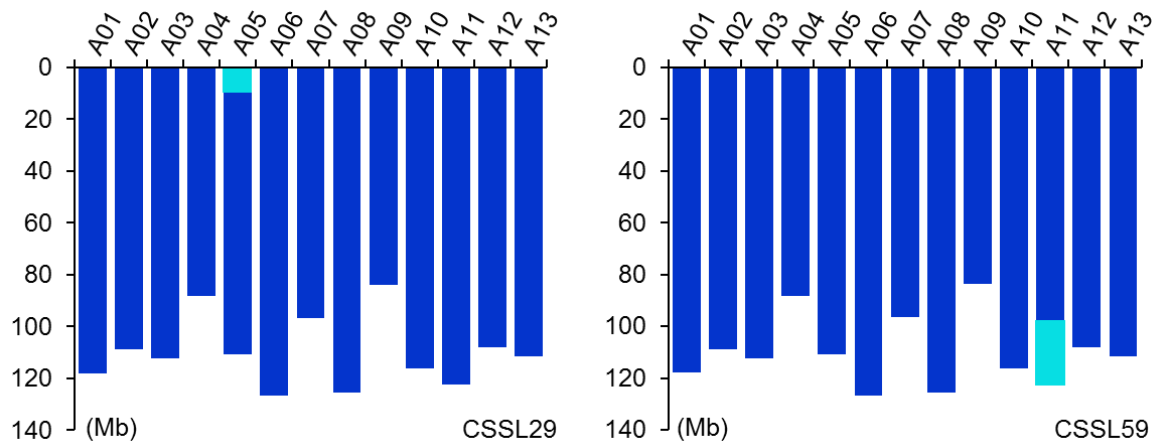
Supplemental Figure 9. Graphical genotype of CSSL29 and CSSL59 based on re-sequencing method. The introgression segments on A05 (from 0.02 Mb to 5.89 Mb) and A11 (from 104.49 Mb to 121.36 Mb) were shown by the light blue box, respectively.
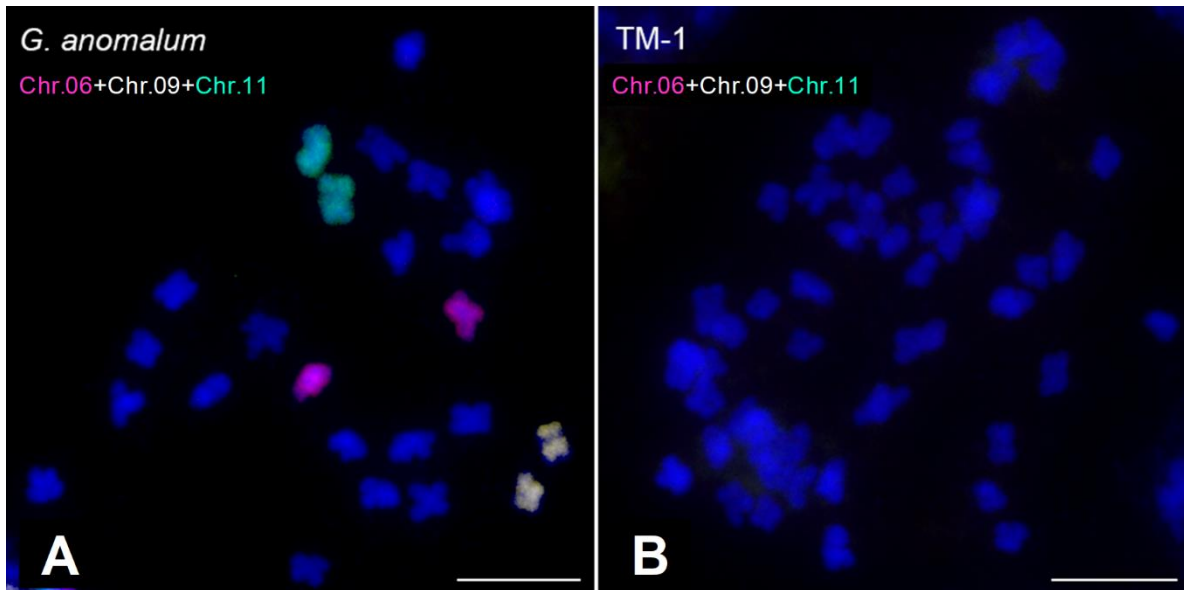
Supplemental Figure 10. FISH analyses using *G. anomalum*-specific oligo probes of chromosomes 6, 9, and 11. FISH image showing that *G. anomalum*-specific oligo probes of chromosomes 6, 9, and 11 were hybridized in the same mitotic metaphase cell of *G. anomalum* (**A**) and *G. hirsutum* TM-1 (**B**). The signals of the chromosomes 6, 9, and 11 probes are pseudocolored in magenta, white, and light blue, respectively. Scale bar, 10 µm.
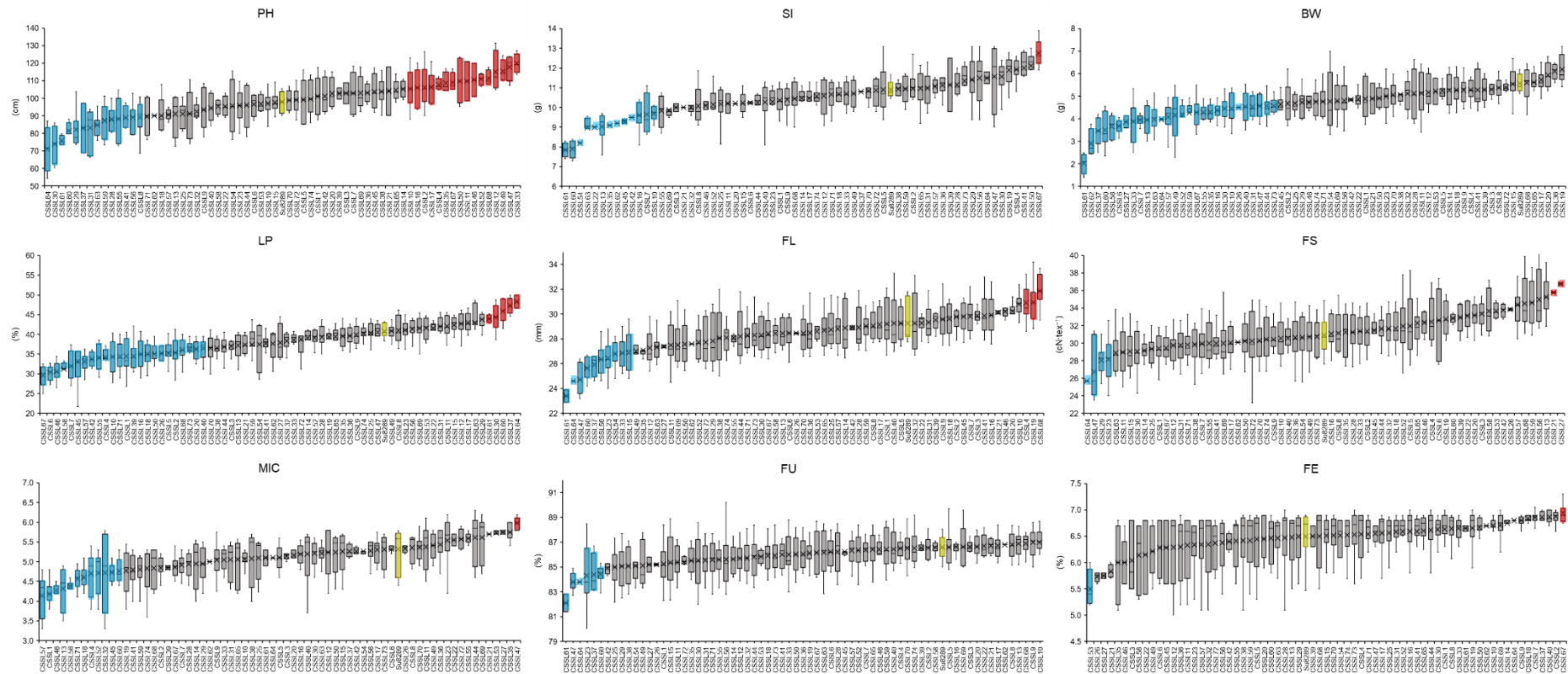
Supplemental Figure 11. Boxplots of phenotypic distribution of nine traits in the CSSL population. The horizontal axis indicates CSSL, and the vertical axis represents phenotypic value. The yellow graph represents the recurrent parent Su8289, the red graph indicates the CSSL with phenotypic value significantly higher ($P < 0.05$) than that of Su8289, while the blue graph indicates the CSSL with phenotypic value significantly lower ($P < 0.05$) than that of Su8289.
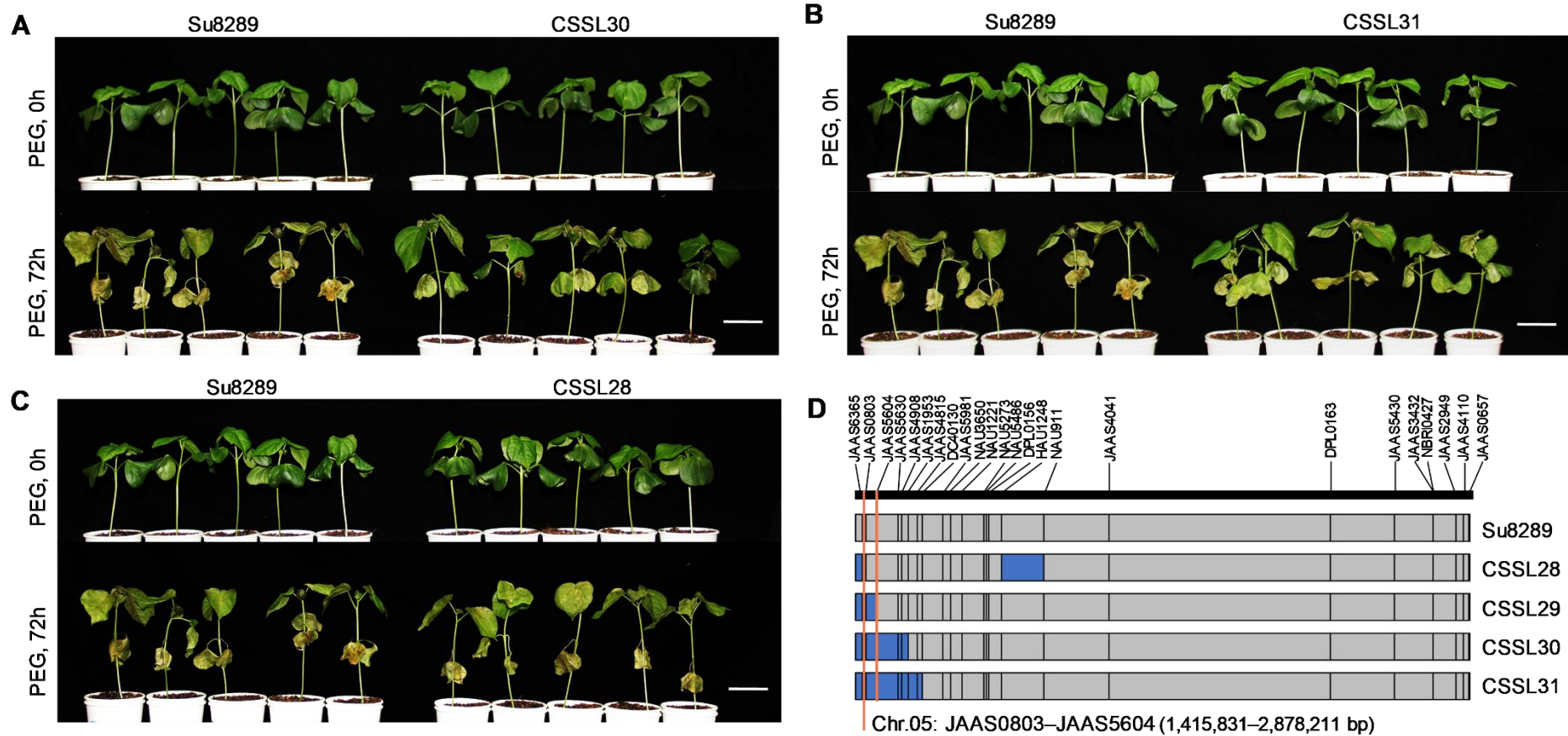
Supplemental Figure 12. Identification of valuable substitution segment associated with drought tolerance. (**A, B, and C**) Phenotypic comparison of CSSL30, CSSL31, CSSL28, and Su8289 cotton seedlings in response to drought stress in the form of PEG treatment. Photographs were taken at 0 h and 72 h after PEG treatment. Scale bars, 5 cm. (**D**) Graphical genotypes of SSR interval for drought stress on Chr.05.
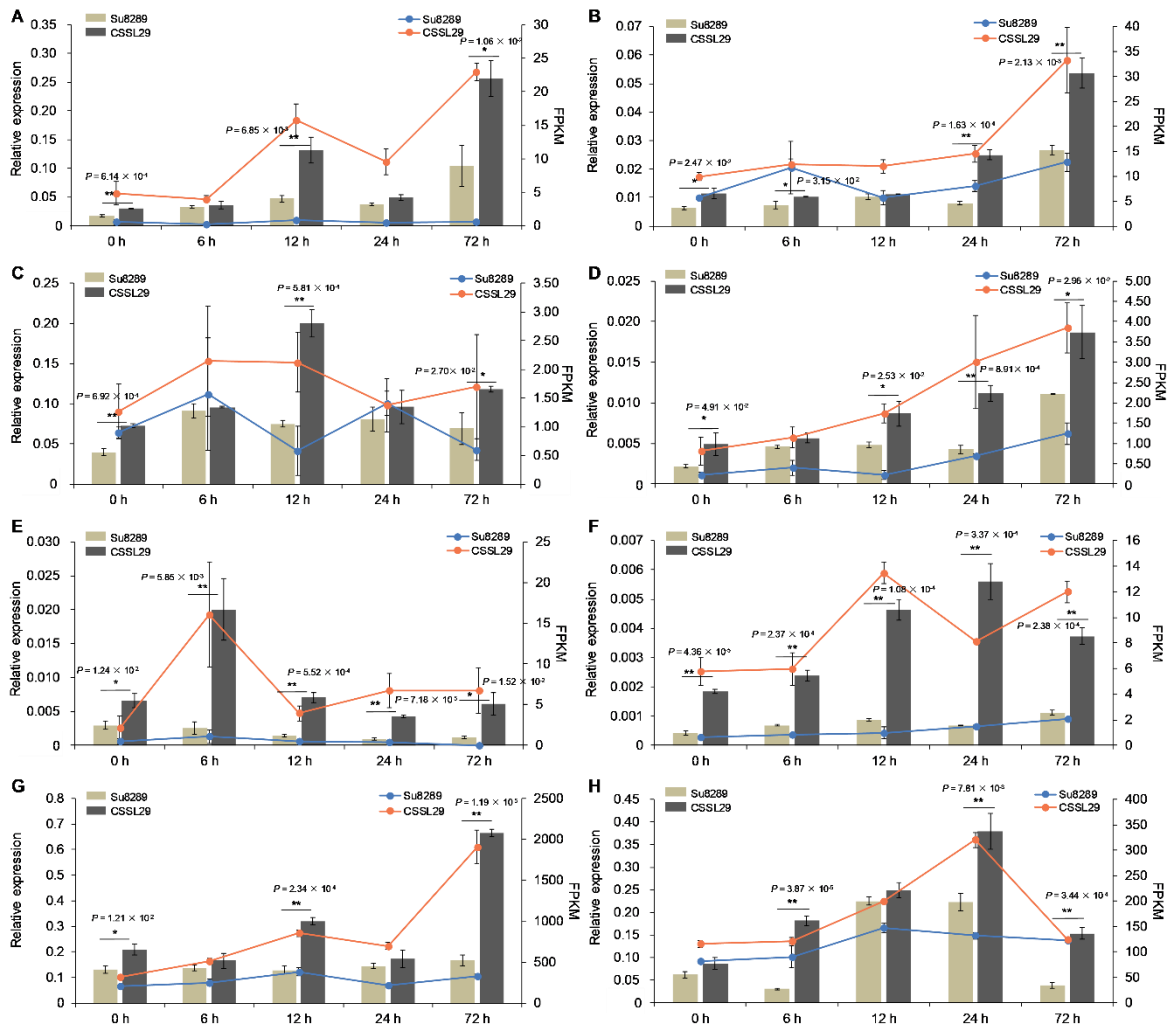
Supplemental Figure 13. Expression level of eight candidate genes under PEG stress in CSSL29 and Su8289. **(A-H)** Expression level of *Goano05G0164*, *Goano05G0170*, *Goano05G0207*, *Goano05G0222*, *Goano05G0235*, *Goano05G0236*, *Goano05G0301*, and *Goano05G0319* in CSSL29 and Su8289 under PEG stress at 0, 6, 12, 24, and 72 hours. The left y-axis shows the relative expression obtained from qRT-PCR, and the right y-axis shows the FPKM value obtained from RNA-Seq data. *$P < 0.05$, **$P < 0.01$, Student's *t* test.
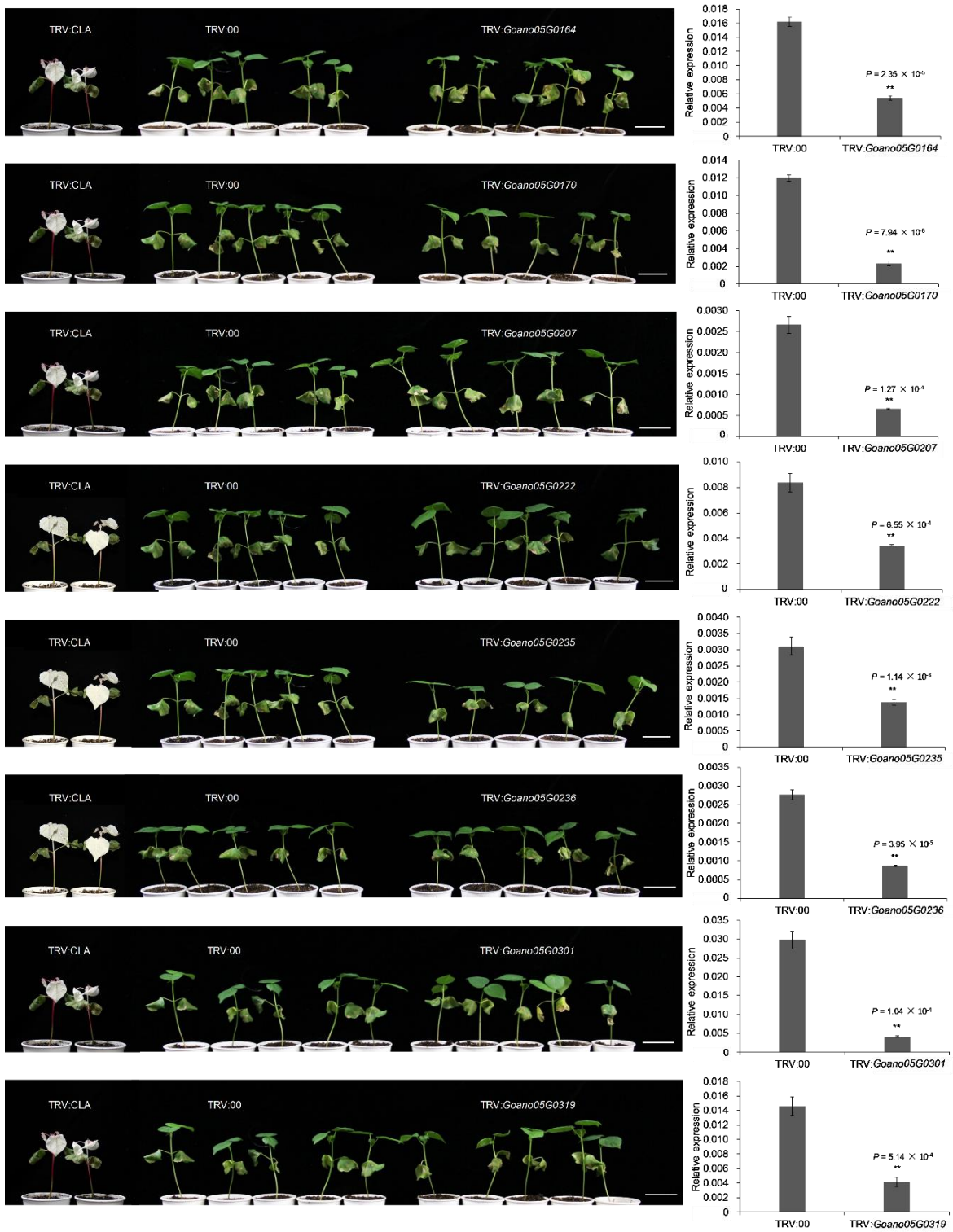
Supplemental Figure 14. Functional verification of eight candidate genes by VIGS. Scale bars, 5 cm. **P* < 0.01, Student's *t* test.
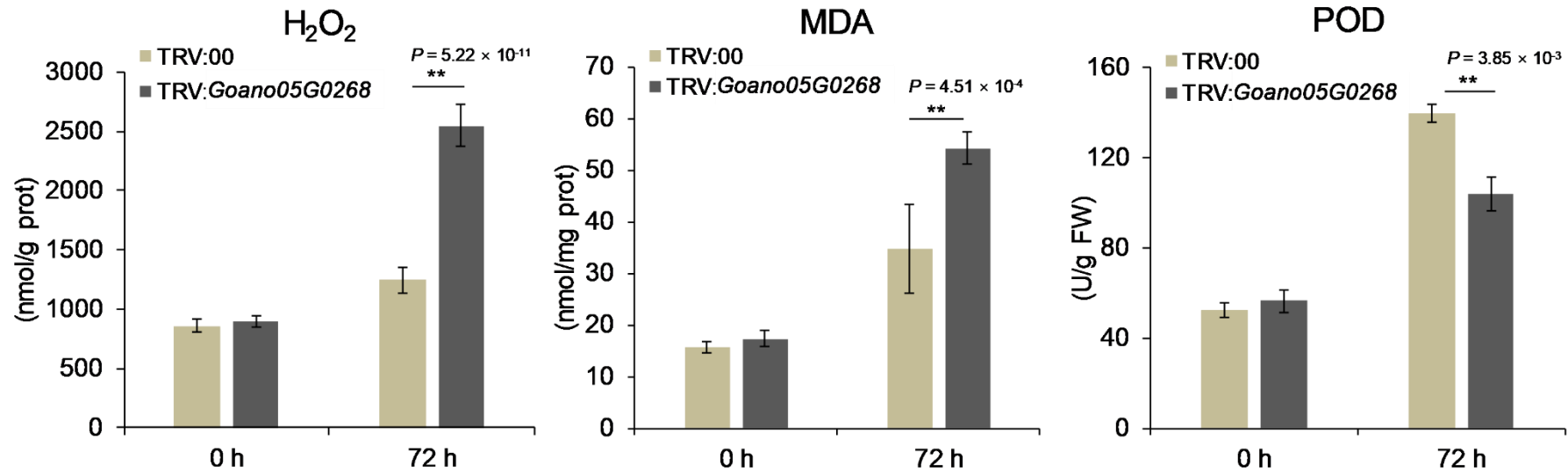
Supplemental Figure 15. Endogenous H_2O_2, MDA, and POD content in TRV:00 and TRV:*Goano05G0268* cotton plants under PEG stress. **$P < 0.01$, Student's *t* test.
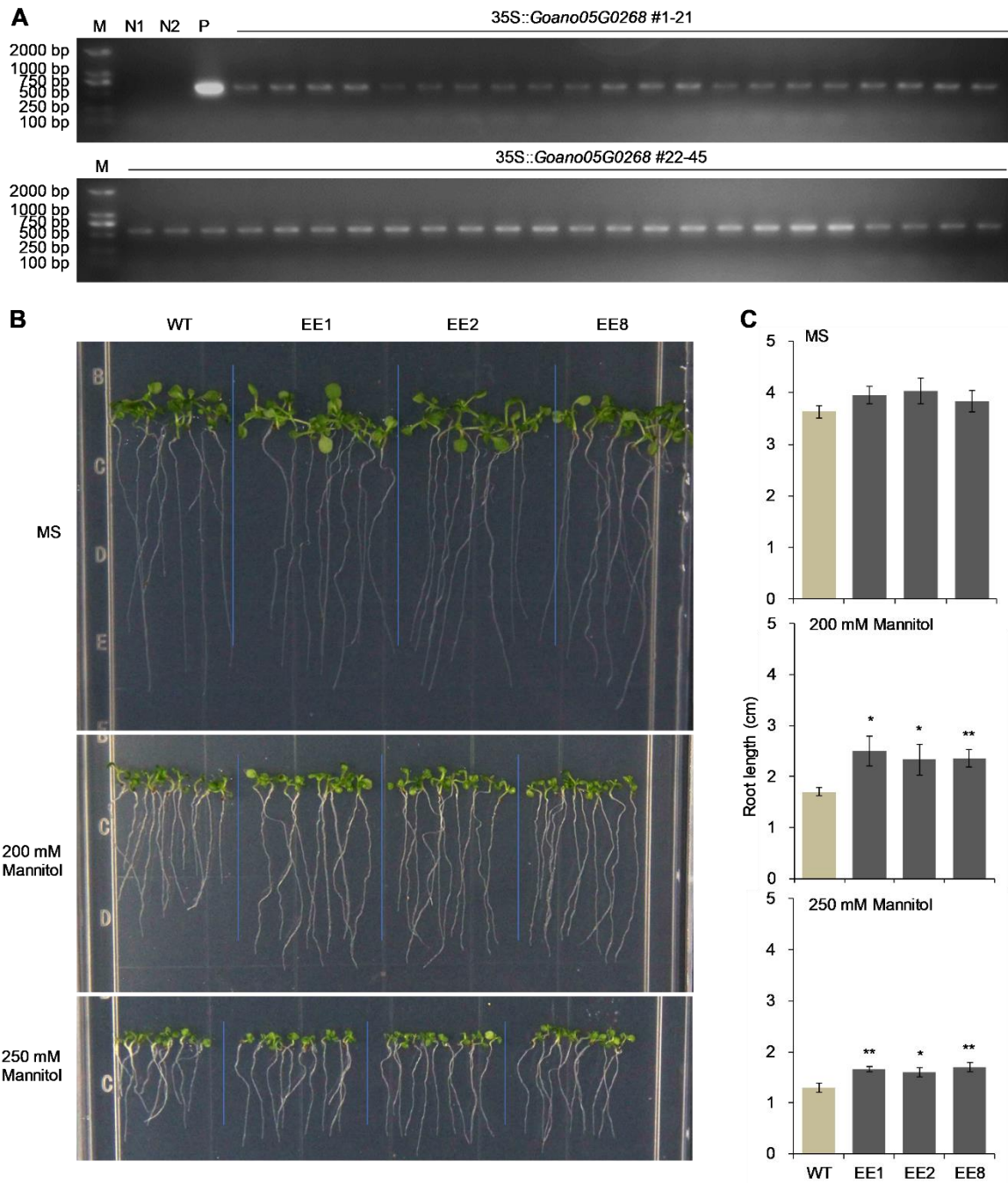
Supplemental Figure 16. Ectopic expression of *Goano05G0268* modulates drought response in *Arabidopsis*. **(A)** Identification of the EE transgenic lines by PCR amplification. M: DNA marker, N1: H_2O control, N2: WT, P: 35::*Goano05G0268* plasmid. **(B and C)** Root elongation of seedlings grown on MS medium with 0, 200, and 250 mM mannitol after eight days. *$P < 0.05$, **$P < 0.01$, Student's *t* test.
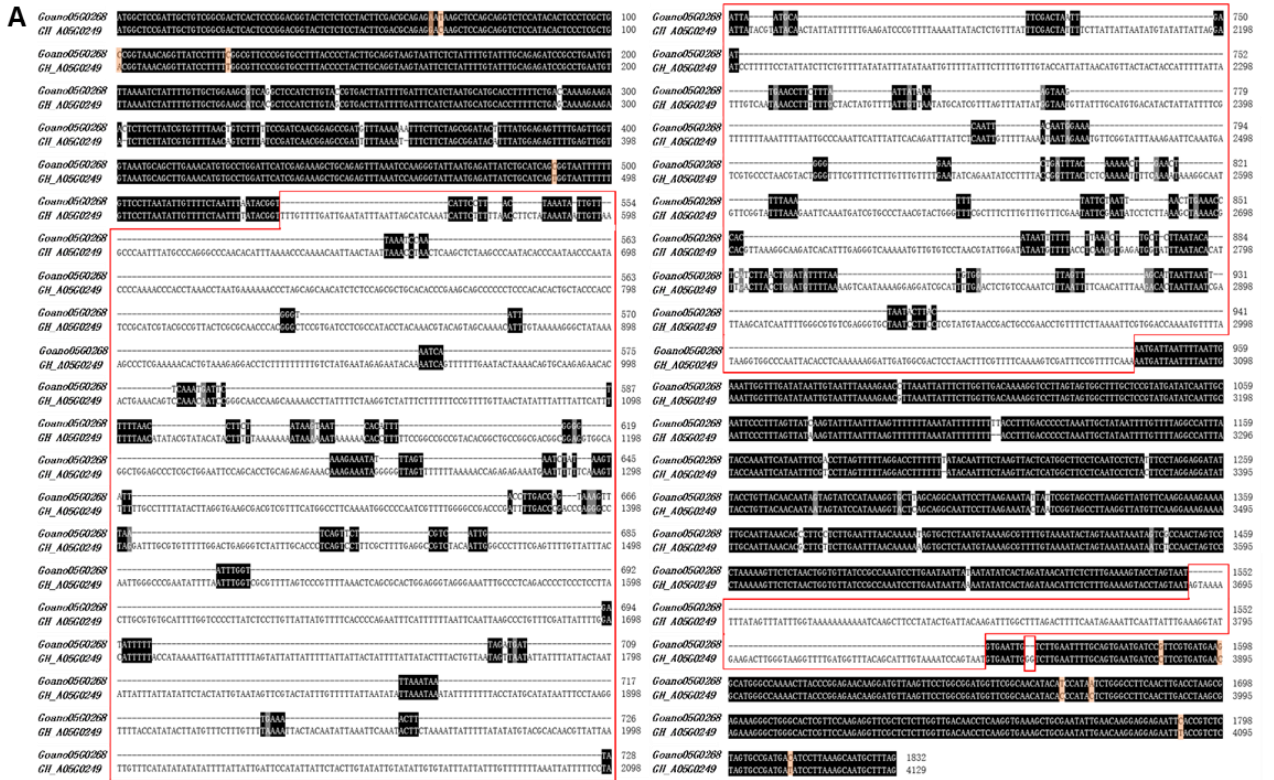
Supplemental Figure 17. (A and B) Structural variation of *Goano05G0268* and *GH_A05G0249* by PCR amplification in CSSL 29 and Su8289, respectively. Sequences highlighted in red boxes are structural variation sites **(A)**. The blue lines represent structural variation sites in exons **(B)**.
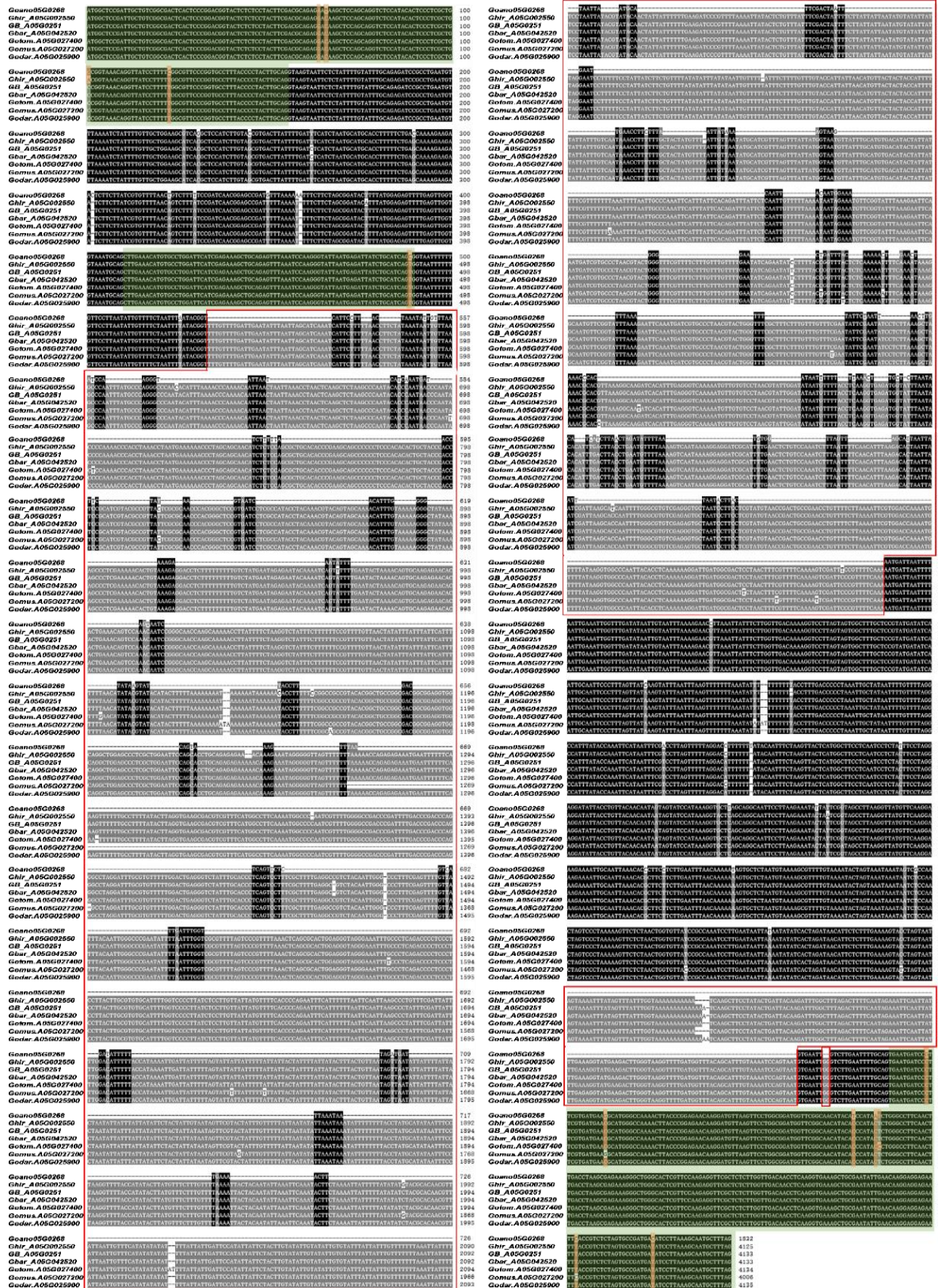
Supplemental Figure 18. Structural variation of *Goano05G0268* and its orthologous gene in *G. hirsutum* L. acc. TM-1, *G. barbadense* L. cv. Hai7124, *G. barbadense* L. accession 3–79, *G. tomentosum*, *G. mustelinum*, and *G. darwinii*, respectively. Sequences highlighted in red boxes are structural variation site.
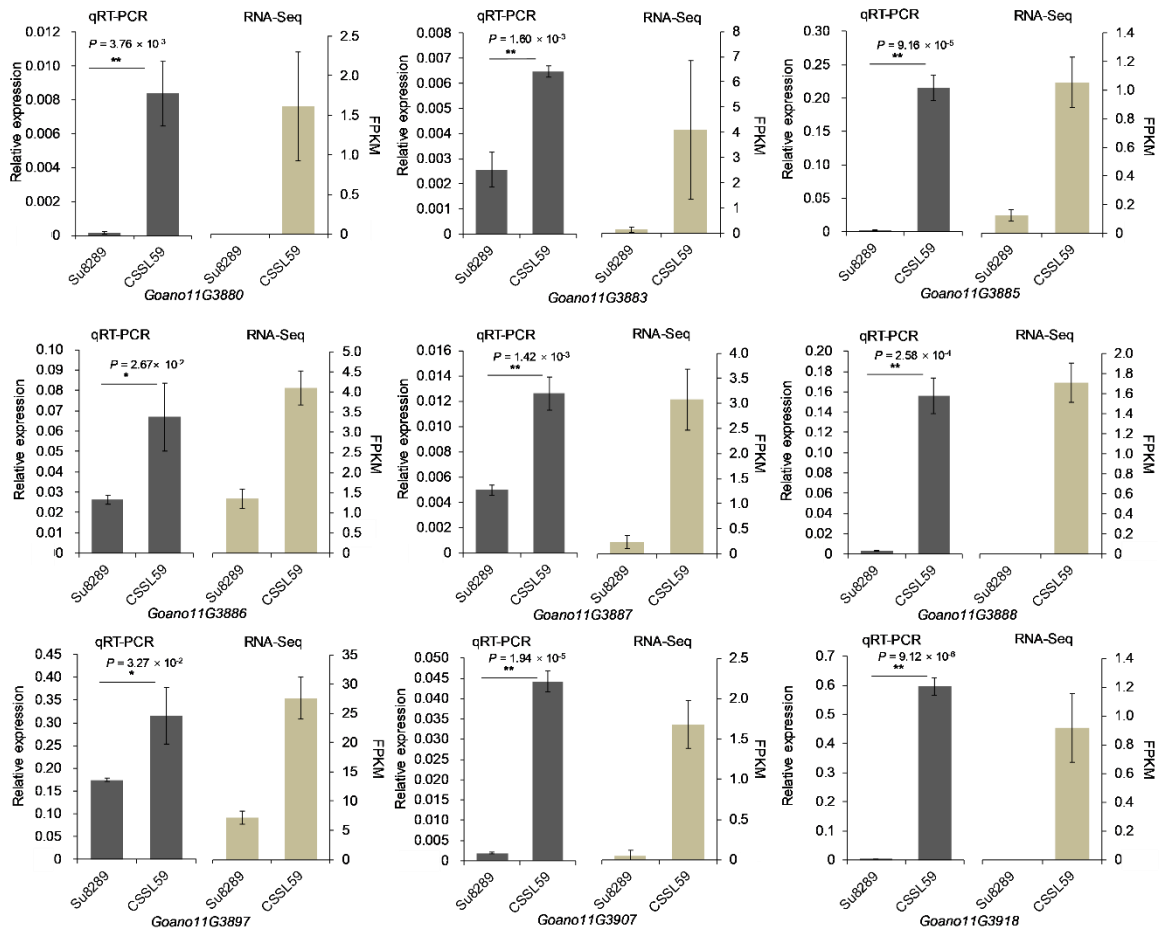
Supplemental Figure 19. The expression level of nine genes in fibers at 20 DPA of CSSL59 and Su8289. The left y-axes show the relative expressions according to qRT-PCR, and the right y-axes show the FPKM values obtained from RNA-seq data. $**P < 0.01$, Student's t test.
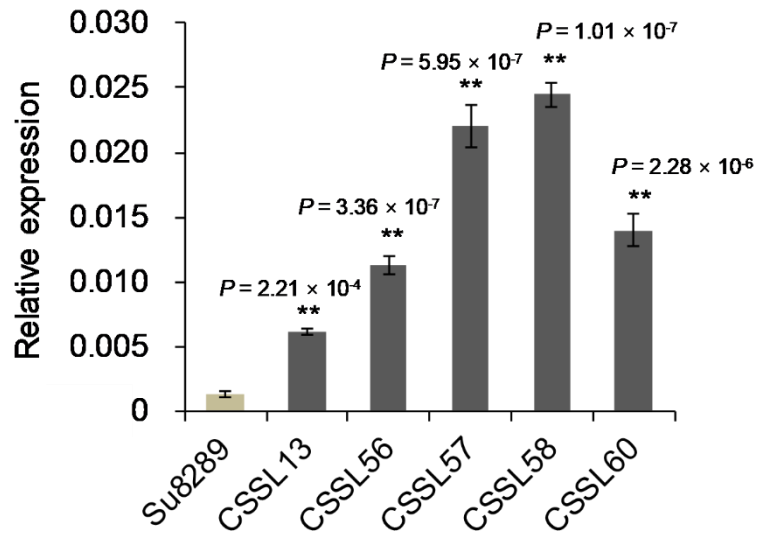
Supplemental Figure 20. Expression level of *Goano11G3883* in fibers at 20 DPA from Su8289, CSSL13, CSSL56, CSSL57, CSSL58, and CSSL60. **$P < 0.01$, Student's *t* test.
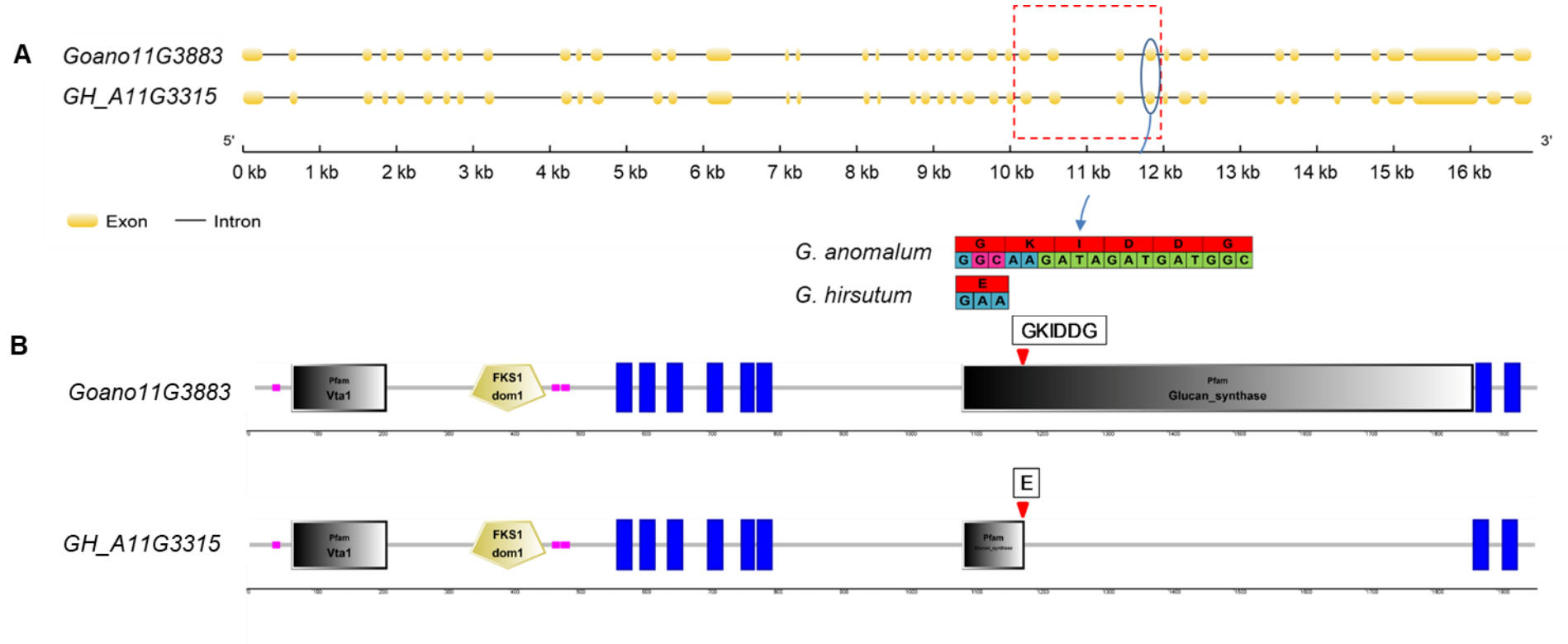
Supplemental Figure 21. (A and B) Structural and glucan_synthase domain variation of *Goano11G3883* and *GH_A11G3315* in *G. anomalum* and *G. hirsutum* TM-1, respectively.
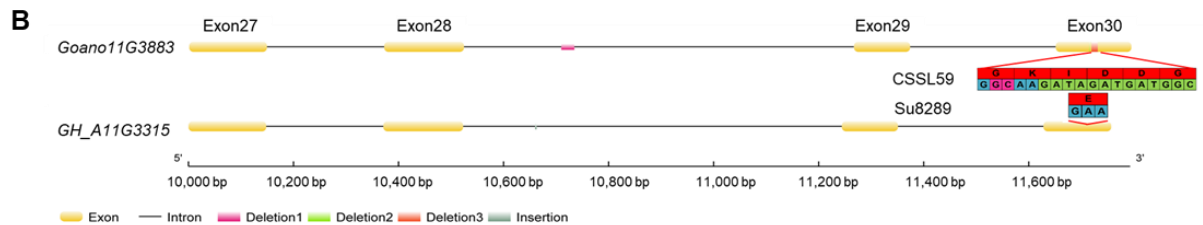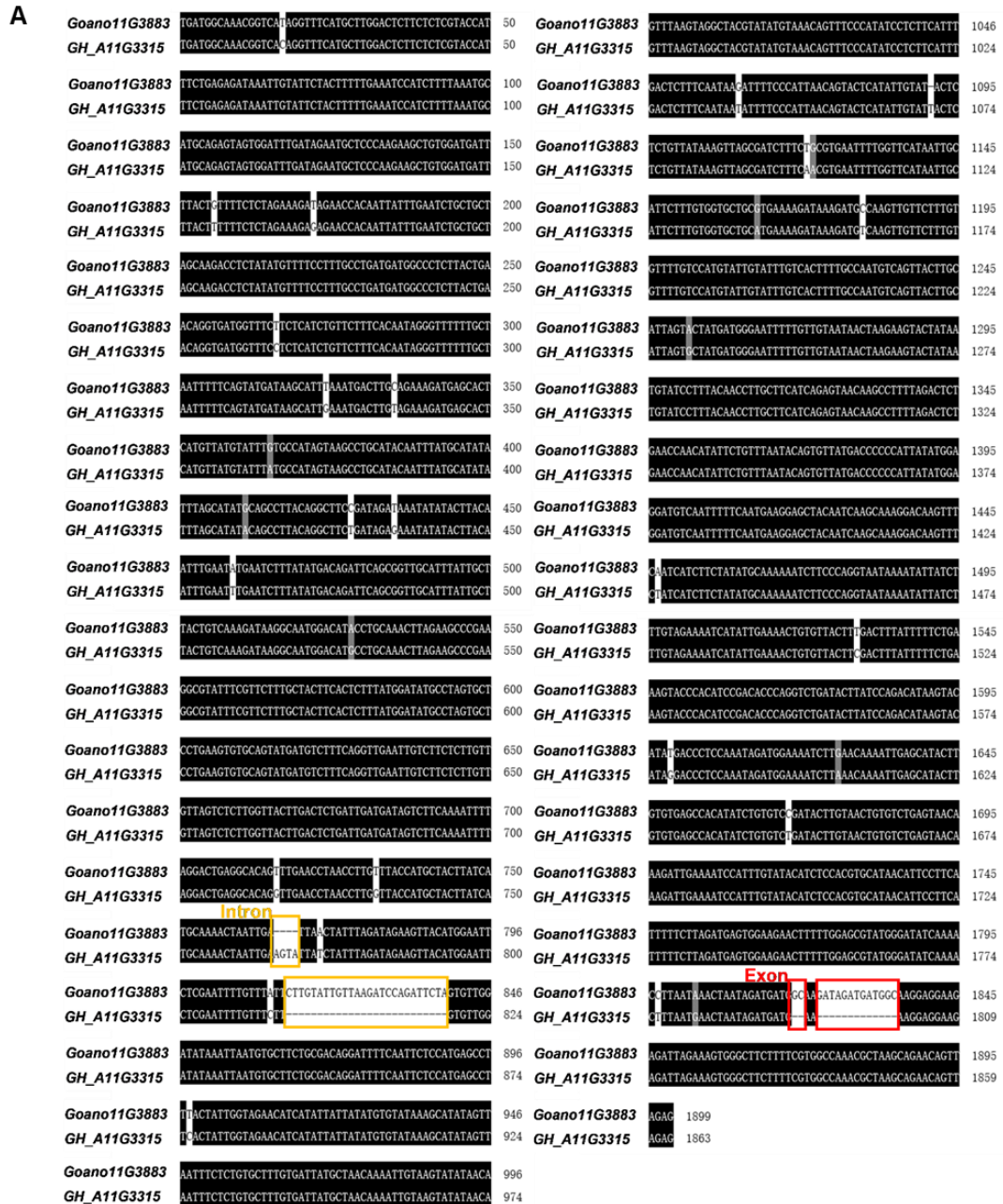
Supplemental Figure 22. (A and B) Structural variation of *Goano11G3883* and *GH_A11G3315* by PCR amplification in CSSL59 and Su8289, respectively. Sequences highlighted in yellow and red boxes are structural variation sites in intron and exon, respectively **(A)**.

Supplemental Table 1. Statistic of sequencing data from the PacBio Sequel and Illumina platform for *G. anomalum*.

Type of sequencing libraries	Insert size	Read length (bp)	Read number	Total data (Gb)	Sequence depth (×)
PacBio	40 kb	133,688 (Max) 9,561 (Mean) 16,309 (N50)	8,647,163	82.68	64.09
Illumina	230 bp	150	155,987,099	46.60	36.12
	350 bp	150	137,948,704	41.33	32.04
	500 bp	150	149,730,393	44.68	34.64
Total				132.61	102.80

The statistic was based on the filtered data in PacBio library and clean data in Illumina library. Sequence depth was calculated according to the genome size of 1.29 Gb.

Supplemental Table 2. Summary of genome assemblies from PacBio to Hi-C for *G. anomalum*.

Category	PacBio (Contig)	BioNano (Contig)	BioNano (Scaffold)	Hi-C (Contig)	Hi-C (Scaffold)
Contig/Scaffold number	611	687	396	685	364
Contig/Scaffold length (bp)	1,202,351,677	1,202,351,677	1,208,495,100	1,202,351,483	1,208,248,306
Contig/Scaffold Max (bp)	25,786,769	25,786,769	99,134,223	25,786,769	107,236,083
Contig/Scaffold N50 (bp)	7,783,093	6,738,719	73,503,624	6,738,719	99,188,525
Contig/Scaffold N90 (bp)	2,031,882	1,724,051	21,629,783	1,724,051	73,503,624
Gap size (bp)	0	0	6,143,423	0	5,896,823

Supplemental Table 3. Summary of BioNano data quality and assembly results for *G. anomalum*.

Category	*G. anomalum*
BioNano data quality	
Filtering threshold	150 kb
Enzyme	DLE-1
Molecule number	1,153,134
Total length (Mb)	291,883
Average length (kb)	253
Molecule N50 (kb)	255
Label density (/100 kb)	13.980
Genome coverage ($\times$)	242.800
BioNano assembly results	
Number Genome Maps	144
Total Genome Map Length (Mb)	1,290.326
Mean Genome Map Length (Mb)	8.961
Median Genome Map Length (Mb)	0.746
Genome Map N50 (Mb)	58.216

Supplemental Table 4. Summary of Hi-C mapping data for *G. anomalum.*

Category	Number of reads	Percentage (%)	
Unique paired alignments	249,816,916	55.27	
Valid read pairs	240,410,680	53.19	
Invalid read pairs	9,406,236	2.08	
Dangling end pairs	6,689,797		1.48
Re-ligation pairs	2,410,771		0.53
Self-cycle pairs	303,540		0.07
Dumped pairs	2,128		0.00
Multiple pairs alignments	131,108,882	29.01	
Pairs with singleton	50,686,398	11.21	
Unmapped pairs	20,353,034	4.50	
Total read pairs	451,965,230	100.00	

Supplemental Table 5. Characteristics of the 13 pseudo-chromosomes in the final genome assembly of *G. anomalum*.

Pseudo-chromosome	Number of scaffolds[a]	Size (bp)
Chr.01	3	99,188,525
Chr.02	1	73,503,624
Chr.03	3	107,236,083
Chr.04	1	85,652,033
Chr.05	3	86,126,817
Chr.06	10	99,994,996
Chr.07	4	83,060,063
Chr.08	3	104,733,974
Chr.09	2	72,545,696
Chr.10	2	101,529,486
Chr.11	4	102,973,240
Chr.12	3	88,411,662
Chr.13	2	93,713,888
Total	41	1,198,670,087

[a]The scaffolds here were obtained from the BioNano assembly and some of them were truncated by others in the following Hi-C process.

Supplemental Table 6. Summary of genome base content for *G. anomalum*.

Base type	Number (bp)	Percentage of genome (%)
A	395,389,494	32.72
T	395,196,517	32.71
C	206,049,696	17.05
G	205,715,776	17.03
N	5,896,823	0.49
Total	1,208,248,306	100.00
G + C	411,765,472	34.25[a]

[a]GC content of the genome without N.

Supplemental Table 7. Evaluation of genome assembly integrity of *G. anomalum* using the BUSCO method.

BUSCOs	Number	Percentage (%)
Complete BUSCOs (C)	2,303	99.01
Complete single-copy BUSCOs (S)	2,086	89.68
Complete duplicated BUSCOs (D)	217	9.33
Fragmented BUSCOs (F)	10	0.43
Missing BUSCOs (M)	13	0.56
Total BUSCO groups searched	2,326	100

BUSCO: Benchmarking Universal Single-Copy Ortholog.

Supplemental Table 8. Evaluation of genome assembly integrity of *G. anomalum* using the CEGMA method.

	Complete + Partial CEGMAs		Total CEGMA groups searched	
	number	Percentage (%)	number	Percentage (%)
G. anomalum	242	97.58	248	100
G. anomalum (Grover et al., 2021)	243	97.98	248	100

CEGMA: Core Eukaryotic Genes Mapping Approach.

Supplemental Table 9. Summary of Illumina short reads mapping in *G. anomalum* genome.

	Category	Percentage (%)
Reads	Mapping rate	97.25
	Average sequencing depth	96.93
	Coverage	99.90
Genome	Coverage at least 4×	99.84
	Coverage at least 10×	99.74
	Coverage at least 20×	99.42

Supplemental Table 10. The sequence coverage of the *G. anomalum* genome by homologous search using RNA-seq data.

Dataset	Number	Total length (bp)	Bases covered by assembly (%)	Sequences covered by assembly (%)	with > 90% sequence in one scaffold		with > 50% sequence in one scaffold	
					Number	Percent	Number	Percent
> 500 bp	141,863	239,349,699	90.948	88.648	113,205	79.799	124,766	87.948
> 1,000 bp	91,314	203,490,609	93.068	93.668	75,296	82.458	85,040	93.129
> 2,000 bp	42,452	132,559,496	94.850	96.928	34,882	82.168	40,925	96.403
> 5,000 bp	2,683	16,839,815	95.370	98.658	2,091	77.935	2,626	97.876

Supplemental Table 11. Summary of gene structure annotation in *G. anomalum* genome.

	Gene set	Number	Average gene length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
De novo	Augustus	58,306	2,113.65	969.23	4.04	239.88	376.39
	GlimmerHMM	110,304	9,227.34	583.06	2.79	208.71	4,819.37
	SNAP	38,739	8,617.61	548	3.75	146.15	2,934.88
	Geneid	85,816	4,341.16	694.95	3.82	181.78	1,291.55
	Genscan	58,848	12,145.66	1,010.16	5.15	196.30	2,685.93
Homolog	*Arabidopsis thaliana*	75,480	1,294.09	679.65	2.50	271.49	408.71
	Gossypium arboreum	39,986	2,687.86	1,190.45	3.96	300.86	506.43
	Gossypium hirsutum	66,844	2,412.87	997.83	3.32	300.67	610.27
	Gossypium raimondii	70,338	1,622.10	791.99	3.07	258.21	401.56
	Populus trichocarpa	39,204	2,619.07	1,169.22	3.72	314.34	533.12
	Theobroma cacao	34,369	2,808.17	1,310.04	4.01	326.70	497.73
RNA-seq	PASA	82,914	2,728.24	1,001.98	4.89	204.72	443.28
	Cufflinks	54,828	4,768.84	2,110.71	6.42	328.77	490.43
EVM		64,660	2,266.69	906.12	3.84	235.89	478.87
Pasa-update[*]		64,373	2,236.80	907.45	3.82	237.49	471.23
Final set[*]		42,752	3,140.07	1,099.98	4.75	291.92	464.73

[*] It contains UTR region, others do not contain UTR region.

Supplemental Table 12. Summary of gene function annotation by using different databases.

	Database	Number	Percent (%)
Annotated		41,592	97.29
	Swissprot	30,288	70.85
	Nr	41,561	97.21
	KEGG	30,706	71.82
	InterPro	33,900	79.29
	GO	21,680	50.71
	Pfam	31,663	74.06
Unannotated		1,160	2.71

Supplemental Table 13. Summary of non-coding RNAs in the *G. anomalum* genome.

Type	Copy number	Average length (bp)	Total length (bp)
miRNA	262	129.67	33,974
tRNA	1,085	74.73	81,078
rRNA	774	296.19	229,250
18S	110	1,277.05	140,476
28S	307	138.80	42,613
5.8S	82	160.12	13,130
5S	275	120.11	33,031
snRNA	6,064	107.49	651,847
CD-box	5,872	106.47	625,211
HACA-box	52	125.73	6,538
splicing	140	143.56	20,098
Total	8,185		996,149

Supplemental Table 14. The summary of repetitive sequences.

Type	Number	Length (bp)	% (genome)
Class I: Retrotransposon	840,690	728,632,679	60.30
LTR retrotransposon	805,038	714,153,605	59.11
Copia	136,198	66,734,309	5.52
Gypsy	598,287	607,203,108	50.25
Unclassied	70,553	40,216,188	3.33
non-LTR Retroelement	35,652	14,479,074	1.20
LINE	35,566	14,415,724	1.20
SINE	86	63,350	0.0052
Class II: DNA transposon	32,627	13,910,223	1.15
TIR/MULE-MuDR	14,505	6,577,793	0.54
TIR/CMC	6,354	2,606,893	0.22
TIR/P	432	290,271	0.02
TIR/PIF	2,593	1,034,534	0.09
TIR/TcMar	175	69,345	0.0057
TIR/hAT	6,511	2,412,317	0.20
Helitron/Helitron	1,623	775,713	0.06
Unclassfied	434	143,357	0.01
Other	2,172	1,620,171	0.13
Simple_repeat	2,172	1,620,171	0.13
unknown	20,680	12,118,599	1.00
Total	896,169	756,281,672	62.59

Supplemental Table 15. The summary of chromosomal rearrangements between *G. anomalum* and other extant diploid cotton genomes. (included as a separate EXCEL file)

Supplemental Table 16. Primers of 230 SSR markers. (included as a separate EXCEL file)

Supplemental Table 17. Population size for marker-assisted selection.

Chr.	BC_3F_1	BC_4F_1	BC_4F_2	BC_4F_3	BC_4F_4
Chr.01	888 (86)[*]	774 (50)[*]	752 (132)[*]	145	127
Chr.02	495 (77)	464 (119)	487 (69)	117	155
Chr.03	243 (32)	696 (503)	227 (108)	20	55
Chr.04	165 (41)	348 (123)	240 (94)	145	210
Chr.05	170 (64)	318 (40)	347 (48)	113	104
Chr.06	299 (228)	869 (761)	347 (112)	136	214
Chr.07	121 (91)	443 (443)	207 (170)	47	98
Chr.08	275 (275)	392 (392)	183 (112)	58	121
Chr.09	209 (114)	588 (410)	235 (108)	66	181
Chr.10	250 (201)	550 (493)	259 (188)	245	293
Chr.11	827 (63)	1,377 (57)	706 (46)	267	395
Chr.12	128 (65)	795 (589)	247 (69)	73	33
Chr.13	261 (240)	926 (926)	306 (143)	101	239
Total	4,331 (1,577)	8,540 (4,906)	4,543 (1,399)	1,533	2,225

[*]The data in parentheses is the number of plants from alien addition lines.

Supplemental Table 18. The number of recombination types identified in each generation.

Chr.	BC$_2$F$_1$	BC$_3$F$_1$	BC$_4$F$_1$	BC$_4$F$_2$	BC$_4$F$_3$	BC$_4$F$_4$
Chr.01	3	4	9	8	8	13
Chr.02	4	5	6	3 (2)	3 (2)	3
Chr.03	1	1	3	1 (1)	1 (1)	1 (1)
Chr.04	1	5	3	2 (1)	2 (1)	3
Chr.05	4	6	6	6	7	7
Chr.06	2	2	3	5	5	6 (1)
Chr.07	1	0	1	1	1	1
Chr.08	1	0	3	1	1	2 (1)
Chr.09	3	2	2	3 (1)	3 (1)	3 (1)
Chr.10	1	2	2	1	1	3
Chr.11	11	11	11	13 (1)	14 (1)	17
Chr.12	3	2	3	3	3	3
Chr.13	1	0	4	4	4	4
Total	36	40	56	51 (6)	53 (6)	74* (4)

The data in the parentheses denotes the number of hybrid recombination types in the generations produced by selfing.

*Since some recombination types containing substitution segments from different chromosomes, the total number of recombination types was greater than the total number accumulated in chromosome units.

Supplemental Table 19. Summary of introgression segments. (included as a separate EXCEL file)

Supplemental Table 20. Genome coverage of *G. anomalum* substitution segments in the CSSL population.

Chr.	B_1	B_1-introgression	Coverage (%)
Chr.01	176.40	176.40	100.00
Chr.02	191.50	156.14	81.54
Chr.03	118.70	71.10	59.90
Chr.04	101.60	101.60	100.00
Chr.05	263.90	116.90	44.30
Chr.06	167.80	146.10	87.07
Chr.07	195.40	119.60	61.21
Chr.08	183.50	183.50	100.00
Chr.09	172.60	157.50	91.25
Chr.10	165.50	29.60	17.89
Chr.11	216.20	216.20	100.00
Chr.12	159.50	53.25	33.39
Chr.13	198.10	140.80	71.08
Total	2,310.70	1,668.69	72.22

Supplemental Table 21. Summary of Illumina re-sequencing data for CSSLs. (included as a separate EXCEL file)

Supplemental Table 22. Descriptive statistics of nine traits in the *G. anomalum*-derived CSSL population in multiple environments. (included as a separate EXCEL file)

Supplemental Table 23. The detected segments/QTLs of the agronomic traits in the *G. anomalum*-derived CSSL population. (included as a separate EXCEL file)

Supplemental Table 24. Summary of RNA-Seq data for CSSL29 and Su8289 during PEG stress.

Sample	Raw reads	Clean reads	Clean base (Gb)	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
CSSL29_0h_1	61,426,990	60,361,996	9.05	0.03	97.82	93.65	44.75
CSSL29_0h_2	44,191,544	43,628,714	6.54	0.03	97.73	93.43	44.47
CSSL29_0h_3	50,666,142	49,874,820	7.48	0.03	97.78	93.54	44.40
CSSL29_6h_1	52,910,090	51,782,160	7.77	0.03	97.73	93.49	44.81
CSSL29_6h_2	58,005,932	56,894,338	8.53	0.03	97.83	93.69	44.89
CSSL29_6h_3	47,752,394	47,012,056	7.05	0.03	97.86	93.77	44.97
CSSL29_12h_1	53,507,968	52,325,294	7.85	0.03	97.88	93.81	44.51
CSSL29_12h_2	55,945,848	54,851,274	8.23	0.03	97.86	93.76	44.52
CSSL29_12h_3	50,944,974	50,096,754	7.51	0.03	97.74	93.48	44.55
CSSL29_24h_1	49,578,464	48,538,146	7.28	0.03	97.06	91.96	43.86
CSSL29_24h_2	44,966,346	44,162,348	6.62	0.03	97.76	93.58	43.75
CSSL29_24h_3	46,933,152	45,592,854	6.84	0.03	97.74	93.49	43.78
CSSL29_72h_1	51,561,758	50,548,304	7.58	0.03	97.69	93.37	44.36
CSSL29_72h_2	39,450,540	38,676,436	5.80	0.03	97.69	93.34	44.48
CSSL29_72h_3	47,329,794	46,003,272	6.90	0.03	97.75	93.53	44.34
Su8289_0h_1	50,433,980	48,946,092	7.34	0.03	97.40	92.68	43.77
Su8289_0h_2	57,284,664	55,813,900	8.37	0.03	97.68	93.38	43.70
Su8289_0h_3	50,878,936	49,550,684	7.43	0.03	97.81	93.64	43.84
Su8289_6h_1	54,160,104	52,570,184	7.89	0.03	97.80	93.64	44.83
Su8289_6h_2	48,345,034	47,398,424	7.11	0.03	97.77	93.50	45.03
Su8289_6h_3	47,582,286	46,571,368	6.99	0.03	97.93	93.94	44.27
Su8289_12h_1	45,899,858	44,919,500	6.74	0.03	97.69	93.41	44.26
Su8289_12h_2	54,921,350	53,900,484	8.09	0.03	97.69	93.41	44.33
Su8289_12h_3	42,819,618	41,180,122	6.18	0.03	97.78	93.60	44.35
Su8289_24h_1	49,719,686	48,971,020	7.35	0.03	97.69	93.39	44.52
Su8289_24h_2	71,473,464	70,755,984	10.61	0.03	97.78	93.59	44.66
Su8289_24h_3	43,695,654	43,176,316	6.48	0.03	97.48	92.79	44.61
Su8289_72h_1	53,836,832	52,959,380	7.94	0.03	97.76	93.47	44.13
Su8289_72h_2	53,247,538	52,145,682	7.82	0.03	97.61	93.15	44.15
Su8289_72h_3	53,473,164	52,345,388	7.85	0.03	97.68	93.35	44.08

Supplemental Table 25. The annotation and expression level of 20 genes in CSSL29 and Su8289 during PEG stress. (included as a separate EXCEL file).

Supplemental Table 26. The qRT-PCR primers for nine candidate genes on the JAAS0803–JAAS5604 interval.

Gene ID	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
Goano05G0268	AAGCTGCGAATATTGAACA	CAAGATGACCTCCGTAGATA
Goano05G0164	CTATGTACGCCGAGTTAATATG	CTTCTTCTTCTTCAGCATCC
Goano05G0170	AAAGTTCCCATTCATCATCAC	TCAAGCCTTAGACAAGTCAT
Goano05G0207	CACCAACAATAGCAACAATC	TTGAACTCATTAGCATCGTT
Goano05G0222	CTTTCCTTCCATTCTCCAAA	TCATCCATATCTTCATTCATAGAG
Goano05G0235	TGGATGGAGTGACCTCATACAA	ATCAAGAAGGAGATGTCCAAAGC
Goano05G0236	TTCTCAACTGTGCCTCCTAC	GTCCAGCATTATTAACACCATCTT
Goano05G0301	AGCAAGACAAGAAGAAGAAG	TCTCAGATGAGTTAGGTCC
Goano05G0319	TCTCAATCTCAAAGTAGAAAGAAG	CAACCATGAACAAGACCATAT

Supplemental Table 27. The VIGS primers for nine candidate genes on the JAAS0803–JAAS5604 interval.

Gene ID	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
Goano05G0268	CGGGATCCGGTCTCCATACACTCCCTCG	CGAGCTCATGACCTCCGTAGATAAAAG
Goano05G0164	CGGAATTCTTCGTGAAAGCCGAGGAATA	CGGGATCCGAGATTGCGTCATTGGTCAT
Goano05G0170	CGGGATCCAAGTTCCCATTCATCATCAC	CGAGCTCACATATGGACTTCTTAACTC
Goano05G0207	CGGAATTCAGCAACAATCAGACGAGCCA	CGGGATCCCTCAAGTTACTAGCATACGC
Goano05G0222	CGGGATCCGCACCTCCTCCTCACTTCTG	CGAGCTCACCTAACCCTAAACCCTAAT
Goano05G0235	CGGGATCCAGGCTCTTGTTCTTGTTTCT	CGAGCTCTCTTCATTGTTCATCCATAT
Goano05G0236	CGGGATCCCACAAGGATGGCAAACTGGC	CGAGCTCAGGATGGCATTGCTCTTTAT
Goano05G0301	CGGAATTCCTGCTCACCGCCATAAGCGAAAT	CGGGATCCCTTCTTGCCTTCTTCACCTACAT
Goano05G0319	CGGGATCCCAGGAGCAGAAACAACCATC	CGAGCTCGCTAAGGTGGTGGGGAGAAT

Supplemental Table 28. The PCR primers of *Goano05G0268* and its orthologous gene *GH_A05G0249* in Su8289.

Gene ID	Sample	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
Goano05G0268	CSSL29	CAACCGACTTACATCATCACCTTC	AACAAGATGACCTCCGTAGATAAA
GH_A05G0249	Su8289	CAACTGCCTTAGCAACCGACTT	TATTGCCATAACCATCAAGTGC

Supplemental Table 29. Structural variation statistics of *Goano05G0268* and its orthologous gene in other allotetraploid cotton species.

Genome	Assembly	Gene name	Deletion (bp)	Insertion-1 (bp)	Insertion-2 (bp)	Insertion-3 (bp)	SNP in exon
B_1	*G. anomalum*	*Goano05G0268*					
$(AD)_1$	*G. hirsutum* HAU v1.1	*Ghir_A05G002550*	408	2,544	159	2	10
$(AD)_2$	*G. barbadense* ZJU v1.1	*GB_A05G0251*	408	2,548	162	2	9
$(AD)_2$	*G. barbadense* HAU v2.0	*Gbar_A05G042520*	408	2,548	162	2	9
$(AD)_3$	*G. tomentosum* HGS_v1.1	*Gotom.A05G027400*	408	2,550	161	2	10
$(AD)_4$	*G. mustelinum* JGI_v1.1	*Gomus.A05G027200*	408	2,422	159	2	8
$(AD)_5$	*G. darwinii* HGS_v1.1	*Godar.A05G025900*	408	2,549	163	2	9

Supplemental Table 30. The detected segments/QTLs of the trait of FS in the *G. anomalum*-derived CSSL population.

Environment	Marker interval	Chr.	Start (bp)	End (bp)	LOD	PVE (%)	ADD
E1	JAAS4769–NAU2152	Chr.11	91,616,161	99,990,495	4.96	40.49	2.39
E2	NAU1305–NAU3678	Chr.07	36,412,351	80,217,863	2.91	16.45	-2.79
E2	JAAS4769–NAU2152	Chr.11	91,616,161	99,990,495	3.66	21.39	1.48
E3	JAAS4769–NAU2152	Chr.11	91,616,161	99,990,495	5.91	43.97	1.96
E4	JAAS4769–NAU2152	Chr.11	91,616,161	99,990,495	3.24	20.51	1.72
Mean	JAAS2076–JAAS2022	Chr.04	27,206,433	79,551,098	2.71	12.72	2.28
Mean	JAAS4769–NAU2152	Chr.11	91,616,161	99,990,495	4.38	21.77	1.78

Supplemental Table 31. Primers of 19 SSR markers for fine mapping of the *qFS11*.

Marker name	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
NAU3703	AACTCCCATCATCAGCAAGT	TATCAATCGGTTCAGCAATG
JAAS8011	TGTGTGTGTGTTATTGGCTGC	GACCCAGTTGTGTCTAATGTTTTG
JAAS8027	TGCCCCACCATTCTATTGAT	GAAGCTGCCTTTCACTCACC
JAAS8033	CAATATGGATAAGGTCAATTCTTTTT	GGGGAAAATGGAGTTTGAGG
JAAS8034	TGACAAGAAATTCGGCAGTG	TGTCCAAAACACATGCTCGT
JAAS8036	TAGTTTGTTGCCTCACGTCG	CGAATCGAGTTTTGCTCACC
JAAS8037	TGGTGTAGTTTAACCGTGCAA	TGATCGACTGATTCGACCTG
JAAS8040	CACATCATCAACCCATACTATGCT	TGAACTAAACACCTATTCACCTGC
JAAS8084	TGGCATAATAATGTCAAACCTTTT	TAAATCATTGCCTGGGCTTC
JAAS8099	TTCGGATAGGTCCGAAAGTG	ATGTTGCTTCACATCCAAGC
JAAS8103	CCTGCTGGTTTTTATGGCA	TCATTTGAAAAAGGGAGCCA
JAAS8104	TTGTGGTGCGGTTTGTATGT	TGACAATTTTAAGTAAAGAGAGCAAG
JAAS8116	TCTCTTTCTTTCGCCTTGGA	GTGCCCAGCAATATCGTTTT
JAAS8125	TGACTTTGCCTTATTTCCGC	ACACGTGACTGTGTGCTCGT
JAAS8129	TGCATGTCACAATCCTTGAC	TTTCTTGGTTAAGGCGAAGG
JAAS8143	CTGTACACGACTCTCCGCAA	GGGATAGCATCAGTTAAGCAGC
JAAS8144	AAATGGACCAAGTTGCTGCT	TGCCCAACTTGAGCTTGAAT
JAAS8157	GTGGGACCTTTGCTTTTGAA	TTTGGGGTCAAGTTTTGATTC
JAAS8164	GCCCATGCTCACCTCTAGTC	TGGAAGGGAAACAAACATGA

Supplemental Table 32. Summary of RNA-Seq data for CSSL13, CSSL59, and Su8289.

Sample	Raw reads	Clean reads	Clean base (Gb)	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
CSSL13_20d_1	22,418,501	21,680,087	6.50	0.03	97.95	93.95	44.87
CSSL13_20d_2	26,870,022	26,001,940	7.80	0.03	98.00	94.09	44.68
CSSL13_20d_3	20,963,564	20,227,253	6.07	0.02	98.05	94.19	44.73
CSSL59_20d_1	25,673,966	24,568,436	7.37	0.03	97.95	93.73	44.54
CSSL59_20d_2	27,648,078	26,755,427	8.03	0.03	97.88	93.59	44.64
CSSL59_20d_3	25,127,994	23,966,272	7.19	0.03	98.05	93.95	44.74
Su8289_20d_1	26,678,116	25,693,401	7.71	0.03	97.91	93.65	44.75
Su8289_20d_2	22,309,720	21,487,496	6.45	0.03	97.92	93.69	45.02
Su8289_20d_3	31,491,960	30,147,965	9.04	0.03	98.00	93.81	46.11

Supplemental Table 33. The annotation and expression level of 12 genes at 20 DPA fibers in CSSL13, CSSL59, and Su8289. (included as a separate EXCEL file).

Supplemental Table 34. The sequence variants of nine genes between between *G. hirsutum* A_{t1}-subgenome and *G. anomalum*. (included as a separate EXCEL file).

Supplemental Table 35. The qRT-PCR primers for nine candidate genes on the *qFS11* interval JAAS8037–JAAS8040.

Gene ID	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
Goano11G3880	GATAAATTGGAGAAAGGAGGAA	TAAGAAGCAAACAAAGTAGGAA
Goano11G3883	AGATGATGGCAAGATAGATGA	TTCCTCTAACTGTTCTGCTTA
Goano11G3885	CTAAGTTTCCTTCTTCTTCAAATG	ACTAGATATACGAGATAAGATGCT
Goano11G3886	AGATACGTTCTCAATCTTCAAT	TAAACATAACCTGTCTCCTCTA
Goano11G3887	GAAGATGAAAGGGAAATGTGTA	CAAGCGGATTCAAAGAAATTC
Goano11G3888	AAGATATGAGGTATGGCAAGAA	GTGTTTGTAGTACGAACTCTATG
Goano11G3897	CGGAATAGGGTTAAGTAAATCTTT	ATTGGAACAGCATAGCTCTAAA
Goano11G3907	ACCAGATGTCTCAGATAGAA	CTGTAGTTGGAGTTCGTTAT
Goano11G3918	ATAGCAAATGAATTGATCCATTC	CCATCTTGATTAACACTAACTCTA

Supplemental Table 36. The PCR primers of *Goano11G3883* and *GH_A11G3315* in CSSL59 and Su8289, respectively.

Gene ID	Sample	Forward PCR primer (5'–3')	Reverse PCR primer (5'–3')
Goano11G3883	CSSL59	AGAGGGTTGTATCAAACATCGTG	CTCTAACTGTTCTGCTTAGCGTTT
GH_A11G3315	Su8289	TGATGGCAAACGGTCACAGAGTA	CTCTAACTGTTCTGCTTAGCGTTT

Supplemental Table 37. The 130 *G. anomalum*-specific SSR markers for detecting recombinant individuals.

Chr.	SSR marker
Chr.01	JAAS1098, JAAS0006, NAU3690, NAU3615, JAAS1148, NAU2182, NAU5100, NAU3714, NAU2083, JAAS0392, JAAS3221
Chr.02	NAU2929, JAAS1489, NAU5421, NAU1190, JAAS0426, NBRI0092, JAAS2050, NAU3820, HAU1219
Chr.03	NAU2908, NAU1072, JAAS2579, JAAS4015, NAU895, NAU1167, JAAS4003, JAAS0861, JAAS4512, NAU3083
Chr.04	JAAS5943, JAAS5256, NAU2120, JAAS2076, NAU3508, JAAS2022, JAAS2977, DPL494, JAAS2662
Chr.05	JAAS0803, JAAS4908, JAAS1953, JAAS4815, DC40130, NAU3650, NAU5273, HAU1248, NAU911, JAAS3432, JAAS2949, JAAS0657
Chr.06	NAU2714, NAU4969, DPL101, JAAS6138, JAAS2590, NAU905, DPL365, NAU2397, JAAS6227, NBRI3941
Chr.07	JAAS4754, JAAS2966, NAU2686, NAU4956, NBRI0047, NAU1305, JAAS5041, JAAS0945
Chr.08	JAAS0199, JAAS1049, JAAS3476, JAAS4458, JAAS4933, JAAS4907, JAAS5813, JAAS5861, NAU1037, NAU1262, NAU1322, NAU2169, JAAS6420, NAU2914
Chr.09	JAAS3113, NAU3100, JAAS2010, JAAS5567, JAAS5248, JAAS4094, JAAS2567, JAAS0163, JAAS1469, JAAS0125, JAAS1923
Chr.10	JAAS1256, JAAS3070, NAU4881, NAU1236, JAAS2768, CGR6110, JAAS0321
Chr.11	NAU5192, JAAS4259, JAAS2330, JAAS3310, JAAS3820, JAAS3088, NAU1063, JAAS1048, NAU3703, JAAS0852, NAU2152, JAAS4829
Chr.12	NAU3109, NAU1274, NAU1119, JAAS2266, NAU3294, JAAS6372, JAAS4943, JAAS3735, JAAS6269
Chr.13	JAAS2956, JAAS3946, JAAS6264, JAAS5724, NAU2697, NAU3398, NBRI0431, NBRI0978

REFERENCES

Chen, Z.J., Sreedasyam, A., Ando, A., Song, Q.X., De Santiago, L.M., Hulse-Kemp, A.M., Ding, M.Q., Ye, W.X., Kirkbride, R.C., Jenkins, J., et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat. Genet. **52**:525-533. 10.1038/s41588-020-0614-5.

Deng, X., Gong, J., Liu, A., Shi, Y., Gong, W., Ge, Q., Li, J., Shang, H., Wu, Y., and Yuan, Y. (2019). QTL mapping for fiber quality and yield-related traits across multiple generations in segregating population of CCRI 70. J. Cotton Res. **2**:13. 10.1186/s42397-019-0029-y.

Diouf, L., Magwanga, R.O., Gong, W., He, S., Pan, Z., Jia, Y.H., Kirungu, J.N., and Du, X. (2018). QTL mapping of fiber quality and yield-related traits in an intra-specific upland cotton using genotype by sequencing (GBS). Int. J. Mol. Sci. **19**:441. 10.3390/ijms19020441.

Guo, Y., Guo, X., Wang, F., Wei, Z., Zhang, S., Wang, L., Yuan, Y., Zeng, W., Zhang, G., Zhang, T., et al. (2014). Molecular tagging and marker-assisted selection of fiber quality traits using chromosome segment introgression lines (CSILs) in cotton. Euphytica. **200**:239-250. 10.1007/s10681-014-1150-0.

Huang, C., Shen, C., Wen, T., Gao, B., Zhu, D., Li, X., Ahmed, M.M., Li, D., and Lin, Z. (2018). SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population. Mol. Genet. Genomics. **293**:793-805. 10.1007/s00438-018-1419-4.

Jamshed, M., Jia, F., Gong, J., Palanga, K.K., Shi, Y., Li, J., Shang, H., Liu, A., Chen, T., Zhang, Z., et al. (2016). Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in *Gossypium hirsutum* recombinant inbred line population. BMC Genomics. **17**:197. 10.1186/s12864-016-2560-2.

Jia, X., Wang, H., Pang, C., Ma, Q., Su, J., Wei, H., Song, M., Fan, S., and Yu, S. (2018). QTL delineation for five fiber quality traits based on an intra-specific *Gossypium hirsutum* L. recombinant inbred line population. Mol. Genet. Genomics. **293**:831-843. 10.1007/s00438-018-1424-7.

Lacape, J., Nguyen, T., Courtois, B., Belot, J., Giband, M., Gourlot, J., Gawryziak, G., Roques, S., and Hau, B. (2005). QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations. Crop Sci. **45**:123-140. 10.2135/cropsci2005.0123a.

Li, C., Yu, H., Li, C., Zhao, T., Dong, Y., Deng, X., Hu, J., Zhang, Y., Zhang, F., Daud, M.K., et al. (2018). QTL mapping and heterosis analysis for fiber quality traits across multiple genetic populations and environments in upland cotton. Front. Plant Sci. **9**:1364. 10.3389/fpls.2018.01364.

Liu, R., Gong, J., Xiao, X., Zhang, Z., Li, J., Liu, A., Lu, Q., Shang, H., Shi, Y., Ge, Q., et al. (2018). GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. Front. Plant Sci. **9**:1067. 10.3389/fpls.2018.01067.

Ma, L., Zhao, Y., Wang, Y., Shang, L., and Hua, J. (2017). QTLs analysis and validation for fiber quality traits using maternal backcross population in Upland cotton. Front. Plant Sci. **8**:2168. 10.3389/fpls.2017.02168.

Rong, J., Feltus, F.A., Waghmare, V.N., Pierce, G.J., Chee, P.W., Draye, X., Saranga, Y., Wright, R.J., Wilkins, T.A., May, O.L., et al. (2007). Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics. **176**:2577-2588. 10.1534/genetics.107.074518.

Saranga, Y., Jiang, C.X., Wright, R.J., Yakir, D., and Paterson, A.H. (2004). Genetic dissection of cotton physiological responses to arid conditions and their inter-relationships with productivity. Plant, Cell Environ. **27**:263-277. 10.1111/j.1365-3040.2003.01134.x.

Shang, L., Wang, Y., Wang, X., Liu, F., Abduweli, A., Cai, S., Li, Y., Ma, L., Wang, K., and Hua, J. (2016). Genetic analysis and stable QTL detection on fiber quality traits using two recombinant inbred line populations and their backcross progeny in upland cotton. G3 (Bethesda). **6**:2717-2724. 10.1534/g3.116.031302.

Shappley, Z.W., Jenkins, J.N., Zhu, J., and McCarty, J.C. (1998). Quantitative trait loci associated with agronomic and fiber traits of upland cotton. J. Cotton Sci. **2**:153-163.

Shen, X., Guo, W., Zhu, X., Yuan, Y., Yu, J., Kohel, R., and Zhang, T. (2005). Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol. Breeding. **15**:169-181. 10.1007/s11032-004-4731-0.

Tang, S., Teng, Z., Zhai, T., Fang, X., Liu, F., Liu, D., Zhang, J., Liu, D., Wang, S., Zhang, K., et al. (2015). Construction of genetic map and QTL analysis of fiber quality traits for Upland cotton (*Gossypium hirsutum* L.). Euphytica. **201**:195-213. 10.1007/s10681-014-1189-y.

Tan, Z., Zhang, Z., Sun, X., Li, Q., Sun, Y., Yang, P., Wang, W., Liu, X., Chen, C., Liu, D., et al. (2018). Genetic map construction and fiber quality QTL mapping using the CottonSNP80K array in upland cotton. Front. Plant Sci. **9**:225. 10.3389/fpls.2018.00225.

Wang, B., Guo, W., Zhu, X., Wu, Y., Huang, N., and Zhang, T. (2007). QTL mapping of yield and yield components for elite hybrid derived-RILs in upland cotton. J. Genet. Genomics. **34**:35-45. 10.1016/S1673-8527(07)60005-8.

Wang, H., Huang, C., Guo, H., Li, X., Zhao, W., Dai, B., Yan, Z., and Lin, Z. (2015). QTL mapping for fiber and yield traits in upland cotton under multiple environments. PLoS One. **10**:e0130742. 10.1371/journal.pone.0130742.

Wang, H., Jin, X., Zhang, B., Shen, C., and Lin, Z. (2015). Enrichment of an intraspecific genetic map of upland cotton by developing markers using parental RAD sequencing. DNA Res. **22**:147-160. 10.1093/dnares/dsu047.

Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., L, J., Liu, F., Pei, L., Wang, P., Zhao., G., et al. (2019). Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. Nat. Genet. **51**:224-229. 10.1038/s41588-018-0282-x.

Yu, J., Zhang, K., Li, S., Yu, Y., Zhai, H., Wu, M., Li, X., Fan, S., Song, M., Yang, D., et al. (2013). Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. Theor. Appl. Genet. **126**:275-287. 10.1007/s00122-012-1980-x.

Yu, J.Z., Ulloa, M., Hoffman, S.M., Kohel, R.J., Pepper, A.E., Fang, D.D., Percy, R.G., and Burke, J.J. (2014). Mapping genomic loci for cotton plant architecture, yield components, and fiber properties in an interspecific (*Gossypium hirsutum* L. × *G. barbadense* L.) RIL population. Mol. Genet. Genomics. **289**:1347-1367. 10.1007/s00438-014-0930-5.

Zhang, K., Zhang, J., Ma, J., Tang, S., Liu, D., Teng, Z., Liu, D., and Zhang, Z. (2012). Genetic mapping and quantitative trait locus analysis of fiber quality traits using a three-parent composite population in upland cotton (*Gossypium hirsutum* L.). Mol. Breeding. **29**:335-348. 10.1007/s11032-011-9549-y.

Zhang, S., Feng, L., Xing, L., Yang, B., Gao, X., Zhu, X., Zhang, T., Zhou, B. and Jenkins, J. (2016). New QTLs for lint percentage and boll weight mined in introgression lines from two feral landraces into *Gossypium hirsutum* acc TM-1. Plant Breed. **135**:90-101. 10.1111/pbr.12337.

Zhang, Z., Ge, Q., Liu, A., Li, J., Gong, J., Shang, H., Shi, Y., Chen, T., Wang, Y., Palanga, K.K, et al. (2017). Construction of a high-density genetic map and its application to QTL identification for fiber strength in upland cotton. Crop Sci. **57**:774-788. 10.2135/cropsci2016.06.0544.