

Supplementary Information for

Evolutionary divergence of duplicated genomes in newly described allotetraploid

Renhai Peng^{a,1}, Yanchao Xu^{b,1}, Shilin Tian^{c,1}, Turgay Unver^{d,1}, Zhen Liu^{a,1}, Zhongli Zhou^{b,1}, Xiaoyan Cai^{b,1}, Kunbo Wang^b, Yangyang Wei^a, Yuling Liu^a, Heng Wang^b, Guanjing Hu^{b,e}, Zhongren Zhang^c, Corrinne E. Grover^f, Yuqing Hou^b, Yuhong Wang^b, Pengtao Li^a, Tao Wang^a, Quanwei Lu^a, Yuanyuan Wang^h, Justin L. Conover^f, Hassan Ghosal^g, Qinglian Wang^h, Baohong Zhang^{i,*}, Marc Van Montagu^{j,k,*}, Yves Van de Peer^{j,k,l,m,*}, Jonathan F. Wendel^{f,*}, and Fang Liu^{b,*}

Email: liufcri@163.com (F.L.), zhangb@ecu.ed (B.Z.), marc.vanmontagu@ugent.be (M.V.M.), yves.vandeppeer@psb.vib-ugent.be (Y.V.d.P.), or jfw@iastate.edu (J.F.W.)

This PDF file includes:

Figures S1 to S18
Tables S1 to S19
Legends to Datasets S1 and S2

Other supplementary materials for this manuscript include the following:

Datasets S1 to S2

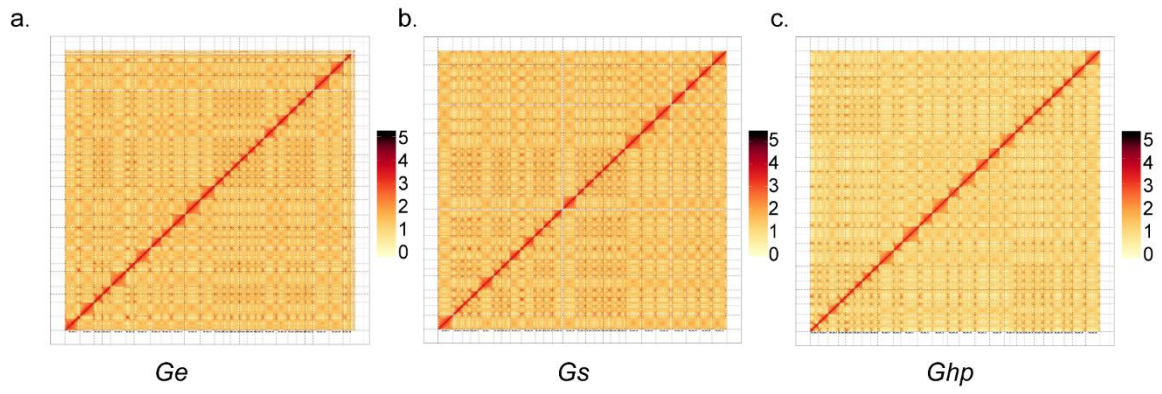


Fig. S1. Heatmap of chromatin interactions in each chromosome of *Ge* (a), *Gs* (b), and *Ghp* (c).

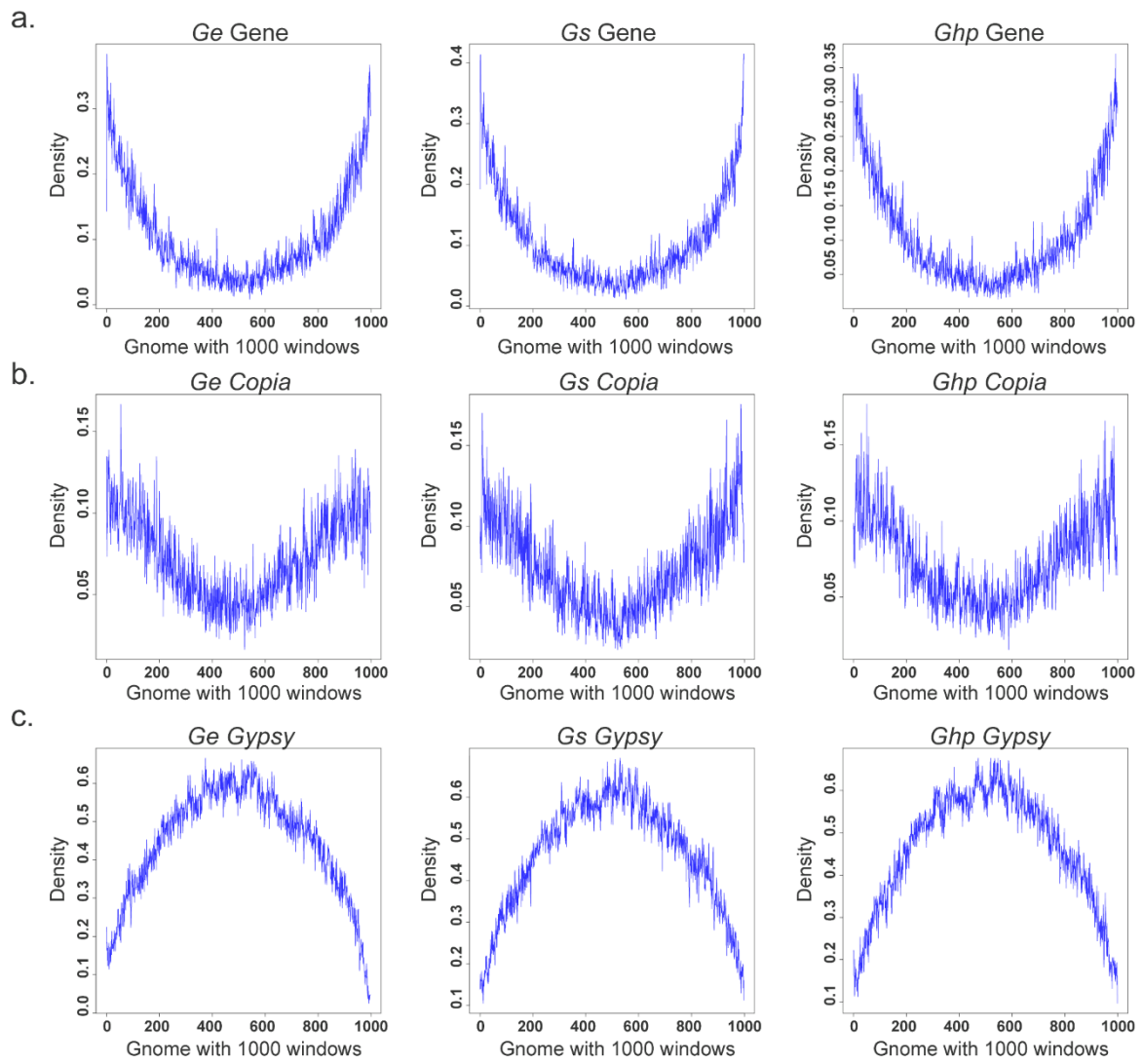


Fig. S2. Density distribution of genes (a), *Copia* (b) and *Gypsy* (c) in three cotton genomes.

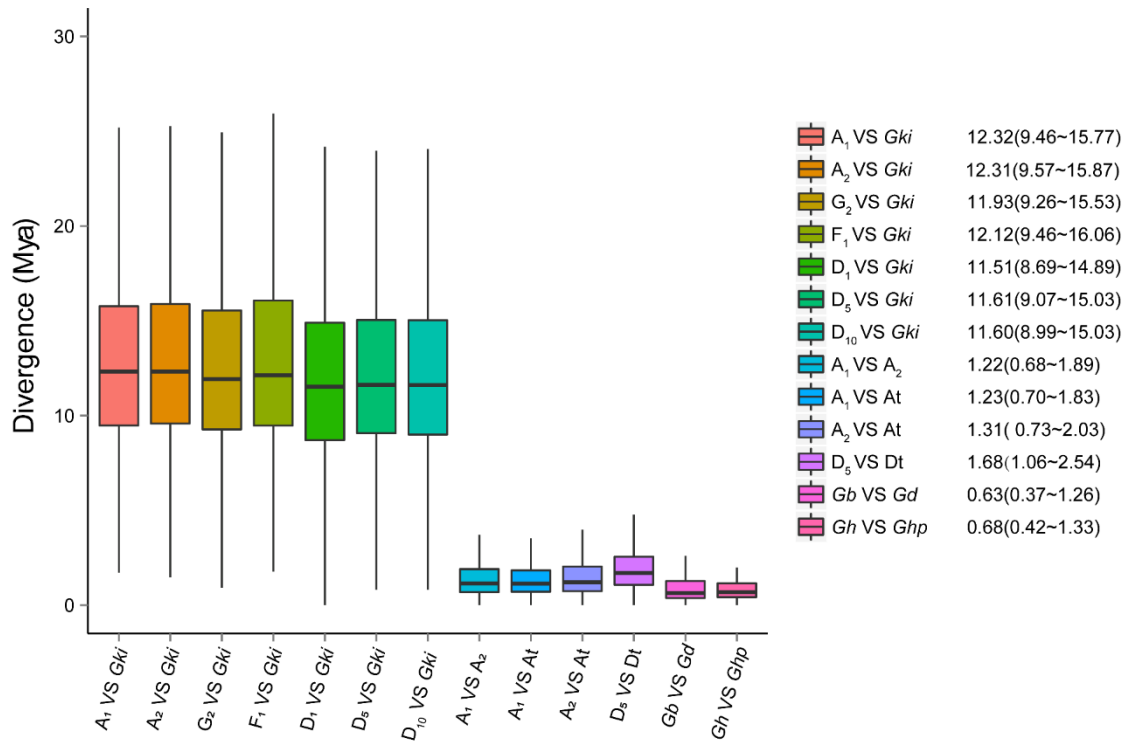


Fig. S3. Divergence between *Gossypium* genomes. Divergence times were estimated by *Ks* values for orthologous genes among *Gossypium* genomes. Ranges for divergence times are shown in the legend.

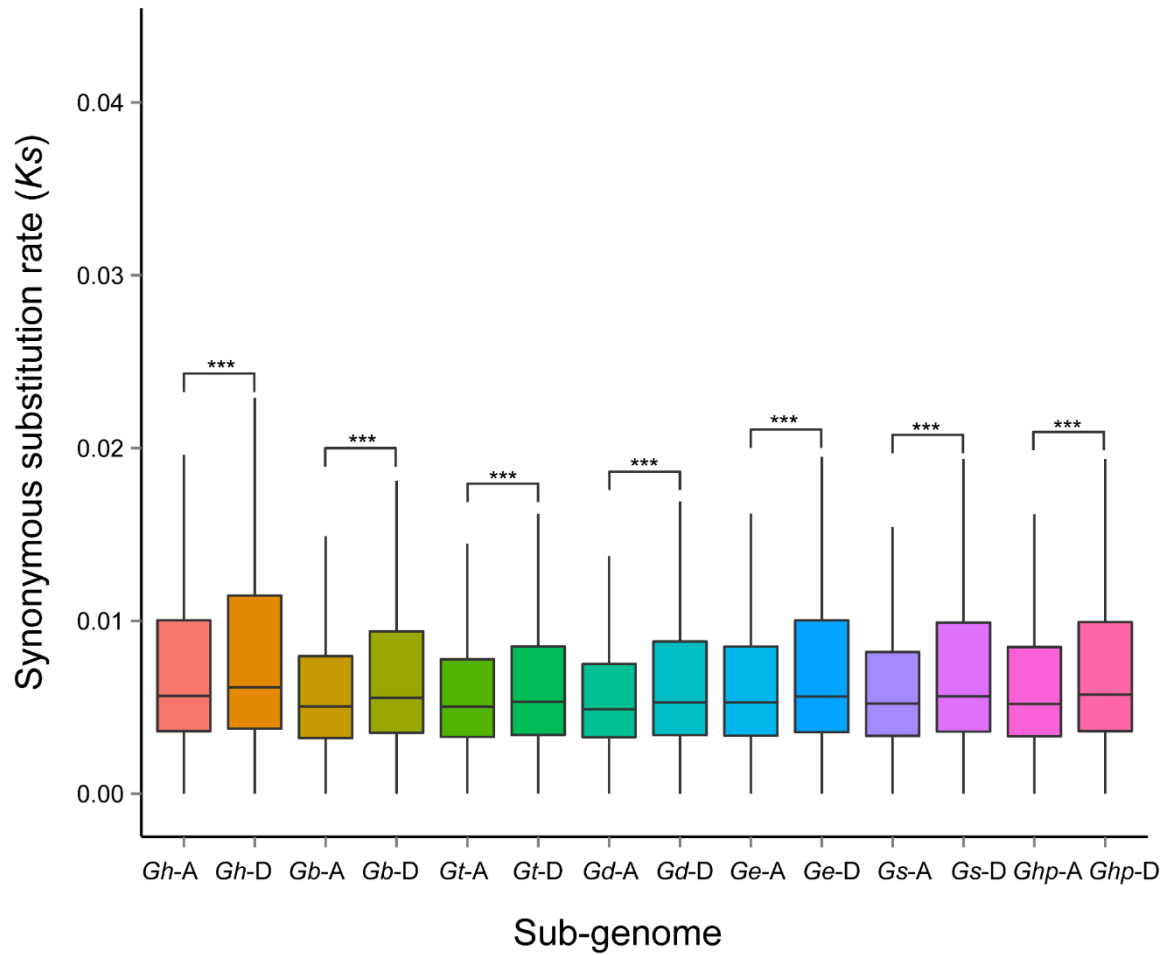


Fig. S4. Divergence between Gm with other tetraploid cotton genomes in A/D subgenome. Divergence times were estimated by K_s values for orthologous genes. Differences between the A- and D- subgenomes were analyzed by Wilcoxon signed-rank test. *** indicates P value < 0.001 .

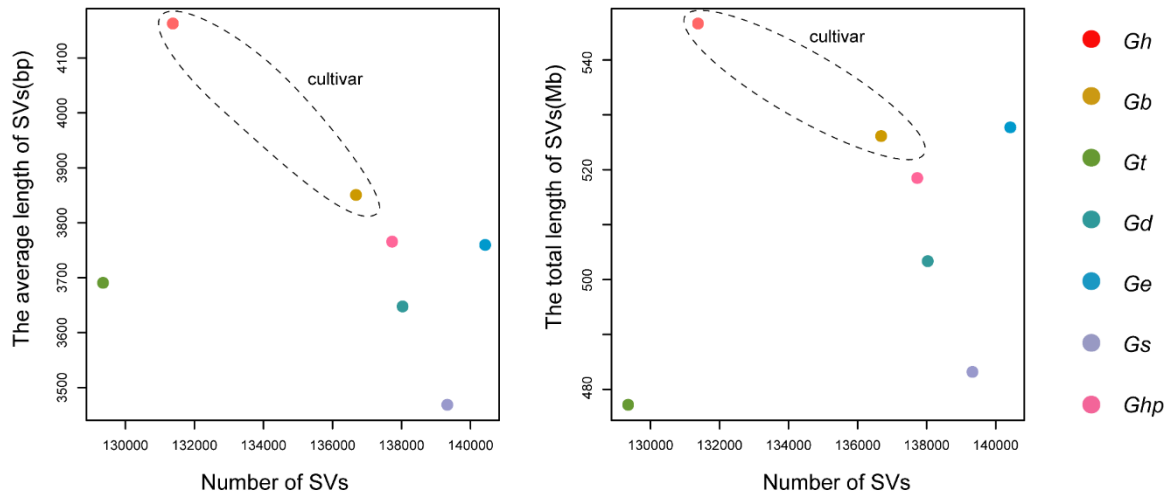


Fig. S5. PAV number and length distribution of seven allotetraploid cotton genomes relative to *Gm*.

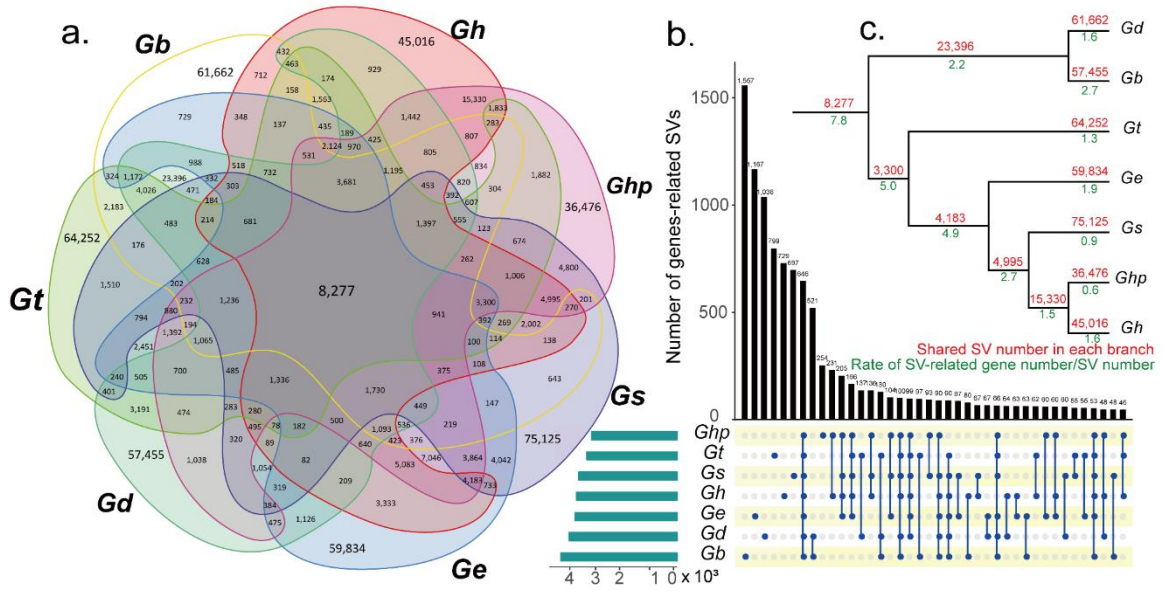


Fig. S6. PAVs and PAV-related genes in seven cotton genomes relative to *Gm*. (a) Venn diagram shown PAVs in seven genomes. (b) PAV-related genes in seven genomes. (c) PAVs evolution in seven genomes.

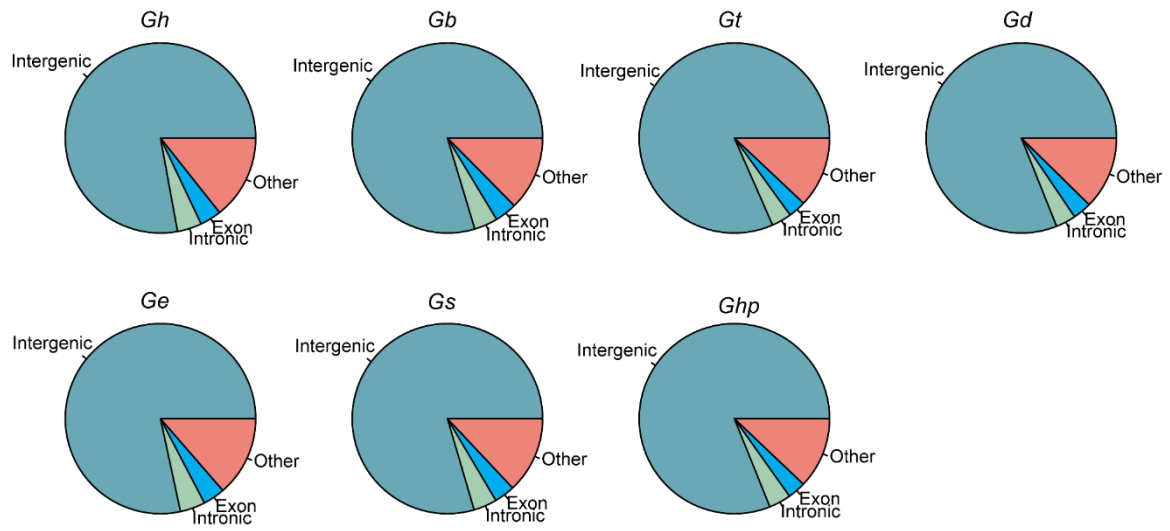


Fig. S7. PAV annotation in allotetraploid cotton, relative to *Gm*.

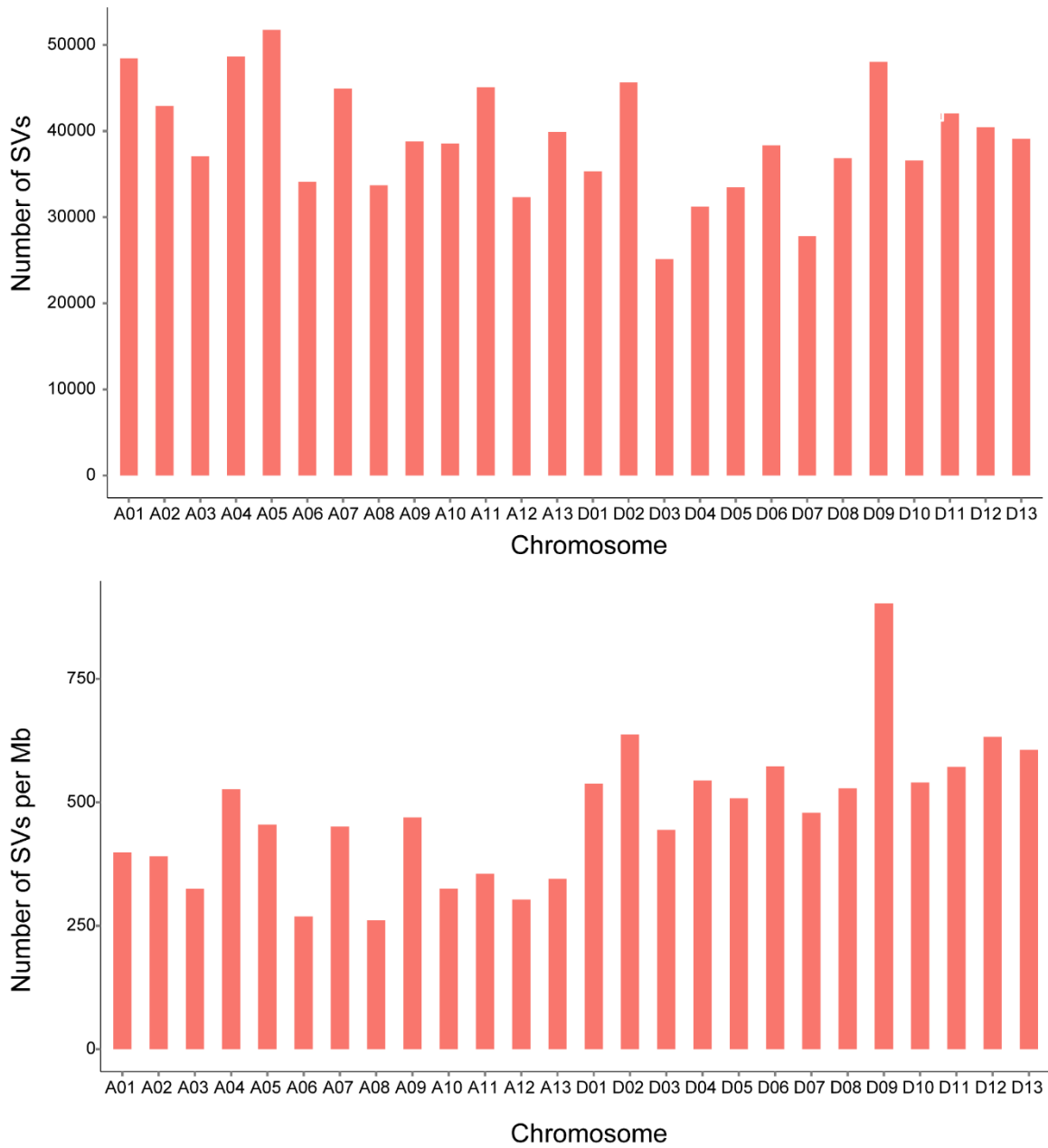


Fig. S8. SV density and number in 26 chromosomes.

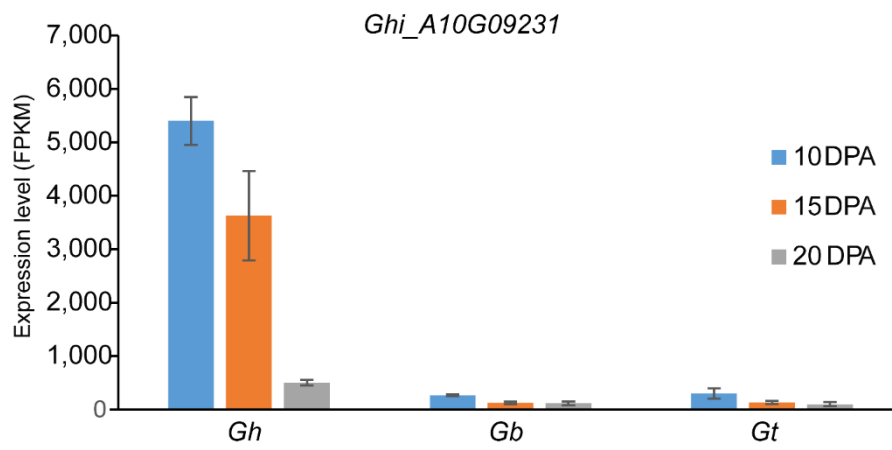
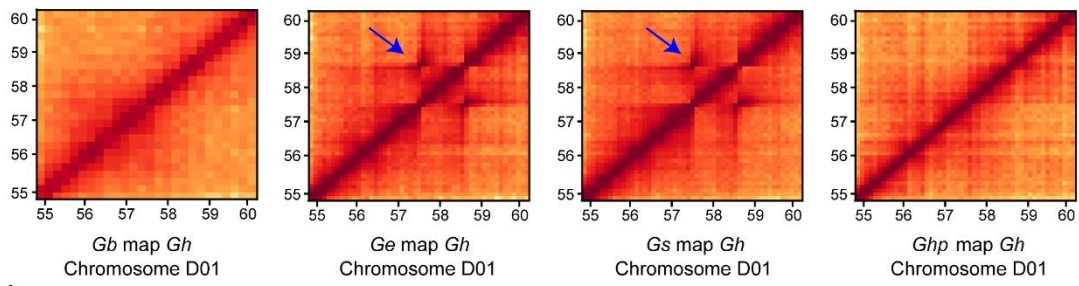


Fig. S9. Gene expression of *Ghi_A10G09231* (DPA, day post anthesis).

a.



b.

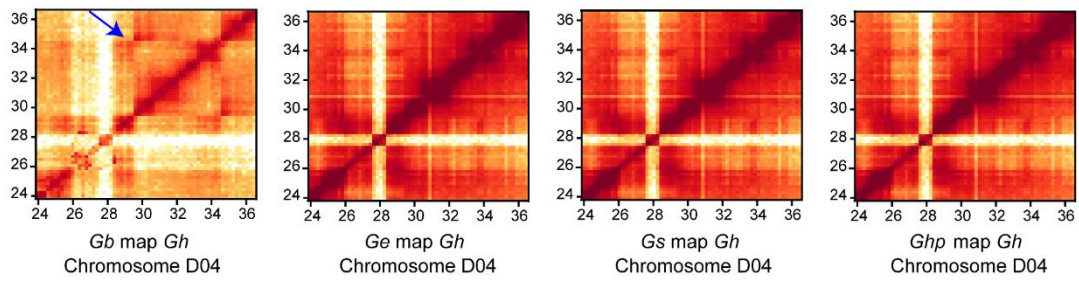


Fig. S10. SV identified by Hi-C data. *Gb*, *Ge*, *Gs* and *Ghp* Hi-C data mapped to TM-1 D01 (a) and D04 (b). The (a) and (b) dotted boxes indicate the inversions between D01 and D04, respectively.

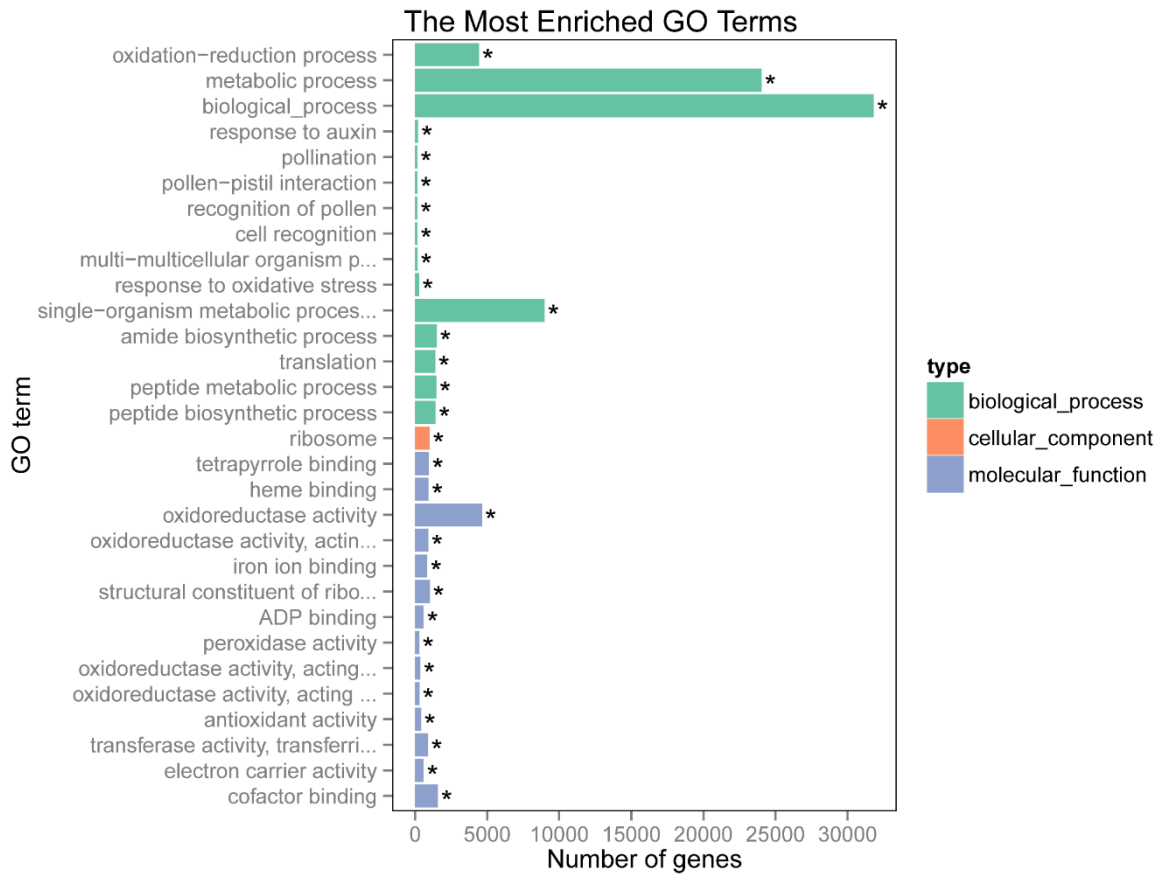


Fig. S11. GO enrichment analysis of genes in core gene families. * $P < 0.05$ (Fisher exact test).

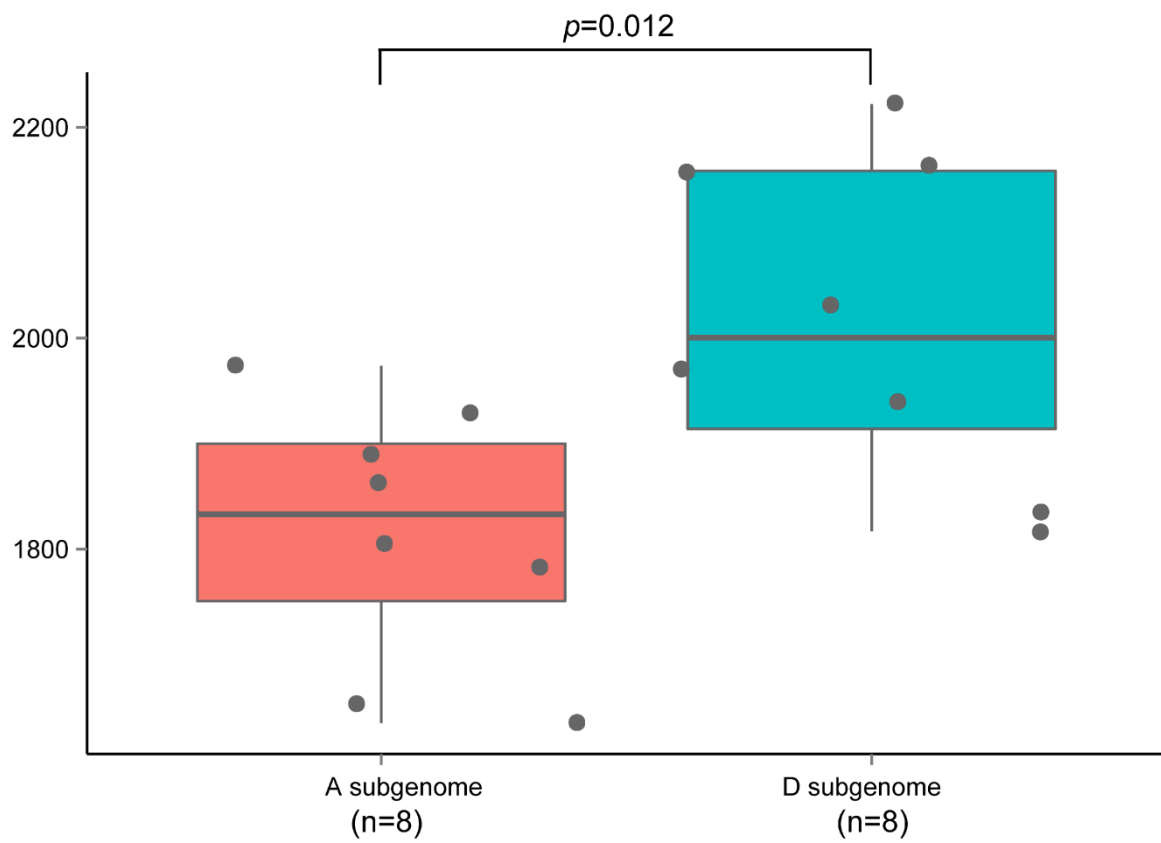


Fig. S12. NLR gene content of A subgenome and D subgenome in eight species/accessions. *P* value, Student *t* test.

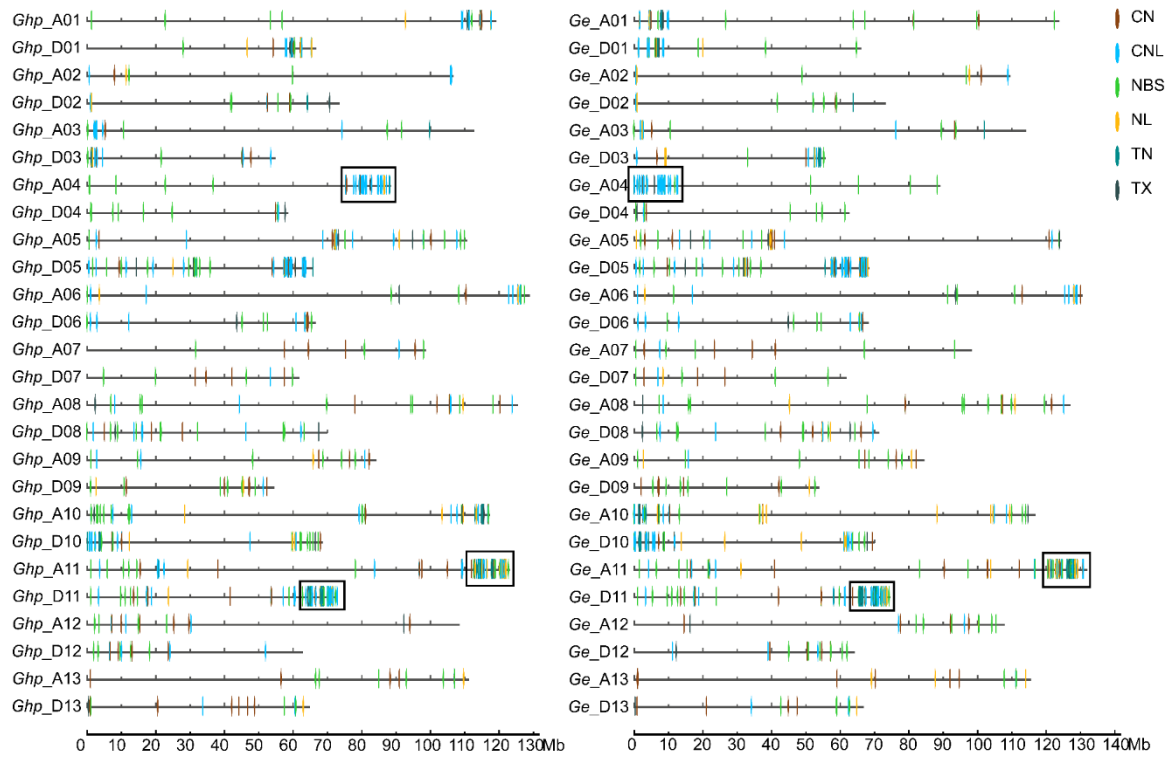


Fig. S13. NLR distribution in *Ghp* and *Ge* genome.

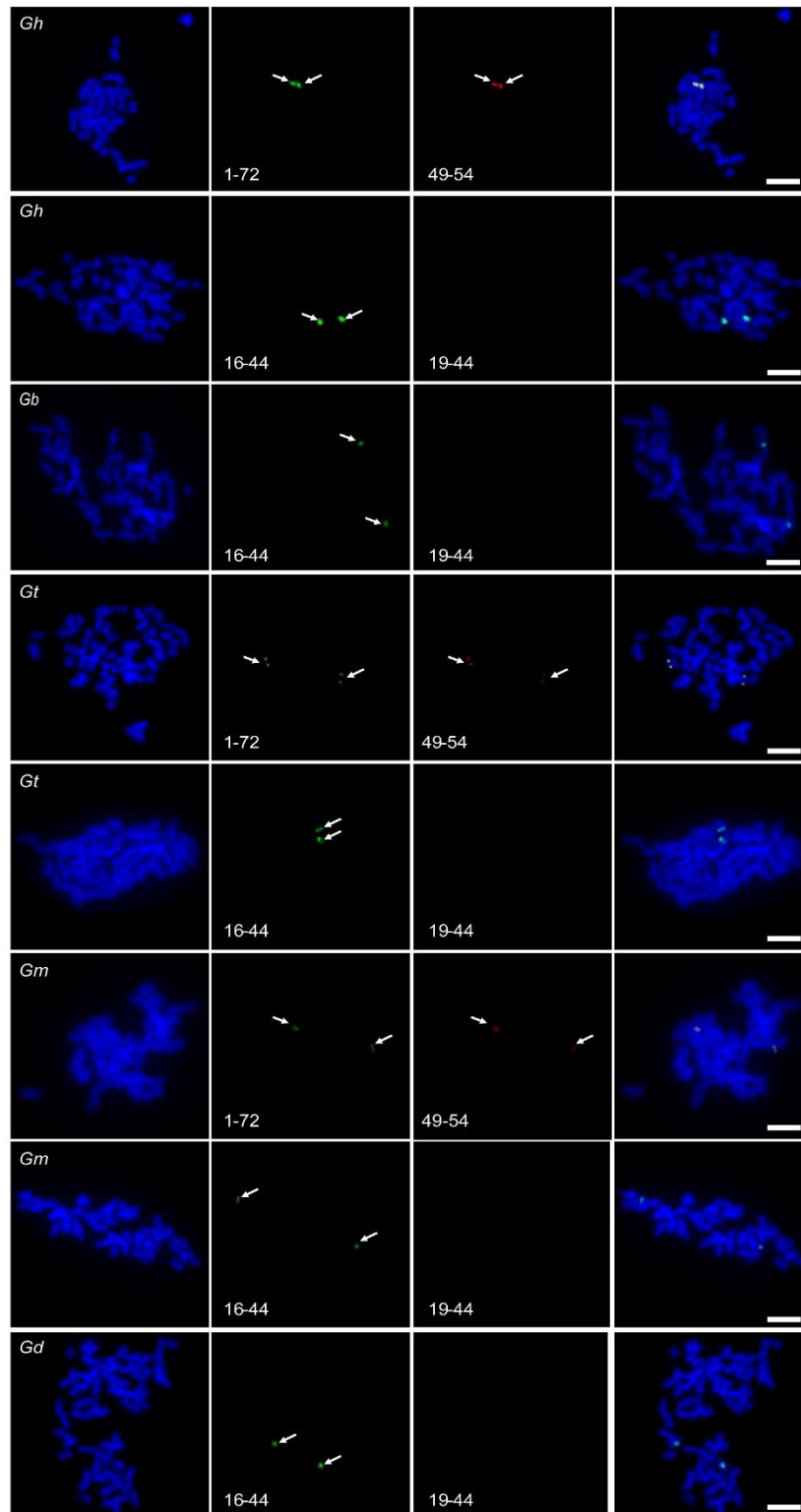


Fig. S14. Cytological observation using probes specific for NLR clusters confirms the presence of NLR clusters in allotetraploid cotton by Oligo FISH. Probes 1-72 was designed according to R gene cluster in A04 of *Ge*. Probes 16-44, 19-44 and 49-54 were designed according to R gene cluster in A11 of *Ge*. White bars, 5 μ m.



Fig. S15. GO enrichment analysis of *Ghp*-specific genes.

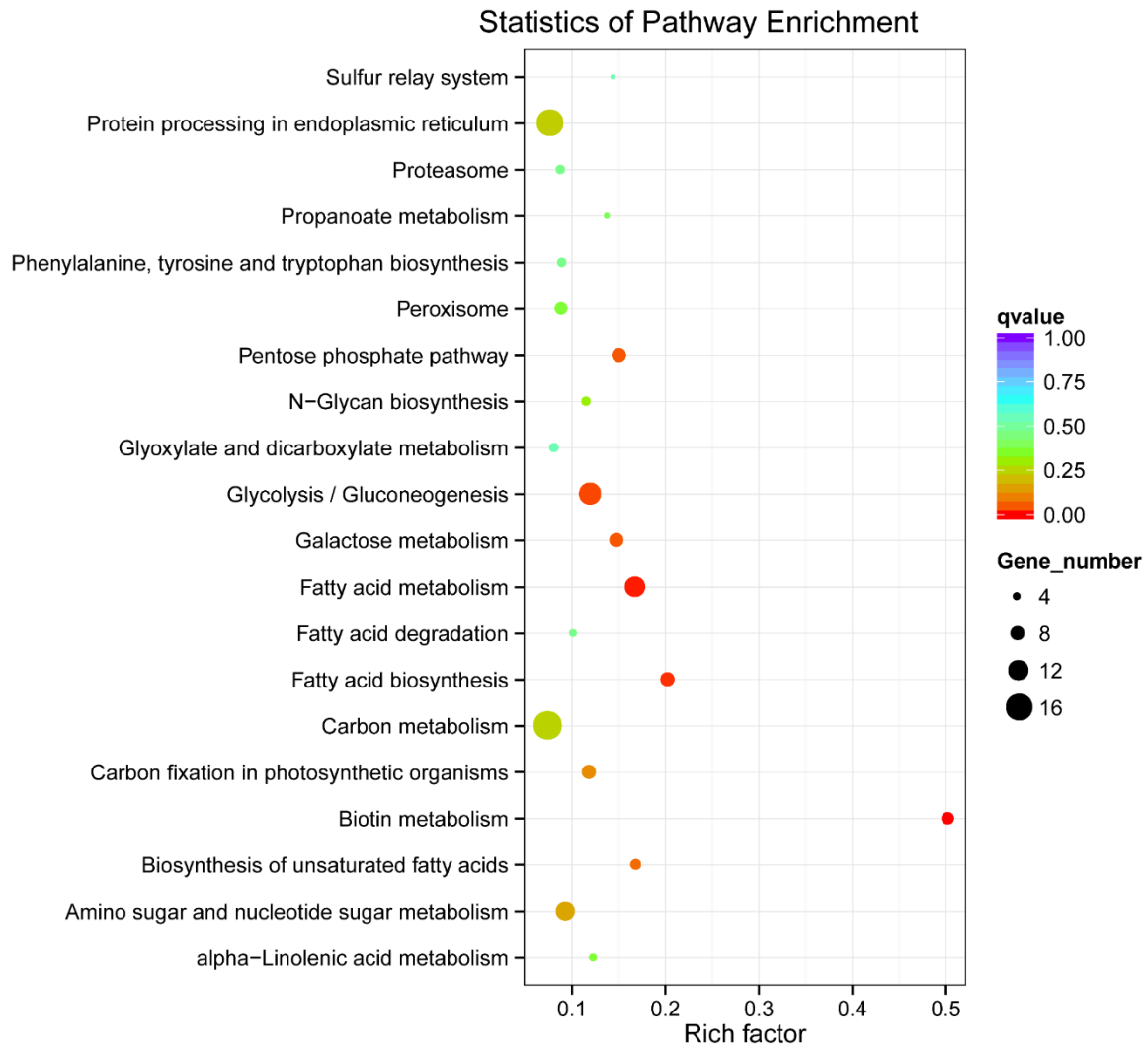


Fig. S16. KEGG enrichment analysis of *Ghp*-specific genes.

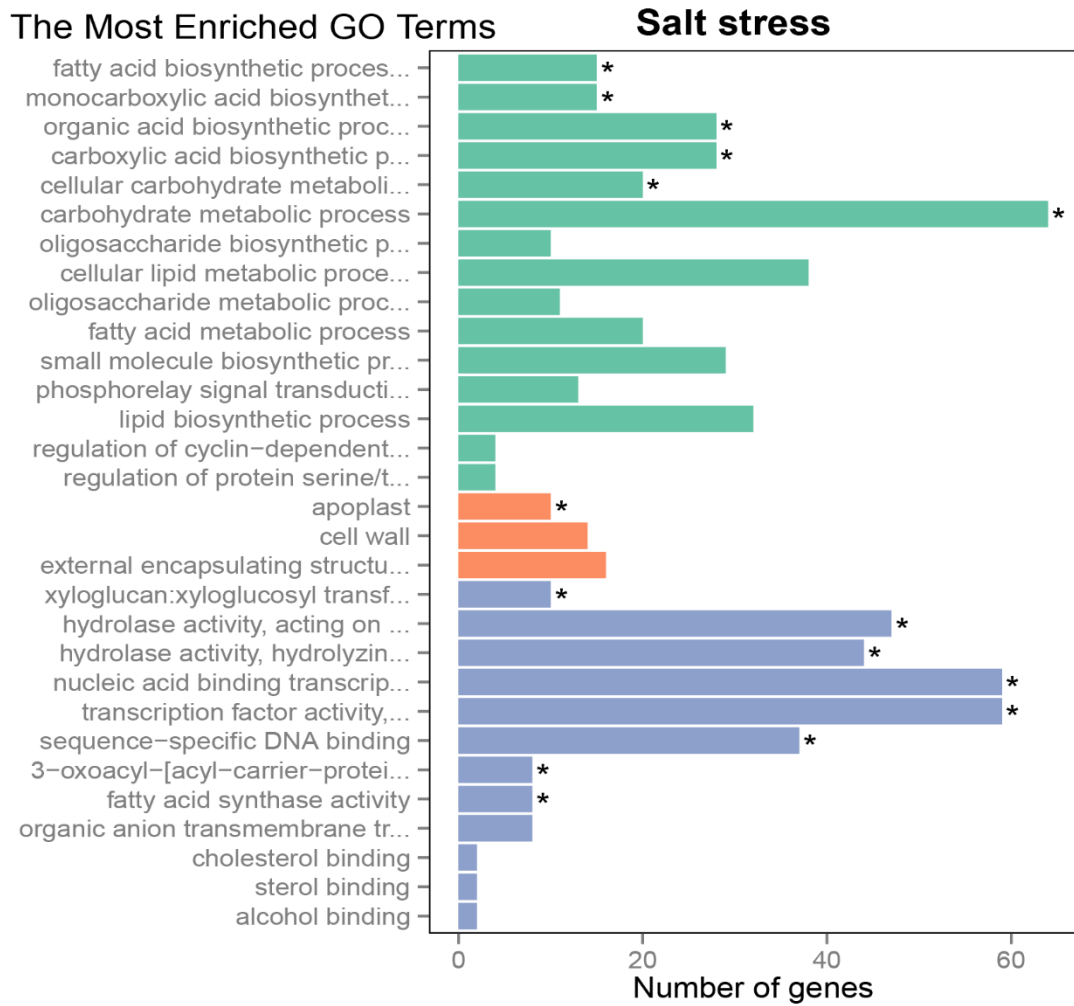


Fig. S17. GO enrichment analysis of shared DEGs under salt treatments at two time periods (12 h and 24 h) for *Ghp*.

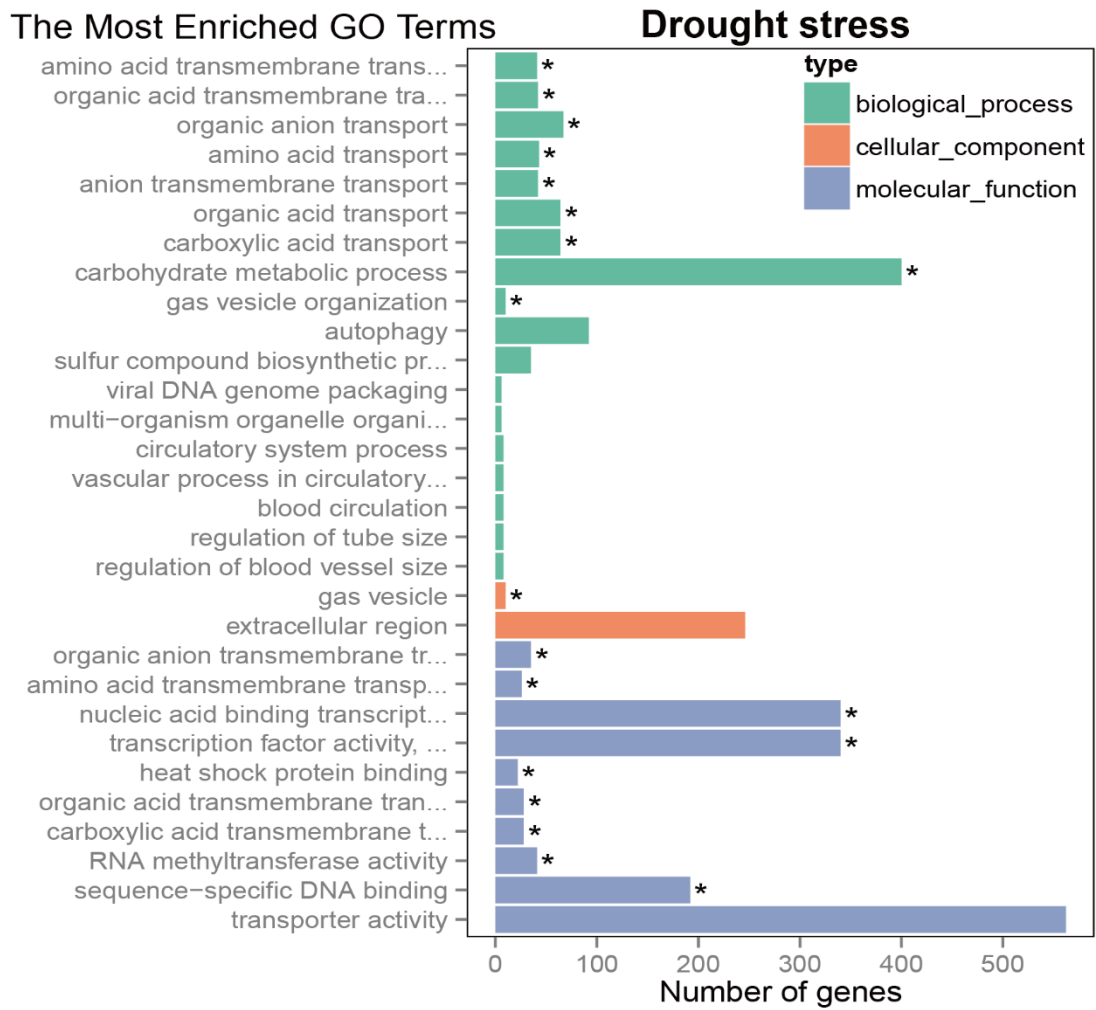


Fig. S18. GO enrichment analysis of shared DEGs under drought treatments at two time periods (12 h and 24 h) for *Ghp*.

Table S1. Summary of PacBio sequencing data

Sequencing features	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
Number of flow cells	6	41	51
High quality bases (Gb)	255.52	255.07	260.20
High quality reads	22,527,280	20,829,397	23,662,273
Average length (bp)	11,342	12,245	10,996
Max length (bp)	242,683	112,330	103,537
Reads N50* (bp)	13,915	13,186	17,296

*PacBio reads were accumulated from long to short; when the accumulated length of the reads reaches 50% of the total reads, the length of this read is Reads N50.

Table S2. Summary of sequencing data derived from the Illumina platform

Sample	Category	Raw data (Gb)	Read length (bp)	High quality data			
				Base (Gb)	Pair number	Proportion of Q30* (%)	Coverage depth (X)
<i>Ge</i>	350 bp	249.84	150	249.64	832,144,029	92.26	106
	Hi-C	274.44	150	270.19	900,629,846	93.41	117
	Total	524.28	150	519.83	1,732,773,875	92.86	223
<i>Gs</i>	350 bp	290.44	150	289.93	966,445,047	96.00	127
	Hi-C	276.59	150	274.22	914,062,537	95.68	120
	Total	567.04	150	564.15	1,880,507,584	95.84	247
<i>GhP</i>	350 bp	293.89	150	293.41	978,038,196	96.15	128
	Hi-C	275.77	150	271.82	906,067,161	95.35	120
	Total	569.66	150	565.23	1,884,105,357	95.77	248

*Q30: The Phred quality is 30.

Table S3. Statistics of the corrected assembly

Sample	Type	Length (bp)	Number	N50 (bp)	N90 (bp)	Max length (bp)	GC content (%)	Gap length (bp)
<i>Ge</i>	Scaffold	2,341,873,899	160	108,057,877	62,932,355	132,145,079	34.34	362,100
	Contig	2,341,511,799	3,781	1,567,542	396,068	8,007,102		
<i>Gs</i>	Scaffold	2,291,836,038	243	108,204,208	59,415,676	125,976,056	34.45	368,400
	Contig	2,291,467,638	3,927	1,233,739	354,533	8,587,061		
<i>GhP</i>	Scaffold	2,292,484,390	277	106,956,574	62,054,521	129,207,784	34.45	83,400
	Contig	2,292,400,990	1,111	11,489,524	2,913,997	58,819,365		

Note: Contigs or Scaffolds were accumulated from long to short; when the accumulated length of the Contig/Scaffold reaches 50% of the total genome size, the length of this read is Contig/Scaffold N50; when the accumulated length of the Contig/Scaffold reaches 90% of the total genome size, the length of this read is Contig/Scaffold N90.

Table S4. Summary of Illumina Hi-C sequences

Sample	Read pair numbers	Base number (bp)	GC content (%)	Q20 ratio (%)	Q30 ratio (%)	Mapped ratio (%)	Unique mapped read pairs
<i>Ge</i>	900,629,846	270,188,953,800	36.49	97.67	93.41	97.70	20,691,111
<i>Gs</i>	914,062,537	274,218,761,100	36.16	98.50	95.68	97.45	23,352,325
<i>GhP</i>	906,067,161	271,820,148,300	37.47	98.34	95.35	96.46	32,061,238

Table S5. Evaluation of valid data for Hi-C sequences

Type	Number (ratio %)		
	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
Total pairs processed	900,629,846 (100)	914,062,537 (100.00)	906,067,161 (100.00)
Unique paired alignments	393,911,811 (43.74)	414,551,254 (45.35)	293,713,316 (32.42)
Valid interaction pairs	375,173,085 (41.66)	398,867,601 (43.64)	254,363,107 (28.07)
Dangling end pairs	10,408,999 (0.01)	10,570,384 (1.16)	22,259,030 (2.46)
Re-ligation pairs	7,854,487 (0.01)	4,776,060 (0.52)	16,527,553 (1.82)
Self-cycle pairs	456,317 (<0.01)	322,869 (0.04)	537,653 (0.06)
Dumped pairs	18,923 (<0.01)	14,340 (<0.01)	25,973 (<0.01)

Table S6. Summary of mapping and coverage depth for realign to genome

Category	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
High quality reads	527,796,298	962,364,810	990,365,051
Mapping rate (%)	99.45	99.68	99.55
Average sequencing depth (X)	32.27	60.57	64.01
Coverage (%)	98.97	97.47	99.82
Coverage at least 10X (%)	95.91	96.01	99.66

Table S7. Assessment of the core gene coverage rate in assembly using BUSCO

Sample	BUSCO assessment results
<i>Ge</i>	C:99.2% [S:13.1%,D:86.1%],F:0.1%,M:0.7%,n:1614
<i>Gs</i>	C:97.4% [S:16.0%,D:81.4%],F:0.1%,M:2.5%,n:1614
<i>GhP</i>	C:99.3% [S:12.1%,D:87.2%],F:0.1%,M:0.6%,n:1614

Table S8. Assessment of the LTR completeness using LAI

Sample	LTR length (Mb)	LTR counts	Complete LTR length (Mb)	Complete LTR counts	LAI
<i>Ge</i>	1,140.52	66,450	155.74	18,030	13.66
<i>Gs</i>	1,079.63	61,535	137.81	16,279	12.76
<i>GhP</i>	1,199.86	63,723	151.87	17,709	12.66

Table S9. Summary of the transposon element for three cotton genomes

Classification	Length (Mb) (ratio %)		
	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
Transposon element	1,575.35 (64.86)	1,489.20 (63.01)	1,487.58 (64.89)
Class I: Retroelement	1,512.39 (62.27)	1,429.24 (60.47)	1,429.29 (62.35)
LTR Retroelement	1,477.80 (60.84)	1,396.48 (59.09)	1,398.55 (61.01)
Gypsy	1,169.24 (48.14)	1,117.97 (47.30)	1,117.61 (48.75)
Copia	162.92 (6.71)	156.21 (6.61)	155.66 (6.79)
Other	145.65 (6.00)	122.30 (5.17)	125.29 (5.47)
Non-LTR retrotransposons	34.59 (1.42)	32.76 (1.39)	30.73 (1.34)
LINE	34.32 (1.41)	32.45 (1.37)	30.47 (1.33)
SINE	0.27 (0.01)	0.30 (0.01)	0.26 (0.01)
Class II: DNA transposons	54.18 (2.23)	50.76 (2.15)	48.15 (2.10)
DNA transposons superfamily	49.80 (2.05)	46.67 (1.97)	43.32 (1.89)
CACTA	14.34 (0.59)	15.45 (0.65)	11.66 (0.51)
hAT	7.44 (0.31)	6.71 (0.28)	6.74 (0.29)
MULE	21.23 (0.87)	19.02 (0.80)	17.45 (0.76)
PIF–Harbinger	5.26 (0.22)	4.22 (0.18)	5.80 (0.25)
Ginger	1.11 (0.05)	1.03 (0.04)	1.13 (0.05)
Tc1/Mariner	0.41 (0.02)	0.25 (0.01)	0.54 (0.02)
Maverick	0.69 (0.03)	0.12 (0.01)	0.30 (0.01)
Helitron	2.10 (0.09)	2.47 (0.10)	3.17 (0.14)
Other DNA transposons	1.60 (0.07)	1.49 (0.06)	1.36 (0.06)
Other/Unspecified/Unknown	8.77 (0.36)	9.20 (0.39)	10.15 (0.44)

Table S10. General statistics of predicted protein-coding genes

Statistical item	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
Number of genes	74,178	74,970	74,520
Genes anchored to pseudomolecules	74,038	73,324	74,283
Average mRNA length (CDS+intron)	2775.64	2917.46	2951.37
Average CDS length	1105.35	1185.42	1176.78
Average exon length	226.28	236.73	235.20
Average exon number per gene	4.88	5.01	5.00
Average intron length	429.95	432.20	431.20

Table S11. Statistics of gene functional annotation

Database	Number (ratio %)		
	<i>Ge</i>	<i>Gs</i>	<i>GhP</i>
NR	71,673 (96.60)	73,467 (98.00)	72,652 (97.50)
SwissProt	55,609 (75.00)	59,940 (80.00)	57,789 (77.50)
KEGG	52,592 (70.90)	56,893 (75.90)	55,490 (74.50)
GO	41,141 (55.50)	44,944 (59.90)	42,778 (57.40)
Pfam	55,030 (74.20)	59,615 (79.50)	58,054 (77.90)
Annotated	71,718 (96.70)	73,488 (98.00)	72,672 (97.50)
Total	74,178	74,970	74,520

Table S12. Homologous PCG models analysis in allotetraploid cotton

Identity \geq 80	<i>Gh</i>	<i>Gb</i>	<i>Gt</i>	<i>Gm</i>	<i>Gd</i>	Total
<i>Ge</i>	69,233	69,356	69,899	69,407	69,813	72,213
<i>Gs</i>	68,776	68,893	69,051	68,656	69,163	71,032
<i>GhP</i>	68,465	68,450	68,033	67,575	68,232	70,850

Table S13. Summary of identified SVs in seven *Gossypium* genomes

Category	PAV number		Inversion number	PAV length (bp)		Inversion length (bp)	PAV density (number/Mb)		Inversion density (number/Mb)	
	Insertion	Deletion		Insertion	Deletion		Insertion	Deletion		
<i>Gh</i>	At	34,394	30,481	263	156,359,316	178,188,975	131,028,104	23.59	20.90	0.18
	Dt	37,277	29,182	105	108,782,320	103,241,816	40,238,909	44.66	34.96	0.13
	ALL	71,671	59,663	368	265,141,636	281,430,791	171,267,013	31.26	26.02	0.16
<i>Gb</i>	At	37,354	35,434	245	131,100,632	199,396,556	92,525,868	25.62	24.30	0.17
	Dt	34,212	29,641	104	81,410,720	114,166,188	52,866,107	40.99	35.51	0.12
	ALL	71,566	65,075	349	212,511,352	313,562,744	145,391,975	31.21	28.38	0.15
<i>Gt</i>	At	33,846	32,682	205	116,035,179	164,453,279	78,370,756	23.21	22.41	0.14
	Dt	34,039	28,744	95	89,850,072	106,788,148	34,701,047	40.78	34.44	0.11
	ALL	67,885	61,426	300	205,885,251	271,241,427	113,071,803	29.61	26.79	0.13
<i>Gd</i>	At	35,997	34,771	191	117,104,253	175,450,982	55,911,062	24.69	23.85	0.13
	Dt	36,324	30,899	108	96,073,570	114,653,042	17,794,554	43.52	37.02	0.13
	ALL	72,321	65,670	299	213,177,823	290,104,024	73,705,616	31.54	28.64	0.13
<i>Ge</i>	At	42,969	35,726	282	164,576,896	156,948,804	132,958,646	29.47	24.50	0.19
	Dt	34,787	26,902	128	113,786,877	92,342,342	46,656,970	41.68	32.23	0.15
	ALL	77,756	62,628	410	278,363,773	249,291,146	179,615,616	33.91	27.31	0.18
<i>Gs</i>	At	39,461	35,531	241	120,461,733	160,990,669	139,774,502	27.06	24.37	0.17
	Dt	34,263	30,032	107	97,269,075	104,393,226	26,939,617	41.05	35.98	0.13
	ALL	73,724	65,563	348	217,730,808	265,383,895	166,714,119	32.15	28.59	0.15
<i>GhP</i>	At	41,534	35,022	207	148,058,995	155,911,239	120,889,188	28.48	24.02	0.14
	Dt	34,297	26,835	90	119,779,357	94,673,600	38,899,757	41.09	32.15	0.11
	ALL	75,831	61,857	297	267,838,352	250,584,839	159,788,945	33.07	26.98	0.13

Table S14. Distribution of fragment size of PAVs in *Gossypium* genomes

			Fragment size (bp) of PAVs			
			50–100	100–500	500–1000	≥ 1000
<i>Gh</i>	Total	Number	42,350	33,848	7,475	47,661
		Rate	32.25%	25.77%	5.69%	36.29%
	At	Number	20,161	15,657	3,511	25,546
		Rate	31.08%	24.13%	5.41%	39.38%
	Dt	Number	22,189	18,191	3,964	22,115
		Rate	33.39%	27.37%	5.96%	33.28%
<i>Gb</i>	Total	Number	46,105	35,498	7,789	47,249
		Rate	33.74%	25.98%	5.70%	34.58%
	At	Number	24,402	18,006	3,889	26,491
		Rate	33.52%	24.74%	5.34%	36.39%
	Dt	Number	21,703	17,492	3,900	20,758
		Rate	33.99%	27.39%	6.11%	32.51%
<i>Gt</i>	Total	Number	39,973	32,907	7,633	48,798
		Rate	30.91%	25.45%	5.90%	37.74%
	At	Number	19,841	16,123	3,655	26,909
		Rate	29.82%	24.23%	5.49%	40.45%
	Dt	Number	20,132	16,784	3,978	21,889
		Rate	32.07%	26.73%	6.34%	34.86%
<i>Gd</i>	Total	Number	43,126	35,474	8,296	51,095
		Rate	31.25%	25.71%	6.01%	37.03%
	At	Number	21,204	17,294	4,083	28,187
		Rate	29.96%	24.44%	5.77%	39.83%
	Dt	Number	21,922	18,180	4,213	22,908
		Rate	32.61%	27.04%	6.27%	34.08%
<i>Ge</i>	Total	Number	44,514	35,986	7,957	51,927
		Rate	31.71%	25.63%	5.67%	36.99%
	At	Number	24,475	19,366	4,325	30,529
		Rate	31.10%	24.61%	5.50%	38.79%
	Dt	Number	20,039	16,620	3,632	21,398
		Rate	32.48%	26.94%	5.89%	34.69%
<i>Gs</i>	Total	Number	45,093	35,753	8,120	50,321
		Rate	32.37%	25.67%	5.83%	36.13%
	At	Number	23,625	18,644	4,218	28,505
		Rate	31.50%	24.86%	5.62%	38.01%
	Dt	Number	21,468	17,109	3,902	21,816
		Rate	33.39%	26.61%	6.07%	33.93%
<i>GhP</i>	Total	Number	41,544	34,545	8,031	53,568
		Rate	30.17%	25.09%	5.83%	38.91%
	At	Number	22,781	18,517	4,285	30,973
		Rate	29.76%	24.19%	5.60%	40.46%
	Dt	Number	18,763	16,028	3,746	22,595
		Rate	30.69%	26.22%	6.13%	36.96%

Table S15. Annotation of PAVs in seven *Gossypium* genomes

Category	<i>Gh</i>		<i>Gb</i>		<i>Gt</i>		<i>Gd</i>		<i>Ge</i>		<i>Gs</i>		<i>GhP</i>	
	At	Dt	At	Dt	At	Dt	At	Dt	At	Dt	At	Dt	At	Dt
Intergenic	46,901	47,587	54,452	46,405	50,561	46,763	53,944	49,780	58,674	43,508	56,116	46,704	58,808	44,719
Upstream/downstream	4,458	7,594	4,215	6,042	3,874	5,871	3,933	6,856	5,909	6,782	5,512	6,810	5,112	6,294
Intronic	1,973	3,066	2,227	2,836	1,678	2,372	1,977	2,389	2,593	2,972	2,325	2,807	2,205	2,579
Splice sites	39	42	37	38	29	39	37	30	44	48	47	37	48	41
Stop gain	666	790	663	836	440	619	592	667	759	843	762	704	669	760
Stop loss	-	-	-	-	-	-	-	-	1	-	-	-	-	-
Exon														
Frameshift deletion	1,073	1,383	1,151	1,507	859	1,312	981	1,403	1,243	1,448	1,138	1,430	931	1,160
Frameshift insertion	66	90	66	100	32	65	45	69	64	87	49	76	42	64
Nonframeshift deletion	115	175	144	219	85	132	125	129	144	199	151	175	104	148
Nonframeshift insertion	19	31	36	34	14	12	31	20	36	40	21	25	19	26
Total	64,875	66,459	72,788	63,853	66,528	62,783	70,768	67,223	78,695	61,689	74,992	64,295	76,556	61,132

Table S16. Construction of pan-genome gene family

Species	Total number		Core set number		Dispensable set number		Specific set number	
	Families	Genes	Families	Genes	Families	Genes	Families	Genes
<i>Gm</i>	43,319	74,660	27,384	52,687	12,519	18,552	3,416	3,421 (4.58%)
<i>Gd</i>	45,815	78,303	27,384	52,505	15,055	22,415	3,376	3,383 (4.32%)
<i>Gb</i>	42,528	71,297	27,384	51,289	12,188	17,048	2,956	2,960 (4.15%)
<i>Gt</i>	46,236	78,281	27,384	52,641	14,822	21,598	4,030	4,042 (5.16%)
<i>Ge</i>	48,818	76,645	27,384	50,285	13,759	18,674	7,675	7,686 (10.03%)
<i>Gs</i>	47,290	74,970	27,384	51,296	13,171	16,914	6,735	6,760 (9.02%)
<i>GhP</i>	47,855	74,520	27,384	49,408	13,911	18,532	6,560	6,580 (8.83%)
<i>Gh</i>	45,309	70,199	27,384	47,547	12,018	16,740	5,907	5,912 (8.42%)

Table S17. Total RGAs in eight tetraploid cottons genomes

Species	Has a NB-ARC or TIR domain (C, CC; N, NBS; L, LRR; T, TIR; X, other)						Has a TM domain			Total
	CN	CNL	NBS	NL	TN	TX	RLK	RLP	TM-CC	
<i>Gm</i>	75	325	128	223	131	39	720	367	1,731	3,739
<i>Gd</i>	84	360	143	218	143	46	744	370	1,773	3,881
<i>Gb</i>	70	252	134	121	81	39	717	288	1,834	3,536
<i>Gt</i>	90	366	136	228	125	48	728	379	1,767	3,867
<i>Ge</i>	125	350	207	124	108	46	734	515	2,103	4,312
<i>Gs</i>	142	362	225	105	114	44	705	443	2,104	4,244
<i>GhP</i>	121	357	189	105	105	40	701	447	2,034	4,099
<i>Gh</i>	79	242	136	148	84	44	678	304	1,747	3,462

Table S18. Sample information for RNA-seq

Sample ID	Species	Treatment	Time point
T1PS0R1	<i>G. hirsutum</i> race <i>Punctatum</i>	Water	0 h
T1PS0R2	<i>G. hirsutum</i> race <i>Punctatum</i>	Water	0 h
T1PS0R3	<i>G. hirsutum</i> race <i>Punctatum</i>	Water	0 h
T1PS12R1	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	12 h
T1PS12R2	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	12 h
T1PS12R3	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	12 h
T1PS12P1	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	12 h
T1PS12P2	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	12 h
T1PS12P3	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	12 h
T1PS24R1	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	24 h
T1PS24R2	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	24 h
T1PS24R3	<i>G. hirsutum</i> race <i>Punctatum</i>	300 mM salt	24 h
T1PS24P1	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	24 h
T1PS24P2	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	24 h
T1PS24P3	<i>G. hirsutum</i> race <i>Punctatum</i>	17% PEG	24 h

Table S19. Primer sequence information

Primer	Genome ID	Sequences (5'→3')
1-72	Ge_A04:76488699-89236152	FAM-TTCAACTCATATTGCGAGAAATGAGTTGAGCCTCAAG
3-44	Ge_A04:76488699-89236152	TAMRA-GATTTGGCGTCTCTGTAGCGGCGGAGAGCAGATCGAAGATAACA
16-44	Ge_A11:120648664-131080377	FAM-AATCTGCTCCACTACTGCTTAGGGAGATAAGATTCACTATCTTC
19-44	Ge_A11:120648664-131080377	TAMRA-ATCTGCTCCACTGTAACCTCAGAGAGATAAGATCTCTGGCTTCA
22-50	Ge_D11:69231246-74401191	FAM-GATTTGGCGTCTCTGTAGCGGCGGAGAGCAGATCGAAGATAACA
49-54	Ge_A11:120648664-131080377	TAMRA-AAAAATCAACTCATATTGCGAGAGGTGAGTTGAGCC

Dataset S1. The PAVs located in the exonic region

Dataset S2. The genes located in inversion region D01:27,928,130-32,409,098 and D04:57,489,773-58,476,195.