

Figure S1: Domain and exon-intron organization of identified LRR-RLK family gene members in *A. thaliana* and *Gossypium*.

Exon-intron structures of all identified LRR-RLK genes from four *Gossypium* species and *A. thaliana*. LRR, KD and TM domain coding regions were marked on exons by different colored rectangles. 21 subfamilies were distinguished by different color branches.

(see file Figure S1.pdf)

Figure S2: NJ tree constructed by MEGA 7 based on LRR-RLK family members of *A. thaliana* and *Gossypium*.

Amino acid sequences of LRR-RLKs from *A. thaliana* and *Gossypium* were aligned by MUSCLE software with default parameters. Aligned sequences were used to construct NJ tree by MEGA 7 with 1000 bootstraps. Only topology of tree was displayed. Bootstrap values lower than 50 were hidden.

(see file Figure S2.pdf)

Figure S3: Chromosomal location of LRR-RLK genes from four *Gossypium* species.

LRR-RLK genes from *G. arboreum*, *G. barbadense*, *G. hirsutum*, *G. raimondii* were mapped to chromosomes based on chromosomal coordinates extracted from gene feature annotation files (shown by A, B, C, D, respectively). LRR-RLK genes located on scaffolds were not shown. Tandem duplication gene sets were highlighted by red border rectangles to represent.

(see file Figure S3.pdf)

Table S1: Statistics of *A. thaliana* and *Gossypium* LRR-RLK genes distribution among different subfamilies.

For both *A. thaliana* and four *Gossypium* species, number of LRR-RLK genes belonged to each subfamily was counted respectively, the corresponding percentage of subfamily in all LRR-RLK from specific species was computed and indicated in brackets.

Table S2: Number and percentage of tandem duplication genes in *Gossypium* LRR-RLK family.

For each LRR-RLK subfamily, the number of tandem duplication genes was counted, the percentage of tandem duplication genes out of all subfamily members was computed and indicated in brackets.

Table S3: Statistics of cis-acting regulatory elements found by PlantCARE in promoter regions of *Gossypium* LRR-RLK genes.

Based on PlantCARE analysis results, 121 kinds of cis-acting regulatory elements were found in promoters of *Gossypium* LRR-RLK genes. Number of each kind of cis-acting elements found in specific species was counted.

Table S4: Statistics of TF binding sites predicted in promoter regions of *Gossypium* LRR-RLK genes.

TF binding sites analysis showed that *Gossypium* LRR-RLK genes could be regulated by TFs belonged to as many as 39 TF families. Number of LRR-RLK genes regulated by each TF family were counted for *G. arboreum*, *G. hirsutum* and *G. raimondii*, respectively.

Table S1

Subfamily	Number (percentage) of LRR-RLK genes in each subfamily				
	<i>A. thaliana</i>	<i>G. arboreum</i>	<i>G. barbadense</i>	<i>G. hirsutum</i>	<i>G. raimondii</i>
I	41 (19.2%)	4 (1.3%)	6 (1.2%)	7 (1.4%)	3 (0.9%)
II	14 (6.6%)	17 (5.7%)	23 (4.5%)	29 (5.6%)	18 (5.7%)
III	41 (19.2%)	45 (15.1%)	68 (13.3%)	83 (16.1%)	46 (14.5%)
IV	3 (1.4%)	5 (1.7%)	10 (2.0%)	10 (1.9%)	5 (1.6%)
V	9 (4.2%)	10 (3.4%)	15 (2.9%)	19 (3.7%)	9 (2.8%)
VI-1	5 (2.3%)	7 (2.3%)	11 (2.2%)	14 (2.7%)	7 (2.2%)
VI-2	4 (1.9%)	2 (0.7%)	2 (0.4%)	6 (1.2%)	3 (0.9%)
VII-1	2 (0.9%)	4 (1.3%)	11 (2.2%)	7 (1.4%)	4 (1.3%)
VII-2	5 (2.3%)	6 (2.0%)	6 (1.2%)	12 (2.3%)	6 (1.9%)
VIII-1	8 (3.8%)	4 (1.3%)	7 (1.4%)	7 (1.4%)	4 (1.3%)
VIII-2	12 (5.6%)	17 (5.7%)	26 (5.1%)	30 (5.8%)	17 (5.4%)
IX	4 (1.9%)	12 (4.0%)	18 (3.5%)	19 (3.7%)	11 (3.5%)
X	15 (7.0%)	18 (6.0%)	32 (6.3%)	35 (6.8%)	18 (5.7%)
XI-1	28 (13.1%)	75 (25.2%)	135 (26.4%)	132 (25.6%)	81 (25.6%)
XI-2	2 (0.9%)	3 (1.0%)	3 (0.6%)	7 (1.4%)	3 (0.9%)
XI-3	2 (0.9%)	1 (0.3%)	1 (0.2%)	2 (0.4%)	1 (0.3%)
XII	7 (3.3%)	49 (16.4%)	102 (20.0%)	61 (11.8%)	63 (19.9%)
XIII-1	3 (1.4%)	3 (1.0%)	6 (1.2%)	5 (1.0%)	2 (0.6%)
XIII-2	3 (1.4%)	3 (1.0%)	6 (1.2%)	6 (1.2%)	3 (0.9%)
XIV	2 (0.9%)	5 (1.7%)	10 (2.0%)	9 (1.7%)	5 (1.6%)
XV	3 (1.4%)	8 (2.7%)	13 (2.5%)	15 (2.9%)	8 (2.5%)

Table S2

Subfamily	Number of tandem duplication genes			
	<i>G. arboreum</i>	<i>G. barbadense</i>	<i>G. hirsutum</i>	<i>G. raimondii</i>
II	5 (29.4%)	2 (8.7%)	5 (17.2%)	5 (27.8%)
III	2 (4.4%)	4 (5.9%)	4 (4.8%)	2 (4.3%)
IX		2 (11.1%)		
VII-1		4 (36.4%)		
VIII-2	10 (58.8%)	11 (42.3%)	14 (46.7%)	11 (64.7%)
XI-1	31 (41.3%)	59 (43.7%)	42 (31.8%)	35 (43.2%)
XI-2			2 (28.6%)	
XII	31 (63.3%)	64 (62.7%)	25 (41.0%)	43 (68.3%)

Table S3

Cis-acting regulatory element name (ID of CARE)	Number of elements found in promoters of LRR-RLK genes			
	<i>G. arboreum</i>	<i>G. barbadense</i>	<i>G. hirsutum</i>	<i>G. raimondii</i>
CAAT-box	298	509	515	310
TATA-box	298	504	514	309
AT-TATA-box	243	413	407	265
TATA	169	268	256	178
A-box	25	42	48	23
Myb	285	469	492	288
AT-rich element	41	71	62	55
MYC	279	450	479	274
as-1	149	216	228	150
MBS	101	166	196	95
DRE core	17	28	33	24
ACTCATCCT sequence	5	6	6	3
DRE1	12	18	17	10
MYB recognition site	50	69	76	57
LTR	79	150	168	98
STRE	202	337	347	177
TCA	80	133	143	68
ARE	236	378	402	229
GC-motif	16	34	32	23
TC-rich repeats	110	167	181	108
W box	149	256	252	158
WUN-motif	128	209	213	128
WRE3	90	157	173	94
box S	27	42	45	23
W box	149	256	252	158
GT1-motif	178	296	305	192
TCT-motif	144	237	248	130
GATA-motif	117	184	191	113
MRE	92	136	141	95
AE-box	86	144	167	76
I-box	69	111	103	65
TCCC-motif	62	88	106	59
ATCT-motif	59	93	95	55
AT1-motif	49	75	77	44
GA-motif	47	84	74	53
chs-CMA1a	46	98	99	57
LAMP-element	32	56	59	27
3-AF1 binding site	29	44	48	28
ACE	24	50	48	45

Gap-box	23	24	31	15
Sp1	23	42	52	32
chs-CMA2a	19	32	27	16
ATC-motif	13	21	28	14
Box II	12	26	28	21
AAAC-motif	11	10	12	6
GTGGC-motif	7	8	10	3
CAG-motif	5	9	8	9
ACA-motif	4	8	9	9
chs-Unit 1 m1	3	5	5	5
L-box	3	9	8	5
LS7	2	6	4	4
4cl-CMA2b	1	2	2	1
Pc-CMA2c	1	2	2	2
circadian	36	83	79	52
ERE	226	354	370	242
ABRE	160	268	299	175
ABRE3a	56	81	101	68
ABRE4	56	81	101	68
ABRE2	2	3	6	2
AT-ABRE	1	6	5	6
P-box	68	117	130	68
TATC-box	43	63	70	41
GARE-motif	39	59	65	35
CARE	19	39	40	21
CGTCA-motif	149	216	228	150
TGACG-motif	149	216	228	150
JERE	2	3	4	2
TCA-element	133	191	218	112
TGA-element	60	93	87	71
TGA-box	10	11	12	10
AuxRR-core	23	50	48	35
AuxRE	1	1	2	2
Myb-binding site	116	197	200	112
CCGTCC motif	24	40	45	21
MSA-like	11	20	18	9
re2f-1	4	7	4	3
NON	3	3	5	3
dOCT	3	6	7	3
E2Fb	2	4	7	4
O2-site	80	126	137	80

MBSI	28	41	39	24
AP-1	17	22	27	16
CAT-box	71	128	132	89
MYB-like sequence	171	283	299	166
GCN4_motif	49	66	73	42
RY-element	13	20	26	17
AC-I	10	9	14	6
AC-II	5	8	9	3
AACA_motif	3	3	7	3
telo-box	2	2	5	2
motif I	1	1	2	3
Box III	20	19	26	14
HD-Zip 3	20	21	23	10
AT-rich sequence	25	33	30	25
HD-Zip 1	15	24	24	13
Unnamed__4	293	499	510	304
Unnamed__6	69	105	117	74
Unnamed__1	136	252	247	146
Unnamed__2	34	55	57	38
Unnamed__10	6	14	14	15
Unnamed__12	6	14	14	15
Unnamed__14	6	14	14	15
Unnamed__8	6	14	14	15
Unnamed__16	2	1	1	1
Unnamed__3	1	1	3	3
AAGAA-motif	210	339	373	186
CCAAT-box	50	69	76	57
CCGTCC-box	24	40	45	21
F-box	22	32	40	21
CTAG-motif	21	45	42	23
3-AF3 binding site	9	7	13	6
Box II -like sequence	1	2	3	1

Table S4

TF family	Number of TFs regulated LRR-RLK gene		
	<i>G. arboreum</i>	<i>G. hirsutum</i>	<i>G. raimondii</i>
AP2	188	335	159
ARF	35	83	49
ARR-B	8	10	3
B3	42	78	58
BBR-BPC	137	275	107
BES1	23	42	35
bHLH	98	197	106
bZIP	73	144	82
C2H2	172	315	172
C3H	6	13	6
CAMTA	15	28	20
CPP	19	43	36
Dof	231	418	215
E2F/DP	11	18	10
EIL	13	29	14
ERF	156	301	150
FAR1	21	20	14
G2-like	57	114	67
GATA	99	168	84
GRAS	132	244	97
HD-ZIP	62	142	62
HSF	25	38	20
LBD	63	123	59
LFY	3	4	4
MIKC_MADS	213	373	193
MYB	195	354	200
MYB_related	49	93	63
NAC	100	169	87
Nin-like	44	115	41
RAV	14	25	14
SBP	13	33	19
SRS	4	9	4
TALE	107	195	89
TCP	101	164	114
Trihelix	84	141	77
WOX	18	41	30
WRKY	93	150	83
YABBY	2	6	5
ZF-HD	28	21	22