

GENOME-WIDE BIOINFORMATIC AND FUNCTIONAL ANALYSIS OF
RECEPTOR-LIKE PROTEINS IN COTTON

A Thesis

by

MUSTAFA CILKIZ

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Libo Shan
Committee Members,	Ping He
	David M. Stelly
Head of Department,	Dirk B. Hays

May 2017

Major Subject: Molecular & Environmental Plant Science

Copyright 2017 Mustafa Cilkiz

ABSTRACT

Cotton (*Gossypium. Spp.*) is an agroeconomically significant crop worldwide involved in textile, oil, and feedstock production. Identifying genetic targets involved with disease resistance in cotton remains a critical objective for the scientific community. There remains much to be elucidated about how plant cells communicate and recognize biotic and abiotic environmental cues. Cell-surface-associated pattern recognition receptors (PRR) play a critical role in pathogen recognition. In plants, PRRs are receptor-like kinases (RLKs) and receptor-like proteins (RLPs). In particular, leucine-rich repeat containing RLPs (LRR-RLPs) have been indicated to play an important role in disease resistance against *Fusarium* infection in *Arabidopsis*. My thesis research is to genome-wide identify cotton RLPs with a bioinformatics approach and characterize their potential involvement in cotton disease resistance using virus-induced gene silencing (VIGS) approach. The 57 *Arabidopsis* RLPs were used to identify the cotton RLPs via sequence homology. Analysis of the cotton genome has identified 86 LRR-RLPs in *G. arboreum*, 107 LRR-RLPs in *G. raimondii*, and 151 LRR-RLPs in *G. hirsutum*. Phylogenetic tree analysis with other plant LRR-RLPs suggested that cotton LRR-RLPs are likely evolved independently and clustered together. LRR-RLPs appear to be evolved rapidly as the homology between cotton and *Arabidopsis* LRR-RLPs is often below 40%. The number of LRRs varies among different LRR-RLP members. Using the VIGS approach, we silenced a set of cotton RLPs. My preliminary assays suggested that GhRLPGSO1-like, GhRLP44, GhRLP6, and GhRLP34 might be required

for defense against *Fusarium oxysporum*. Additional infection assays and evaluation of silencing efficiency will further evaluate their involvement in disease resistance.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Libo Shan, and my committee members, Dr. Ping He, and Dr. David M. Stelly for their guidance and support throughout this research. I would like to thank the Ministry of Food, Agricultural, and Livestock and Department of the General Directorate Agricultural Research and Policy in Turkey. I would like to thank also go to my friends and the department faculty and staff for making my time at Texas A&M University a great experience. I would like to thank to Kevin Babilinia, Ping Wang, Dr. Maoying Li, Dr. Fangjun Li, Kevin Cox, Pierce Jamieson, and colleagues. Finally, I would like to thank to my family for their encouragement.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee consisting of Professor Dr. Libo Shan and Dr. David M. Stelly of the Department of Molecular and Environmental Plant Science and Professor Dr. Ping He of the Department of Biochemistry and Biophysics.

All other work conducted for the thesis was completed by the student independently. Mr. Cilkiz is sponsored by the Ministry of National Education of Turkey.

NOMENCLATURE

MYA	Million years ago
n	haploid
AA	<i>G. arboreum</i> genome
DD	<i>G. raimondii</i> genome
AADD	<i>G. hirsutum</i> genome
R	Resistance
Avr	Avirulence
eLRR	Extracellular leucine-rich repeat
RLK	Receptor-like kinase
RLP	Receptor-like protein
PRR	Pattern recognition receptor
PAMP	Pathogen-associated molecular pattern
SOBIR1	Suppressor Of brassinosteroids insensitive 1
LRR	Leucine-rich repeat
BAK1	Brassinosteroids insensitive 1-associated kinase-1
SERK1	Somatic embryogenesis receptor kinase 1
SP	Signal peptide
LRRNT	Leucine-rich repeat N-terminus
LRRCT	Leucine-rich repeat C-terminus
IsD	Island domain

ID	Identification number
AD	Acidic domain
TM	Transmembrane domain
CT	Cytoplasmic tail
YXXØ	Tyrosine-based sorting signal
AtRLP	Arabidopsis receptor-like protein
Cf	Cladosporium fulvum
EIX1	Ethylene-inducing xylanase 1
EIX2	Ethylene-inducing xylanase 2
HcrVf2	Homologue of the <i>C. fulvum</i> resistance genes of the Vf region
LepR3	Leptosphaeria maculans resistance gene 3
AVRLM1	Leptosphaeria maculans recognition of the elicitor 1
Pst	Puccinia striiformis f. sp. Tritici
Ve1	Verticillium dahlia resistance 1
GbVe1	<i>G. barbandense</i> Verticillium dahlia resistance 1
VIGS	Virus-induced gene silencing
dsRNA	Double-strain RNA
RNA	Ribonucleic acid
pTRV1	Plasmid Tobacco rattle virus 1
PTRV2	Plasmid Tobacco rattle virus 2
RISC	RNA-induced silencing complex
siRNA	Short interfering RNA

TAIR	The Arabidopsis Information Resource website
HMMER	Hidden Markov model
PFAM	The protein families database
SMART	Simple Modular Architecture Research Tool
IPC	Isoelectric point calculator
LRV	Leucine-rich repeat variant
TMHMM	Transmembrane hidden Markov model
MCL	Markov Cluster
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MAFFT	Multiple alignment using fast fourier transform
MEGA6	Molecular Evolutionary Genetics Analysis Version 6.0
ClustalW	multiple sequence alignment through sequence weighting
UPGMA	Unweighted Pair Group Method with Arithmetic
BLOSUM	Blocks substitution matrix
FM	Fibermax
TX	Texas
PCR	Polymerase chain reaction
h	hour
⁰ C	Degree
LB	Luria-Bertani
EcoRI	Escherichia coli restriction enzyme I
KpnI	Klebsiella pneumonia restriction enzyme I

pYL156	VIGS plasmid
μg	microgram
ml	milliliter
mM	millimolar
MgCl ₂	Magnesium chloride
MES	2-(N-morpholino)ethanesulfonic acid
OD	Optical Density
uM	unit-molar
PDB	Potato dextrose broth
rpm	Revolutions per minute
%	Percent
cm	centimeter
dpi	days post inoculation
DI	Disease index
AGI	Arabidopsis Genome Initiative
a.a	Amino acid
pI	Isoelectric point value
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
CDD	Conserved Domain Database
NCBI	National Center for Biotechnology Information
GaRLP	<i>Gossypium arboreum</i> receptor-like protein
GrRLP	<i>Gossypium raimondii</i> receptor-like protein

GhRLP	<i>Gossypium hirsutum</i> receptor-like protein
Da	Dalton
GFP	Green fluorescent protein
CLA1	Cloroplastos alterados 1
GSO1	GASSHO1

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
CONTRIBUTORS AND FUNDING SOURCES.....	v
NOMENCLATURE.....	vi
TABLE OF CONTENTS	xi
LIST OF FIGURES.....	xiii
LIST OF TABLES	xiv
1. INTRODUCTION.....	1
2. METHODS.....	8
2.1 Sequence Database Searches.....	8
2.2 Determination of Protein Domain Structure and pI Value.....	8
2.3 LRRs Domains Identification.....	9
2.4 Transmembrane and Signal Peptide Identification	9
2.5 Orthologs Group in Cotton.....	10
2.6 Multiple Sequence Alignment and Phylogenetic Analysis	10
2.7 Plant Materials And Growth	10
2.8 Construction of VIGS Vectors	11
2.9 Agrobacterium-Mediated VIGS	11
2.10 <i>F. Oxysporum</i> Inoculation and Disease Index	12
3. RESULTS.....	13
3.1 Receptor-like Protein in Arabidopsis	13
3.1.1 Chromosomal Distribution of AtRLPs.....	13
3.2 Receptor-Like Protein in Cotton	19
3.2.1 Identification of RLPs in Cotton	19
3.2.2 GaRLPs Biochemical Features and Chromosomal Distribution.	20
3.2.3 GrRLPs Biochemical Features and Chromosomal Distribution .	27
3.2.4 GrRLPs Biochemical Features and Chromosomal Distribution .	35

3.3	Distribution of LRRs in Arabidopsis and Cotton.....	46
3.4	Classifications and Phylogenetic Analysis of RLPs	47
3.4.1	OrthoMCL Analysis in Cotton.....	47
3.4.2	Determination of Conserved Motifs in RLPs	49
3.4.3	Phylogenetic Trees for Four Species	50
3.4.3.1	Arabidopsis RLPs Phylogenetic Analysis.....	50
3.4.3.2	<i>G. arboreum</i> RLPs Phylogenetic Analysis	51
3.4.3.3	<i>G. raimondii</i> RLPs Phylogenetic Analysis	53
3.4.3.4	<i>G. hirsutum</i> RLPs Phylogenetic Analysis	54
3.4.3.5	Gossypium spp. and Arabidopsis RLPs Phylogenetic Analysis	56
3.5	RLPs Evolves Independently and Rapidly.....	58
3.6	Cloning Receptor-Like Proteins for VIGS.....	60
3.6.1	Determination of Candidate RLPs from GhRLPs.....	60
3.6.2	Silencing of GhRLPs by Agrobacterium-Mediated VIGS.....	62
3.6.3	Phenotype Observation and Disease Severity.....	63
4.	CONCLUSIONS.....	66
	REFERENCES.....	68
	APPENDIX A	76
	APPENDIX B	77
	APPENDIX C	80
	APPENDIX D	83
	APPENDIX E.....	84
	APPENDIX F.....	107

LIST OF FIGURES

	Page
Figure 1. Structure of the various domains of typical RLPs	4
Figure 2. Representative <i>A. thaliana</i> RLPs Chromosomal map	14
Figure 3. Representative <i>G. arboreum</i> RLPs chromosomal map	21
Figure 4. Representative <i>G. raimondii</i> RLPs chromosomal map	27
Figure 5. Representative <i>G. hirsutum</i> A-subgenome RLPs chromosomal map	35
Figure 6. Representative <i>G. hirsutum</i> D-subgenome RLPs chromosomal map	36
Figure 7. Conserved domain structure of RLPs (A-G)	49
Figure 8. Arabidopsis Receptor-like protein tree	51
Figure 9. <i>Gossypium arboreum</i> Receptor-like protein tree	52
Figure 10. <i>Gossypium raimondii</i> Receptor-like protein tree	54
Figure 11. <i>Gossypium hirsutum</i> Receptor-like protein tree	55
Figure 12. Distribution of all RLPs in phylogenetic tree	57
Figure 13. Phylogenetic tree and phenotypic observation of VIGSed plants	64

LIST OF TABLES

	Page
Table 1. The list of AtRLPs	15
Table 2. The number of RLPs in Cotton	20
Table 3. The list of GaRLPs	23
Table 4. The list of GrRLPs	30
Table 5. The list of GhRLPs	39
Table 6. Categorized LRRs in plants	46
Table 7. Used primers in this study	61
Table 8. Disease index of infected cottons	65

1. INTRODUCTION

Cotton (*Gossypium. Spp.*) is an agroeconomically significant crop worldwide involved in textile, oil, and feedstock production. (Chen J.Y., et al. 2015). The *Gossypium* genus contains 46 diploid species that can be categorized into eight groups comprising A-G, and K, and six allotetraploid species (Li, F., et al. 2015). There are commercially only four species, which are the allotetraploid of the *G. hirsutum*, and *G. barbadense* and the diploid of the *G. arboreum*, and *G. herbaceum* (Zhang H.B., et al 2008).

The diploid genomes (A- and D-) stems from the identical eudicot ancestor about 5-10 million years ago (MYA). It is believed that after the hybridization of the A- and D-genome cotton transpired five allotetraploid species polyploidized approximately 1-2 MYA (Li, F. et al. 2015). Evolutionary analysis has illustrated that *G. hirsutum* ($4n = 4x = 52$; AADD genome) is a hybrid of two diploid species, *G. arboreum* ($2n = 2x = 26$; AA genome) and *G. raimondii* (DD genome). The whole genome sequences of *G. hirsutum*, *G. arboreum*, and *G. raimondii* have been released for further studies (Zhang, H.B., et al. 2008; Li, F., et al. 2015; Paterson, A. H., et al. 2012).

Many pathogenic organisms including viruses, bacteria, fungi, and nematodes attack plants. Previous studies have shown that induced defense responses provide resistance capacity to these pathogenic attacks with plant disease resistance (R) proteins, in a gene-for-gene interaction. Our understanding of the molecular mechanism of gene-for-gene disease resistance has been improved due to identifying R genes and the corresponding Avirulence (Avr) genes (Kruijt, M., et al, 2005). In this case, induced

defense responses result from the recognition of Avr proteins by R genes. The *Avr* genes and induce defense responses are required to recognize products by producing plant disease resistance. (Xiao S, et. al. 2003).

Previous studies have been predominantly investigating non-peptide phytohormones, such as auxin, cytokinin, and gibberellin, and their back-and-forth interactions to clarify the principal signal transduction pathways and the different function of these hormones on plant growth, development, and reproduction in plants since 1950s. On the other hand, the recent studies have shown that many signaling events in plants are activated by small diverse secreted peptides as known elicitors. The common consensus is that the elicitors interact with extracellular membrane receptors to trigger activation of defense response and growth related genes (Uebler S., and Dresselhaus T., 2014).

One of the more crucial questions in plant biology is how plant cell interconnects with each other and recognizes signals from environmental cues. Many researchers in plant science declared that numerous surface receptors of cell containing extracellular leucine-rich repeats (eLRR) sense indications at the cellular level to respond to biotic or abiotic factors. Two classes of surface receptors families containing eLRRs in plants are receptor-like kinases (RLK) and receptor-like proteins (RLP), which are also known as pattern recognition receptors (PRR) (Zipfel C., 2014). General consensuses about the function of RLPs in plant species have been emphasized to play crucial roles in development stages and disease response (Tor M., et al, 2009).

Pathogen-associated molecular pattern (PAMP)-triggered immunity is activated through the recognition of microbe-derived patterns by PRRs (Albert I., et al. 2015). Well-studied PRRs include RLKs which have an intracellular kinase domain used to send the signal intracellularly, RLPs lack the kinase domain but the association of RLPs with SOBIR1-type adaptor kinases undergoes parallel functions in most of RLPs lacking cytoplasmic kinases domains for defense response. The association between SOBIR1 and RLPs occurs through leucine-rich repeat (LRR) domains by ionic interaction of their oppositely charged juxtamembrane domains and/or by helix–helix interactions of their transmembrane domains. Bimolecular equivalents of RLKs have similar function RLP/adaptor complexes by binding their ligands. The combination RLP/adaptor with ligands substitutes the formation of complex with co-receptor like BAK1/SERK1- type, which carries cytoplasmic domains. The process generates cytoplasmic signal output activation and phosphorylation (Gust A.A., and Felix G., 2014).

In many innate immunity related protein in organisms, the LRR containing domain is abundant. The eLRRs-RLP contains signal peptide (SP), N-terminus region (LRRNT) and C-terminus region (LRRCT) that include cysteine rich residues, LRR domains, island domain (IsD), linker and acidic domain (AD), transmembrane domain (TM), and cytoplasmic tail (CT). (Figure 1) (Jones D.A., Jones J.D.G., 1997).

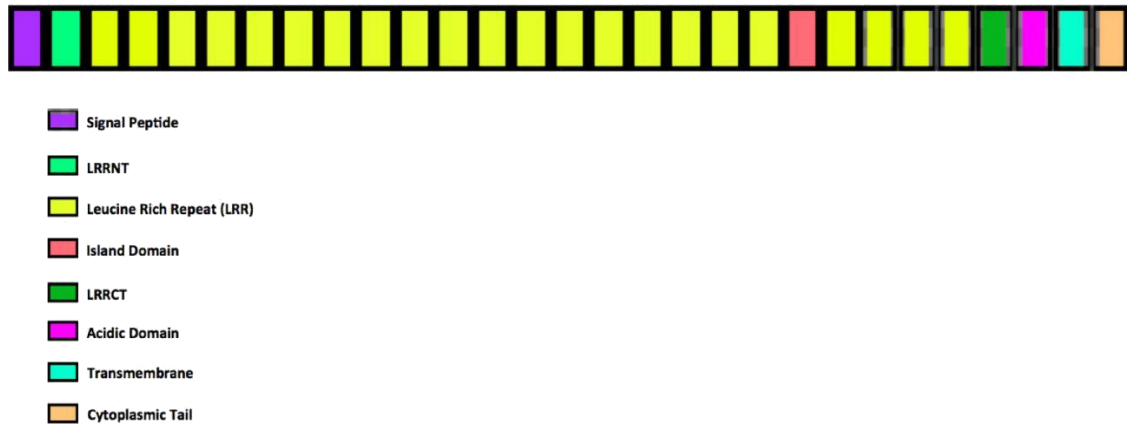


Figure 1. Structure of the various domains of typical RLPs.

Signal peptide in these two classes of proteins is responsible to target protein to plasma membrane. Jones D.A., Jones J.D.G. reported in 1997 that extracellular Leucine rich repeat are generated conserved motif plant specific $LxxLxxLxLxxNxLt/sgxIpxxLG$ and $LxxLxLxxN/CxL$. Extracellular Leucine-rich repeats provide to mediate protein-protein interaction (Wang G., et al, 2010). Even if not all RLPs include an island domain, non-LRR island domain intervenes eLRR subdomains. Acidic domain function has not been clarified until now, however the acidic regions are thought as transactivation domains in most of transcription factors (Schwechheimer C., 1998). Proline-rich acidic domain may helps to the secondary structure of protein in plant and animal cells. Possessing a $(G/S/T)XXX(G/S/T)$ motif in trans-membrane plays a critical role in intramolecular and intermolecular interactions between RLPs and other proteins such as RLK (Fritz-Laylin L.K., et al., 2005). Malectin, is isolated from *Xenopus laevis*, has a function as carbohydrate-binding domain such as lectins (Schallus T., et al., 2008). The Tyr-Xaa-Xaa- \emptyset ($YXX\emptyset$) motif, which may help endocytosis mechanism, is a

tyrosine-based sorting signal in cytoplasmic tail, which is found in the several RLPs but not all (Wang G., et al. 2010). All kingdoms include malectin domains. The binding of two tandem malectins generates malectin-like domain. Many studies reported that malectin-like receptor like kinase contributes to defense response (Chen C.W., et al. 2014; Hok S., et al 2011) whereas malectin-like receptor like protein may require for developmental stage. AtRLP4 is malectin type of receptor like protein in Arabidopsis (Fritz-Laylin L.K., et al., 2005).

Virus-induced gene silencing (VIGS) is commonly preferred to prevent time-consuming, and thank to its ease of handling to engineering phenotypes (Lu, R., et al 2003). The advantages of the method are that partial sequences of gene is enough to silence gene, no required for transformation, easy and rapid, useful for forward and reverse genetics, and targeting multiple copies of gene in polyploidy plant (Senthil-Kumar, M., and Mysore, K.S., 2012). VIGS provides a powerful reverse genetic approach for post-transcriptional gene silencing in plant species by utilizing the biological function of the RNA (ribonucleic acid) defense mechanism. The accrued foreign double-strain RNA (dsRNA) in plant cell is recognized for degradation by the RNA defense (Ruiz, M.T., et al. 1988). VIGS vectors pTRV1 and pTRV2 expressing in the agrobacterium encode target sequences derived from the plant genome. The combinations of the expressed foreign RNAs increase the accumulation of dsRNA. The accumulated dsRNA are transferred into a RNA-induced silencing complex (RISC) to cleave into short interfering (siRNA) duplexes of the length of 21-24 nucleotides by dicer-like protein in plant cell (Dominique R., 2004; Buhrow, L.M., et al. 2016).

Disease losses in cotton are important issues for yield around world. Plantation of infested fields with susceptible cultivars and disease management are highly limited for pathogens (Davis, R. M., et al. 2006). The soil-borne fungal pathogen *Fusarium oxysporum* is colonized in vascular tissues and stunting, wilting, chlorosis and necrosis of leaves, and death cells in a wide variety of crops. This devastating disease influences adversely a large spectrum of economically valuable crops all over the world (Catanzariti, A.M., et al. 2015). Darling Downs had first discovered Fusarium wilt in 1993. Fusarium wilt has long been an important disease.

The Arabidopsis genome and rice genome have 57 and 90 RLPs, respectively (Fritz-Laylin L.K., et al. 2005). Although 66 of *P.trichocarpa* RLP have been clustered up to seven members, *P.trichocarpa* has totally 82 RLPs in genome (Petre, B., et al. 2014). 144 RLPs, including two AtRLP4-like proteins have been identified through a genome-wide analysis of resistance gene analogues in cotton (Chen J.Y., et al. 2015). One of the first identified RLPs involved in plant defense was Cf-9, which confers resistance to *Cladosporium fulvum*. In addition, and Cf-2, Cf-4, Cf-4E, Cf-5, and Cf-9B receptor-like proteins provide resistance abilities to same the fungus by recognizing different or identical *avr* genes (Dixon M.S., et al. 1996; Joosten M.H., et al. 1997; Panter S.N., et al. 2002). Cure1 detects of the parasitic *Cuscuta* to initiate plant defense in tomato (Fürst U., et al. 2016). The ethylene-inducing xylanase 1 (*EIX1*) and ethylene-inducing xylanase 2 (*EIX2*) function as a PRR for the fungus elicitor xylanase (Ron M. and Awni A., 2004). *HcrVf2* (homologue of the *C. fulvum* resistance genes of the Vf region) encodes a receptor-like protein, and has a function against *Venturia inaequalis*

for preventing apple scab disease (Belfanti E., et al. 2004). The *L. maculans* resistance gene 3 (*LepR3*) postulates resistance to the fungus *Leptosphaeria maculans* upon the recognition of the elicitor AVR1M1 (Larkan, N. J., et al. 2013). RLP1 is a pattern recognition receptor for eMAX from *Xanthomonas* spp (Jehle A.K., et al. 2013). In *Arabidopsis*, RLP3 (*Fusarium oxysporum*), RLP23 (*Phytophthora infestans* and *Sclerotinia sclerotiorum*), RLP30 (*Scerotinia sclerotiorum* and bacteria), RLP42 (fungal endopolygalacturonases), RLP51 (resistance pathways), and RLP52 (*Erysiphe cichoracearum*) have resistance features. TaRLP1.1, considered to be induced by *Puccinia striiformis* f. sp. *Tritici* (Pst), was depicted as a significant stripe rust defense gene in wheat (Jiang Z., et al. 2013). The tomato immune receptor Ve1 and GbVe1 (*G. barbandense*) manage resistance to the fungus *Verticillium dahliae* (De Jonge R., et al. 2012; Zhang Z., et al. 2012). All of these examples highlight the importance of RLPs in plant defense response. In this thesis, I will use a combination of bioinformatics and gene functional analysis to identify and characterize cotton RLPs that may play an important role in defense against *Fusarium oxysporum*.

2. METHODS

2.1 Sequence Database Searches

The latest *G. arboreum* (A2_BGI), *G. raimondii* (D5_JGI), and *G. hirsutum* (AD1_NBI), protein sequences were downloaded from the cotton genome database (<https://www.cottongen.org>). The amino acid sequences of Arabidopsis RLPs were downloaded from the Arabidopsis Information Resource website (TAIR) (<http://www.arabidopsis.org>). To get the complete *G. arboretum*, *G. raimondii*, and *G. hirsutum* numbers of RLP proteins, the cotton whole genome sequences were queried with the Arabidopsis using the BLASTP program. All predicted cotton proteins were further screened for potential RLP protein using the HMMER server. Based on the HMMER's manual, the *E* value for the search of candidate protein was set to <0.1 (Finn R. D., et al. 2015).

2.2 Determination of Protein Domain Structure and pI Value

The InterProScan server was used to identify different protein domains (Hunter S. et al. 2009). Hidden Markov models (HMMs) from the PFAM suite (<http://pfam.wustl.edu>) were used to identify the presence of LRRs and other structural domains (Finn R.D., et al. 2016). Receptor-like protein motif search was conducted using SMART (Letunic I., et al. 2015). Isoelectric point of RLPs was calculated by using the isoelectric point calculator (IPC) server (Kozlowski L.P., 2016).

2.3 LRRs Domains Identification

The numbers of leucine-rich repeats were detected by using the LRRsearch webpage (A. Bej, 2014), LRRfinder software (Offord V., et al. 2010), and CDD: NCBI's conserved domain database (Marchler-Bauer A., et al. 2015). The following leucine rich repeats domain families were identified which are LRR_1 – leucine-rich repeat (PF00560), LRRNT – leucine-rich repeat N-terminal domain (PF01462), LRV - leucine-rich repeat variant (PF01816), LRRNT_2 - leucine-rich repeat N-terminal domain (PF08263), LRR_4 - leucine-rich repeat (2 copies) (PF12799), LRR_5 - leucine-rich repeat (6 copies) (PF13306), LRR_8 - leucine-rich repeat (PF13855), LRR_9 - leucine-rich repeat (PF14580), LRRCT - leucine-rich repeat C-terminal domain (PF01463), LRR_2 - leucine-rich repeat (PF07723), and LRR_3 - leucine-rich repeat (PF07725) (Magalhães D.M., et al, 2016).

2.4 Transmembrane and Signal Peptide Identification

The TMHMM website ([http:// www.cbs.dtu.dk/services/TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/)) hosted by the Center for Biological Sequence Analysis, Technical University of Denmark was used to predict transmembrane helices in proteins. The prediction of transmembrane helices in the protein sequences were conducted according to the default parameters of version 2.0 (Sonnhammer E.L., et al. 1998). The SignalP server was used to predict the existence and position of signal peptide cleavage sites in amino acid sequences of proteins (Petersen T.N., et al. 2011).

2.5 Orthologs Group in Cotton

The OrthoMCL applies the Markov Cluster (MCL) algorithm for clustering a set of protein sequences with complicated domain structures based on their sequence similarity. This database contained 116,536 ortholog groups clustered from 1,270,853 protein from 88 eukaryotic, 16 archaeal, and 34 bacterial genomes (<http://orthomcl.org/orthomcl/>). Retrieval proteins were performed by OrthoMCL webpage (Appendix A).

2.6 Multiple Sequence Alignment and Phylogenetic Analysis

In this study, several multiple sequence alignment approaches were performed to illuminate RLPs sequences. MUSCLE (Edgar, R.C., 2004) was used to align global homology group (<http://www.drive5.com/muscle>). MAFFT (Kato K., and Standley D.M., 2013) was used to align the protein sequences for phylogenetic analysis (<http://mafft.cbrc.jp/>). The corresponding protein sequences imported into MEGA6 for multiple sequence alignment by ClustalW and phylogenetic analysis by using the Unweighted Pair Group Method with Arithmetic (UPGMA). The weighing matrix used for ClustalW alignment BLOSUM with the penalty of gap opening 10 and gap extension 0.2. (Tamura K., et al. 2013). Trees were visualized by using the Interactive Tree of Life webpage.

2.7 Plant Materials and Growth

Cotton (*Gossypium hirsutum*) Fibermax (FM) 9160B2F cultivar seeds, which are provided by Bayer Cropsciences at Lubbock, TX, were germinated in pots containing

Metro Mix 900 in a growth room at room temperature for 10 days with a 12h photoperiod until two cotyledon emerged.

2.8 Construction of VIGS Vectors

The cDNA of GhCLA1, Gh_A07G1586 (Gh_RLP23_1) and Gh_D07G1774 (Gh_RLP23_1), Gh_D07G1953 (GhRLP23_2), Gh_A11G2283 (GhRLP23_3A), Gh_D09G2477 (GhRLP23_3D), Gh_D01G0395 (AtRLP34), Gh_D11G2968 (AtRLP27), Gh_D05G3192 (AtRLP7), Gh_A03G0773 (GSO1), Gh_D04G1293 (AtRLP6), Gh_A12G0722 (AtRLP44), and Gh_D13G2467 (AtRLP9) cDNA was amplified by polymerase chain reaction (PCR) from induced cDNA of *G. hirsutum* by several different pathogens (Appendix B). Target amplicons, pTRV-RNA1, pYL156 (pTRV-RNA2) were digested with restriction enzymes EcoRI and KpnI. The digested amplicons were ligated into pYL156 (pTRV-RNA2) vector at 16⁰C overnight. The constructs were transformed into *E. coli* (MC1061) component cells, and plated onto LB media containing kanamycin (50µg/ml) antibiotic at 37⁰C overnight. Plasmid DNA was isolated from the transformed colonies through alkaline lysis method. Sanger sequencing was used to validate the insertion (Used primers are given in Table 7).

2.9 Agrobacterium-Mediated VIGS

Plasmids that were sequenced and found to carry the target insert were transformed into *Agrobacterium tumefaciens* (GV3101) by electroporation. Colonies were inoculated into LB medium containing gentamicin (50µg/ml) and kanamycin (50µg/ml) at 28⁰C for overnight. The cells were collected by centrifugation and re-suspended in infiltration Buffer (10 mM MgCl₂, 10 mM MES and 200 uM

acetosyringone). The *Agrobacterium* culture containing pTRV-RNA1 and pTRV-RNA2 or pTRV-GhGene was mixed at 1:1 ratio (final OD of 0.75 for each one). Mixture cultures inoculated into cotton stems with a needle syringe. The VIGS experiment was performed with more than 15 plants used for each construct.

2.10 *F. Oxysporum* Inoculation and Disease Index

Fusarium oxysporum was grown in potato dextrose broth (PDB) medium at room temperature for 6-7 days. The culture was centrifuged at 4,000 rpm for 10 minutes to collect the pellet. *F. oxysporum* conidial suspensions (1×10^7 spores.mL⁻¹) with 0.001% tween were inoculated into cotton stems by using a needle syringe less than 1 cm below the cotyledons into VIGS plants. The disease symptom was divided into 5 levels (level 0-4) and calculated using the formula $DI = (\sum (n \times \text{number of seedlings at level } n)) / (4 \times \text{number of total seedlings} \times 100)$ (n refers the severity of the disease level of the plants). *Fusarium* inoculated plants were observed daily until 25 days post inoculation (dpi). Disease index indicates the infection ratio of a population.

3. RESULTS

3.1 Receptor-like Protein in Arabidopsis

3.1.1 Chromosomal Distribution of AtRLPs

57 receptor-like proteins were retrieved from Arabidopsis Genome database (www.arabidopsis.org) (Table 1). A chromosomal map tool was used to determine the distributions of AtRLPs genes on the *Arabidopsis* chromosomes with exception of AT1G54480 (AtRLP8/ RPP27) (Fig 2).

(<https://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp>).

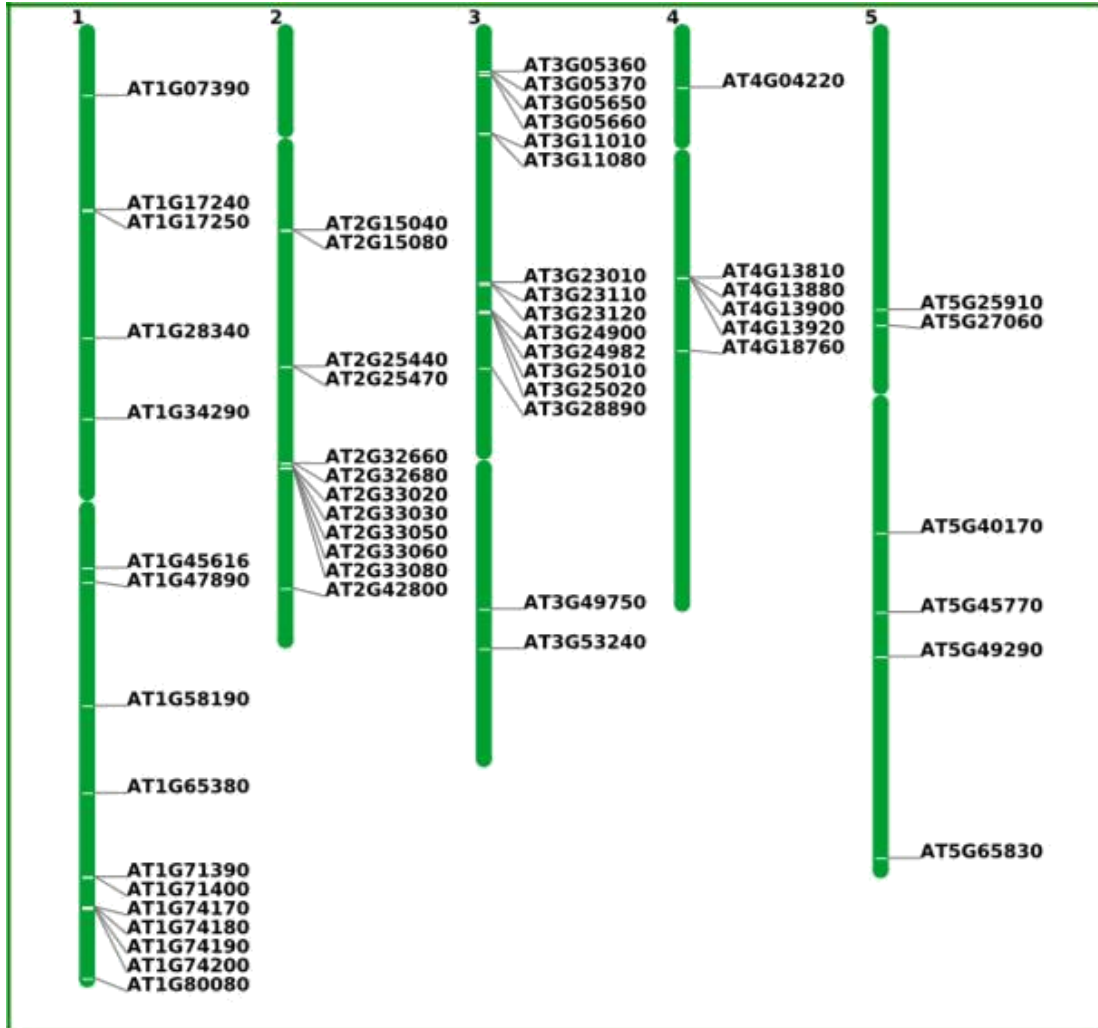


Figure 2. Representative *A. thaliana* RLPs chromosomal map

Table 1. The list of AtRLPs

ATRLPs Gene ID	AGI Code	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
AtRLP1	AT1G07390	Chr1: 2269893-2274820	4924	1083	121878.8904	5.998
AtRLP2	AT1G17240	Chr1: 5,896,416-5,898,717	2302	729	80576.48374	5.577
AtRLP3	AT1G17250	Chr1: 5,901,169-5,903,439	2271	756	83639.80914	6.471
AtRLP4	AT1G28340	Chr1: 9,940,175-9,943,549	3375	626	68418.80934	7.979
AtRLP5	AT1G34290	Chr1: 12,498,000-12,498,800	801	266	30069.19774	8.298
AtRLP6	AT1G45616	Chr1: 17,183,550-17,186,534	2985	994	110157.3123	6.431
AtRLP7	AT1G47890	Chr1: 17,643,976-17,647,035	3006	1019	113342.462	6.431
AtRLP8	AT1G54480	Chr1: 20,344,698-20,348,097	4495	1044	116826.3981	5.358
AtRLP9	AT1G58190	Chr1: 21,540,632-21,548,112	7481	1029	116271.0688	5.421
AtRLP10	AT1G65380	Chr1: 24,286,826-24,289,249	2424	720	79227.80044	5.571
AtRLP11	AT1G71390	Chr1: 26,906,453-26,908,807	2355	784	88731.05054	6.43
AtRLP12	AT1G71400	Chr1: 26,909,835-26,912,597	2763	847	95284.44994	6.042
AtRLP13	AT1G74170	Chr1: 27,891,494-27,896,355	4862	1000	112052.073	6.314
AtRLP14	AT1G74180	Chr1: 27,897,197-27,900,908	3712	976	108615.3817	4.83
AtRLP15	AT1G74190	Chr1: 27,902,590-27,906,158	3569	965	107564.8367	6.114
AtRLP16	AT1G74200	Chr1: 27,906,909-27,909,358	2450	302	33463.97074	8.588
AtRLP17	AT1G80080	Chr1: 30,128,005-30,129,591	1587	496	54505.23084	7.358
AtRLP18	AT2G15040	Chr2: 6,507,523-6,508,565	1043	1011	112336.6984	5.982

Table 1. Continued

ATRLPs Gene ID	AGI Code	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
AtRLP19	AT2G15080	Chr2: 6,533,687-6,536,995	3309	983	109407.3686	5.348
AtRLP20	AT2G25440	Chr2: 10,826,735-10,829,402	2668	671	75187.02434	7.749
AtRLP21	AT2G25470	Chr2: 10,838,420-10,841,881	3462	935	104510.5444	4.884
AtRLP22	AT2G32660	Chr2: 13,853,758-13,856,112	2355	589	65679.39754	4.788
AtRLP23	AT2G32680	Chr2: 13,859,769-13,862,614	2846	890	98482.30754	6.286
AtRLP24	AT2G33020	Chr2: 14,013,874-14,016,516	2643	864	95705.91494	6.395
AtRLP25	AT2G33030	Chr2: 14,017,684-14,018,340	657	218	23952.56164	8.492
AtRLP26	AT2G33050	Chr2: 14,021,870-14,024,398	2529	800	89116.97744	5.406
AtRLP27	AT2G33060	Chr2: 14,025,483-14,028,196	2714	808	90060.33224	6.174
AtRLP28	AT2G33080	Chr2: 14,032,015-14,034,237	2223	740	83137.35354	8.081
AtRLP29	AT2G42800	Chr2: 17,807,942-17,809,590	1649	462	50800.72504	9.182
AtRLP30	AT3G05360	Chr3: 1,530,735-1,533,288	2554	786	87527.28244	5.819
AtRLP31	AT3G05370	Chr3: 1,536,014-1,538,719	2706	860	96867.78514	6.209
AtRLP32	AT3G05650	Chr3: 1,645,680-1,648,573	2894	868	97302.65754	6.311
AtRLP33	AT3G05660	Chr3: 1,648,843-1,652,001	3159	875	97613.69314	7.677
AtRLP34	AT3G11010	Chr3: 3,450,496-3,453,683	3188	894	99013.69504	6.155
AtRLP35	AT3G11080	Chr3: 3,470,304-3,473,685	2873	943	104783.5479	5.907
AtRLP36	AT3G23010	Chr3: 8,174,341-8,176,911	1788	595	65998.82804	5.005

Table 1. Continued

ATRLPs Gene ID	AGI Code	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
AtRLP37	AT3G23110	Chr3: 8,222,364-8,224,871	2508	835	93218.72314	5.704
AtRLP38	AT3G23120	Chr3: 8,227,222-8,229,619	2355	784	87228.94724	6.24
AtRLP39	AT3G24900	Chr3: 9,099,060-9,101,991	2655	884	97449.19524	5.584
ATRLP40	AT3G24982	Chr3: 9,106,000-9,109,971	2748	915	101289.0034	6.001
AtRLP41	AT3G25010	Chr3: 9,110,057-9,112,936	2646	881	96918.01534	5.73
AtRLP42	AT3G25020	Chr3: 9,116,684-9,119,804	2673	890	98439.40374	5.802
AtRLP43	AT3G28890	Chr3: 10,895,977-10,898,900	2136	711	79793.67804	6.236
AtRLP44	AT3G49750	Chr3: 18,450,422-18,451,659	825	274	29993.48894	7.628
AtRLP45	AT3G53240	Chr3: 19,733,766-19,739,232	2676	891	100627.4401	6.397
AtRLP46	AT4G04220	Chr4: 2,033,168-2,035,946	2436	811	90838.07674	5.382
AtRLP47	AT4G13810	Chr4: 8,004,549-8,007,597	2226	741	82310.54414	4.702
AtRLP48	AT4G13880	Chr4: 8,025,804-8,028,614	2178	725	80973.10564	4.722
AtRLP49	AT4G13900	Chr4: 8,035,572-8,041,337	5494	751	83870.43174	4.908
AtRLP50	AT4G13920	Chr4: 8,043,803-8,046,559	2757	891	98334.15034	4.877
AtRLP51	AT4G18760	Chr4: 10,307,810-10,309,591	1296	431	46070.95744	8.218
AtRLP52	AT5G25910	Chr5: 9,038,860-9,041,377	2518	811	90356.51744	5.151
AtRLP53	AT5G27060	Chr5: 9,521,978-9,525,478	2874	957	106675.0052	6.047
AtRLP54	AT5G40170	Chr5: 16,064,859-16,067,628	2379	792	88797.47914	6.874

Table 1. Continued

ATRLPs Gene ID	AGI Code	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
AtRLP54	AT5G40170	Chr5: 16,064,859-16,067,628	2379	792	88797.47914	6.874
AtRLP55	AT5G45770	Chr5: 18,563,333-18,564,921	1278	425	47174.19514	9.278
ATRLP56	AT5G49290	Chr5: 19,979,559-19,984,131	2727	908	102450.1104	4.949
ATRLP57	AT5G65830	Chr5: 26,341,963-26,343,546	840	279	30878.60084	8.114

The receptor-like proteins (RLPs) in *Arabidopsis* are distributed across the five chromosomes. Most of the RLPs are located on chromosome 1 in *Arabidopsis* genome. Chromosome 1 includes seventeen RLPs from AtRLP1 to AtRLP17; chromosome 2 includes twelve RLPs from AtRLP18 to AtRLP29; chromosome 3 includes sixteen RLPs from AtRLP30 to AtRLP45; chromosome 4 includes six RLPs from AtRLP46 to AtRLP51; chromosome 5 includes six RLPs from AtRLP52 to AtRLP57 (Fig2; Table1). AtRLP25 is the shortest of all RLPs whereas AtRLP1 is longest. The pI value of RLPs varies from 4.702 to 9.278. AtRLP47 has the lowest pI value whereas AtRLP55 has the highest.

3.2 Receptor-Like Protein in Cotton

3.2.1 Identification of RLPs in Cotton

The amino acid sequence of all *Arabidopsis* LRR-RLPs provided from the *Arabidopsis* genome database were used as queries to perform a BLASTp analysis against cotton the protein database to identify LRR-RLPs in *G. arboreum*, *G. raimondii*, and *G. hirsutum*. Retrieved protein sequences were uploaded to SMART, PFAM, STRING, and InterProScan, CDD from NCBI to annotate and validate the domain structure of proteins. A candidate LRR-RLP had to at least possess one LRR domain and a transmembrane domain. After eliminating of the unmatched sequences manually by using multiple sequence alignments with known receptor-like protein in the *Arabidopsis* database, 86 LRR-RLPs in *Gossypium arboreum*, 107 LRR-RLPs in *Gossypium raimondii*, and 151 LRR-RLPs in *G. hirsutum* were identified (Table 2).

Table 2. The number of RLPs in Cotton

Species	Receptor-like proteins	Malectin-like RLPs	Total
<i>A. Thaliana</i>	56	1	57
<i>G. Arboreum</i>	85	1	86
<i>G. Raimondii</i>	105	2	107
<i>G. Hirsutum</i>	145	6	151

The peptide length of retrieved LRR-RLPs ranged from 218-1083 amino acids in *A. thaliana*, 181-1125 amino acids in *G. arboreum*, 191-1176 amino acids in *G. raimondii*, and 193-1981 amino acids in *G. hirsutum*. The chromosome location, and the characteristics such as nucleic acid length, amino acid sequences length, protein weight, isoelectronic point, and the number of LRRs for each RLP was listed as detailed information. The number of GaLRR-RLPs, GrLRR-RLPs, and GhLRR-RLPs are approximately 1.50, 1.88, and 2.63 fold higher than the number of AtLRR-RLPs, respectively.

3.2.2 GaRLPs Biochemical Features and Chromosomal Distribution

The RLPs in *Gossypium arboreum* are distributed across the thirteen chromosomes. Most of RLPs are located on chromosome 12 in the genome (Fig 3).

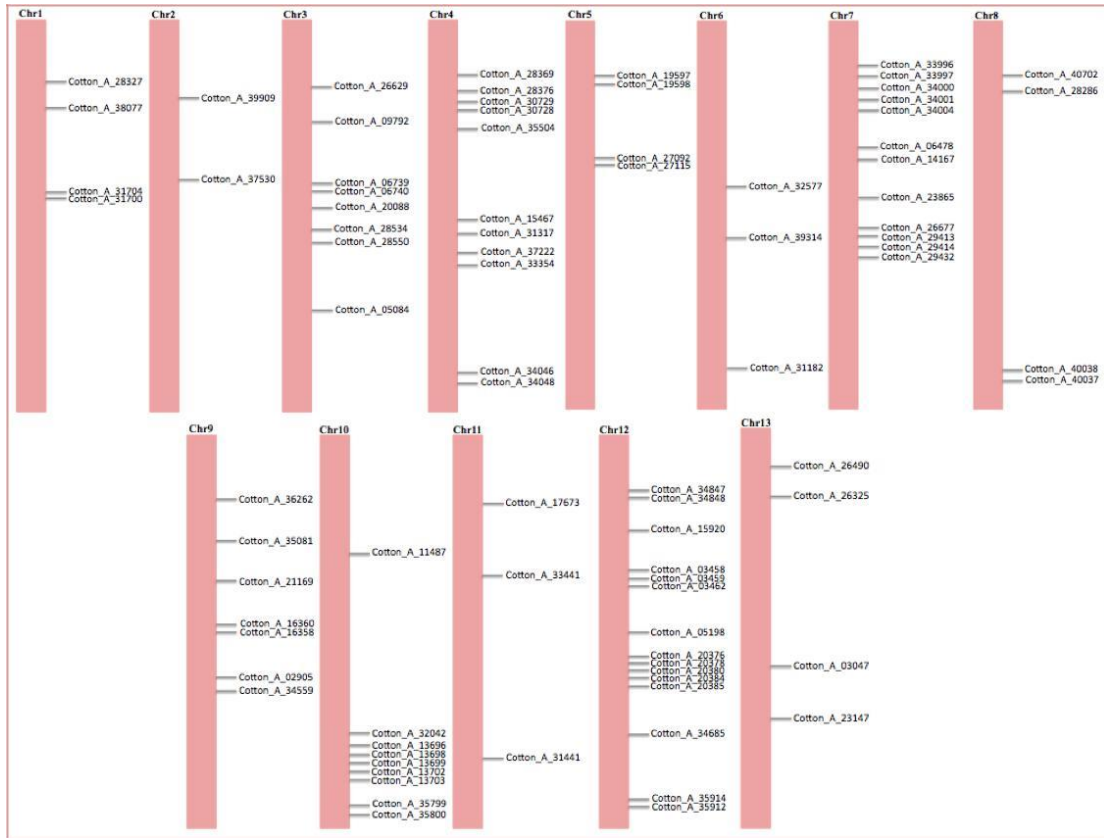


Figure 3. Representative *G. arboreum* RLPs chromosomal map

Chromosome 1 includes four RLPs which are Cotton_A_28327, Cotton_A_38077, Cotton_A_31704, and Cotton_A_31700; chromosome 2 includes two RLPs which are Cotton_A_39909 and Cotton_A_37530; chromosome 3 includes eight RLPs which are Cotton_A_09792, Cotton_A_06739, Cotton_A_06740, Cotton_A_20088, Cotton_A_28534, Cotton_A_28550, Cotton_A_26629, and Cotton_A_05084; chromosome 4 includes eleven Cotton_A_28369, Cotton_A_28376, Cotton_A_34046, Cotton_A_34048, Cotton_A_30729, Cotton_A_30728, Cotton_A_35504, Cotton_A_15467, Cotton_A_31317, Cotton_A_37222, and Cotton_A_33354;

chromosome 5 includes four RLPs which are Cotton_A_19597, Cotton_A_19598, Cotton_A_27092, and Cotton_A_27115; chromosome 6 includes three RLPs which are Cotton_A_32577, Cotton_A_39314, and Cotton_A_31182; chromosome 7 includes twelve RLPs which are Cotton_A_33996, Cotton_A_33997, Cotton_A_34000, Cotton_A_34001, Cotton_A_34004, Cotton_A_06478, Cotton_A_14167, Cotton_A_23865, Cotton_A_26677, Cotton_A_29413, Cotton_A_29414, and Cotton_A_29432; chromosome 8 includes four RLPs which are Cotton_A_40702, Cotton_A_28286, Cotton_A_40038, and Cotton_A_40037; chromosome 9 includes seven RLPs which are Cotton_A_36262, Cotton_A_35081, Cotton_A_21169, Cotton_A_16360, Cotton_A_16358, Cotton_A_02905, Cotton_A_34559; chromosome 10 includes nine RLPs which are Cotton_A_32042, Cotton_A_13696, Cotton_A_13698, Cotton_A_13699, Cotton_A_13702, Cotton_A_13703, Cotton_A_35799, Cotton_A_35800, and Cotton_A_11487; chromosome 11 includes three RLPs which are Cotton_A_31441, Cotton_A_17673, and Cotton_A_33441; chromosome 12 includes fifteen RLPs which are Cotton_A_34847, Cotton_A_34848, Cotton_A_15920, Cotton_A_03458, Cotton_A_03459, Cotton_A_03462, Cotton_A_05198, Cotton_A_20376, Cotton_A_20378, Cotton_A_20380, Cotton_A_20384, Cotton_A_20385, Cotton_A_34685, Cotton_A_35914, and Cotton_A_35912; chromosome 13 includes four RLPs which are Cotton_A_26490, Cotton_A_26325, Cotton_A_03047, and Cotton_A_23147 (Table 3).

Table 3. List of GaRLPs

GaRLPs Gene ID	Chromosome location	Sequences Length	A.a Length	Protein Mass (Da)	pI
Cotton_A_28327	Chr1: 30910956-30917800+	1521	506	55198.21184	6.484
Cotton_A_38077	Chr1: 88368385-88371471-	3087	1028	114992.3887	5.96
Cotton_A_31704	Chr1: 115608089-115610662-	2574	857	95869.46624	7.802
Cotton_A_31700	Chr1: 115655090-115655707-	618	205	22810.07244	8.533
Cotton_A_39909	Chr2: 52128929-52132057+	3129	1042	112682.2216	5.118
Cotton_A_37530	Chr2: 77372098-77373864-	1767	588	65089.83314	7.884
Cotton_A_09792	Chr3: 36831342-36832151+	810	269	28786.06764	4.932
Cotton_A_06739	Chr3: 52138460-52141729-	3270	1089	121239.3429	5.204
Cotton_A_06740	Chr3: 52149648-52152689-	3042	1013	111931.4686	5.228
Cotton_A_20088	Chr3: 53912842-53915841-	3000	999	110940.4208	5.146
Cotton_A_28534	Chr3: 55419939-55426426+	3369	1123	126848.6655	5.379
Cotton_A_28550	Chr3: 55831467-55837245+	3204	1067	119871.514	4.772
Cotton_A_26629	Chr3: 577031-578283+	861	286	31308.86314	5.348
Cotton_A_05084	Chr3: 89627000-89629531+	2532	843	92905.94354	5.911
Cotton_A_28369	Chr4: 10929663-10932707+	3045	1014	112974.5597	5.307
Cotton_A_28376	Chr4: 11240008-11243808+	2847	948	106057.4321	6.993
Cotton_A_34046	Chr4: 124860432-124861157+	726	241	27291.36354	5.998
Cotton_A_34048	Chr4: 125027226-125027873+	648	215	23888.33304	4.708
Cotton_A_30729	Chr4: 12988649-12989695-	1047	348	38032.65464	5.608
Cotton_A_30728	Chr4: 13010739-13013657-	2919	972	106572.8155	5.177
Cotton_A_35504	Chr4: 131499321-131500460+	1140	379	43095.17314	5.084
Cotton_A_15467	Chr4: 57214230-57215936-	1707	568	63317.59314	7.405
Cotton_A_31317	Chr4: 61505598-61506404-	807	268	28916.25284	5.446

Table 3. Continued

GaRLPs Gene ID	Chromosome location	Sequences Length	A.a Length	Protein Mass (Da)	pI
Cotton_A_37222	Chr4: 76882245-76885218-	2772	923	103322.4509	7.33
Cotton_A_33354	Chr4: 79351701-79353221-	1521	506	57209.18634	6.634
Cotton_A_19597	Chr5: 15983701-15986723-	2739	912	100476.9533	6.183
Cotton_A_19598	Chr5: 16020886-16022301-	1416	471	51801.31124	5.227
Cotton_A_27092	Chr5: 58366994-58374758+	3357	1118	125098.456	5.079
Cotton_A_27115	Chr5: 58601656-58604178-	2523	840	94393.65434	7.605
Cotton_A_32577	Chr6: 60225129-60225932+	804	267	28578.83304	5.186
Cotton_A_39314	Chr6: 79318770-79319759-	990	329	37298.61554	4.929
Cotton_A_31182	Chr6: 118743544-118745589-	2046	681	75758.23644	5.584
Cotton_A_33996	Chr7: 11647947-11648693-	747	248	27557.69154	5.145
Cotton_A_33997	Chr7: 11738099-11740773-	783	260	29795.21264	6.486
Cotton_A_34000	Chr7: 11919099-11922101-	3003	1000	111250.487	7.45
Cotton_A_34001	Chr7: 11953648-11958925-	1578	525	58601.50464	7.343
Cotton_A_34004	Chr7: 12058922-12061971-	2079	692	76644.38754	5.363
Cotton_A_06478	Chr7: 29642331-29644376+	2046	681	75758.23644	5.584
Cotton_A_14167	Chr7: 30328783-30331908-	3126	1041	116158.847	4.901
Cotton_A_23865	Chr7: 72473316-72476378-	3063	1020	113084.4431	5.659
Cotton_A_26677	Chr7: 92359605-92365255-	2904	967	108585.283	6.134
Cotton_A_29413	Chr7: 92798578-92806654-	1719	572	63422.47554	6.855
Cotton_A_29414	Chr7: 92864801-92872699-	1569	522	58403.51564	6.986
Cotton_A_29432	Chr7: 93097471-93098205-	735	244	26847.08664	7.467
Cotton_A_40702	Chr8: 42129701-42131452-	1752	583	64885.47144	5.09
Cotton_A_28286	Chr8: 55470994-55478516+	2463	820	91557.42464	4.86

Table 3. Continued

GaRLPs Gene ID	Chromosome location	Sequences Length	A.a Length	Protein Mass (Da)	pI
Cotton_A_40038	Chr8: 107225200-107226936-	1737	578	64522.21274	5.836
Cotton_A_40037	Chr8: 107254897-107256654-	1758	585	65518.00574	5.801
Cotton_A_36262	Chr9: 11100425-11103613+	3189	1062	118890.7195	6.376
Cotton_A_35081	Chr9: 28753594-28756947+	3192	1063	118878.7848	6.153
Cotton_A_21169	Chr9: 40754366-40757329+	2964	987	110279.5149	6.818
Cotton_A_16360	Chr9: 52646496-52650150+	1701	566	62965.77794	4.814
Cotton_A_16358	Chr9: 52736114-52740457+	3378	1125	126920.2344	4.945
Cotton_A_02905	Chr9: 73557324-73558727+	1404	467	50652.44934	5.024
Cotton_A_34559	Chr9: 77565066-77567531+	2334	777	87142.42344	5.052
Cotton_A_32042	Chr10: 108889804-108891273+	1470	489	53441.25964	7.691
Cotton_A_13696	Chr10: 111285691-111288582+	2892	963	108400.9774	7.792
Cotton_A_13698	Chr10: 111310855-111313464+	2610	869	97266.30714	6.425
Cotton_A_13699	Chr10: 111328290-111330524+	1506	501	56538.87464	7.342
Cotton_A_13702	Chr10: 111376232-111376927	696	231	25473.50824	6.196
Cotton_A_13703	Chr10: 111445871-111447790+	1443	480	53592.96244	8.042
Cotton_A_35799	Chr10: 24409574-24410140+	567	188	20228.99054	4.871
Cotton_A_35800	Chr10: 24418955-24419578+	624	207	22764.24024	10.214
Cotton_A_11487	Chr10: 69543993-69550724-	3372	1038	116719.3068	6.884
Cotton_A_31441	Chr11: 101355457-101358412-	2952	984	109441.2415	6.811
Cotton_A_17673	Chr11: 20706754-20710932+	1773	590	64413.08234	8.213
Cotton_A_33441	Chr11: 54390747-54391421-	675	224	25408.04444	5.666
Cotton_A_34847	Chr12: 6100472-6102592-	2037	678	75704.42944	5.98
Cotton_A_34848	Chr12: 6104027-6106181+	1869	622	70627.65584	5.113

Table 3. Continued

GaRLPs Gene ID	Chromosome location	Sequences Length	A.a Length	Protein Mass (Da)	pI
Cotton_A_15920	Chr12: 9326284-9333257-	3357	1119	127228.6078	6.132
Cotton_A_03458	Chr12: 27081331-27084723-	3393	1130	125550.4625	6.495
Cotton_A_03459	Chr12: 27098886-27106089-	1626	541	60219.80384	8.346
Cotton_A_03462	Chr12: 27116827-27120147-	3066	1021	114476.5207	8.197
Cotton_A_05198	Chr12: 47903593-47905527-	1935	644	71934.33444	5.782
Cotton_A_20376	Chr12: 52346781-52348693+	1707	568	63288.78324	6.795
Cotton_A_20378	Chr12: 52393164-52395953+	2790	929	103467.6307	6.239
Cotton_A_20380	Chr12: 52434763-52435524+	762	253	28887.49084	8.853
Cotton_A_20384	Chr12: 52533070-52535700+	2631	876	97520.66754	6.014
Cotton_A_20385	Chr12: 52584501-52587128+	2559	831	93048.35354	5.579
Cotton_A_34685	Chr12: 61587792-61588796-	819	272	30669.15824	5.873
Cotton_A_35914	Chr12: 108161798-108162446+	546	181	20147.88244	4.311
Cotton_A_35912	Chr12: 108440977-108441624+	648	215	24074.54144	4.91
Cotton_A_26490	Chr13: 2273929-2276762+	1527	508	57149.05284	5.088
Cotton_A_26325	Chr13: 33532391-33533737+	1347	448	48292.03404	5.478
Cotton_A_03047	Chr13: 82011906-82013273+	1368	455	50737.16964	5.375
Cotton_A_23147	Chr13: 94465645-94468604-	2892	963	106749.779	5.877

3.2.3 GrRLPs Biochemical Features and Chromosomal Distribution

The RLPs in *Gossypium raimondii* are distributed across the thirteen chromosomes. Most of the RLPs are located on chromosome 2, chromosome 5, chromosome 7, and chromosome 9 in *Gossypium raimondii* genome (19, 15, 18, 15, respectively) (Fig 4).

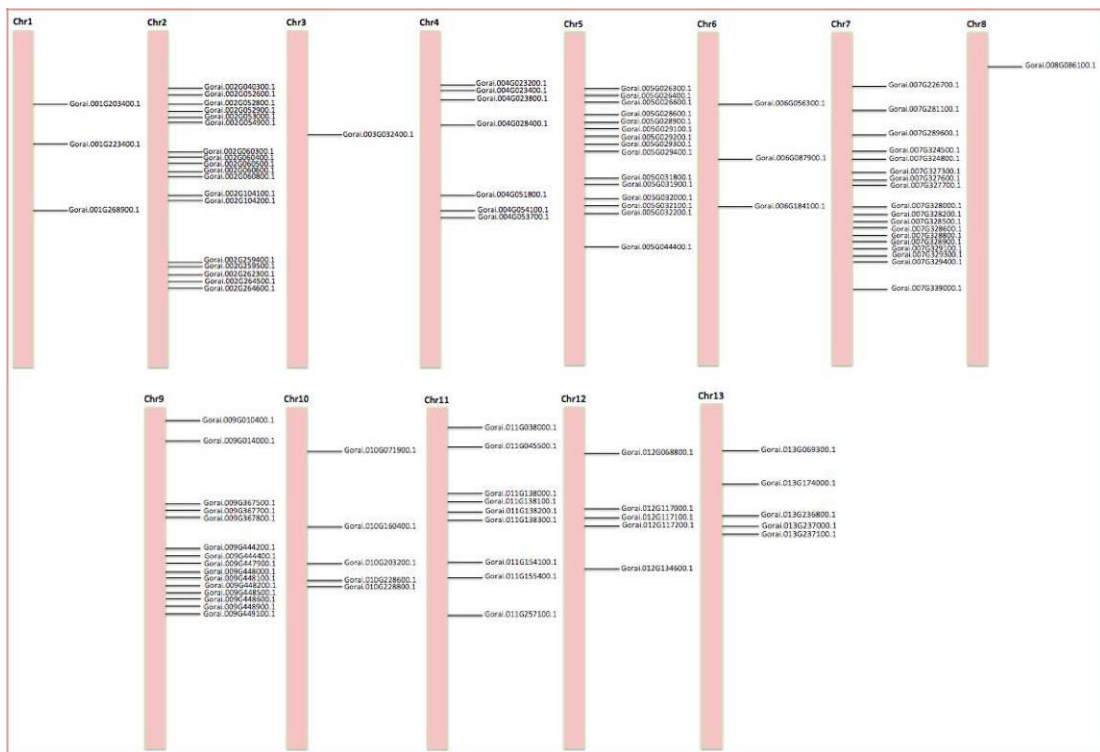


Figure 4. Representative *G. raimondii* RLPs chromosomal map

Chromosome 1 includes four RLPs which are Gorai.001G203400.1, Gorai.001G223400.1, and Gorai.001G268900.1; chromosome 2 includes nineteen RLPs which are Gorai.002G039000.1, Gorai.002G040300.1, Gorai.002G052600.1, Gorai.002G052800.1, Gorai.002G052900.1, Gorai.002G053000.1,

Gorai.002G054900.1, Gorai.002G060300.1, Gorai.002G060400.1,
 Gorai.002G060500.1, Gorai.002G060600.1, Gorai.002G060800.1,
 Gorai.002G104100.1, Gorai.002G104200.1, Gorai.002G259400.1,
 Gorai.002G259500.1, Gorai.002G262300.1, Gorai.002G264500.1, and
 Gorai.002G264600.1; chromosome 3 includes one RLP which is Gorai.003G032400.1;
 chromosome 4 includes seven RLPs which are Gorai.004G023200.1,
 Gorai.004G023400.1, Gorai.004G023800.1, Gorai.004G028400.1,
 Gorai.004G051800.1, Gorai.004G053700.1, and Gorai.004G054100.1; chromosome 5
 includes fifteen RLPs which are Gorai.005G026300.1, Gorai.005G026400.1,
 Gorai.005G026600.1, Gorai.005G028600.1, Gorai.005G028900.1,
 Gorai.005G029100.1, Gorai.005G029200.1, Gorai.005G029300.1,
 Gorai.005G029400.1, Gorai.005G031800.1, Gorai.005G031900.1,
 Gorai.005G032000.1, Gorai.005G032100.1, Gorai.005G032200.1, and
 Gorai.005G044400.1; chromosome 6 includes three RLPs which are
 Gorai.006G056300.1, Gorai.006G087900.1, Gorai.006G184100.1; chromosome 7
 includes eighteen RLPs which are Gorai.007G226700.1, Gorai.007G328800.1,
 Gorai.007G281100.1, Gorai.007G289600.1, Gorai.007G324500.1,
 Gorai.007G324800.1, Gorai.007G327300.1, Gorai.007G327600.1,
 Gorai.007G327700.1, Gorai.007G328000.1, Gorai.007G328200.1,
 Gorai.007G328500.1, Gorai.007G328600.1, Gorai.007G328900.1,
 Gorai.007G329100.1, Gorai.007G329300.1, Gorai.007G329400.1, and
 Gorai.007G339000.1; chromosome 8 includes one RLP which is Gorai.008G086100.1;

chromosome 9 includes fifteen RLPs which are Gorai.009G010400.1, Gorai.009G014000.1, Gorai.009G367500.1, Gorai.009G367700.1, Gorai.009G367800.1, Gorai.009G444200.1, Gorai.009G444400.1, Gorai.009G447900.1, Gorai.009G448000.1, Gorai.009G448100.1, Gorai.009G448200.1, Gorai.009G448500.1, Gorai.009G448600.1, Gorai.009G448900.1, and Gorai.009G449100.1; chromosome 10 includes five RLPs which are Gorai.010G071900.1, Gorai.010G160400.1, Gorai.010G203200.1, Gorai.010G228600.1, and Gorai.010G228800.1; chromosome 11 includes ten RLPs which are Gorai.011G038000.1, Gorai.011G138000.1, Gorai.011G138100.1, Gorai.011G138200.1, Gorai.011G138300.1, Gorai.011G029600.1, Gorai.011G257100.1, Gorai.011G045500.1, Gorai.011G154100.1, and Gorai.011G155400.1; chromosome 12 includes five RLPs which are Gorai.012G117000.1, Gorai.012G117200.1, Gorai.012G068800.1, Gorai.012G117100.1, and Gorai.012G134600.1; chromosome 13 includes five RLPs which are Gorai.013G069300.1, Gorai.013G174000.1, Gorai.013G236800.1, Gorai.013G237000.1, and Gorai.013G237100.1 (Table 4).

Table 4. The list of GrRLPs

GrRLPs Gene ID	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gorai.001G203400.1	Chr01: 39550032-39553170-	3139	1027	115072.7203	6.428
Gorai.001G223400.1	Chr01: 45615405-45618356-	2952	816	90473.09024	7.365
Gorai.001G268900.1	Chr01: 54578272-54580866-	2595	478	53123.76874	7.183
Gorai.002G039000.1	Chr02: 3158504-3160277-	1774	443	49665.61994	7.459
Gorai.002G040300.1	Chr02: 3317155-3325102+	7948	764	86034.67564	6.58
Gorai.002G052600.1	Chr02: 4660875-4663554+	2680	801	89378.19834	5.376
Gorai.002G052800.1	Chr02: 4682517-4685639+	3123	968	107400.9292	4.825
Gorai.002G052900.1	Chr02: 4700325-4704029+	3705	924	103530.6261	5.882
Gorai.002G053000.1	Chr02: 4707562-4709667+	2106	191	20920.15654	5.217
Gorai.002G054900.1	Chr02: 4866426-4867853+	1428	475	53518.65944	5.598
Gorai.002G060300.1	Chr02: 6999595-7009460+	9866	895	99868.45854	6.746
Gorai.002G060400.1	Chr02: 7012644-7021558+	8915	1099	121445.7594	5.765
Gorai.002G060500.1	Chr02: 7024478-7030794+	6317	1066	118327.4753	5.698
Gorai.002G060600.1	Chr02: 7044216-7047195+	2980	786	87788.74274	7.704
Gorai.002G060800.1	Chr02: 7172911-7182017+	9107	839	93636.86054	7.582
Gorai.002G104100.1	Chr02: 13336173-13344166-	7994	559	62135.92994	6.27
Gorai.002G104200.1	Chr02: 13374577-13377738-	3162	950	104815.8164	7.024
Gorai.002G259400.1	Chr02: 62090043-62093267+	3225	1067	118795.0976	6.312
Gorai.002G259500.1	Chr02: 62097052-62100287+	3236	1028	115306.9219	7.925
Gorai.002G262300.1	Chr02: 62280505-62284265-	3761	1081	119061.075	6.78
Gorai.002G264500.1	Chr02: 62467073-62470246+	3174	1057	117543.5104	7.025
Gorai.002G264600.1	Chr02: 62473969-62476021+	2053	683	76277.66504	7.59
Gorai.003G032400.1	Chr03: 3131359-3134452+	3094	268	28645.88274	4.932

Table 4. Continued

GrRLPs Gene ID	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gorai.004G023200.1	Chr04: 1763324-1768158-	4835	980	109121.5224	5.256
Gorai.004G023400.1	Chr04: 1793681-1797005-	3325	1037	115984.4989	6.314
Gorai.004G023800.1	Chr04: 1822103-1825081-	2979	992	109832.6647	5.413
Gorai.004G028400.1	Chr04: 2238391-2241294-	2904	795	87780.00084	5.524
Gorai.004G051800.1	Chr04: 4750737-4755598+	4862	477	53989.17384	7.198
Gorai.004G053700.1	Chr04: 5021964-5028018+	6055	1176	131498.832	4.914
Gorai.004G054100.1	Chr04: 5147959-5152618+	4660	878	98318.46734	5.199
Gorai.005G026300.1	Chr05: 2297846-2301061+	3216	1044	117610.3524	8.025
Gorai.005G026400.1	Chr05: 2313103-2316975+	3873	958	108604.9104	7.685
Gorai.005G026600.1	Chr05: 2349170-2352262+	3093	981	111016.6404	7.637
Gorai.005G028600.1	Chr05: 2501120-2504324+	3205	948	107378.5966	7.364
Gorai.005G028900.1	Chr05: 2553716-2557308+	3593	884	100704.5298	7.817
Gorai.005G029100.1	Chr05: 2585224-2589560+	4337	856	97034.80504	8.149
Gorai.005G029200.1	Chr05: 2596425-2600231+	3807	913	103816.9082	8.191
Gorai.005G029300.1	Chr05: 2613108-2617044+	3937	981	111035.5915	7.569
Gorai.005G029400.1	Chr05: 2632315-2635858+	3544	1044	117913.8224	6.798
Gorai.005G031800.1	Chr05: 2833071-2838460-	5390	1086	121993.8257	5.394
Gorai.005G031900.1	Chr05: 2852829-2859789-	6961	761	85763.00764	6.532
Gorai.005G032000.1	Chr05: 2870255-2876978-	6724	1056	118852.0706	5.9
Gorai.005G032100.1	Chr05: 2880034-2881241-	1208	366	40886.46684	5.053
Gorai.005G032200.1	Chr05: 2898234-2905491-	7258	974	108250.5009	5.56
Gorai.005G044400.1	Chr05: 4225606-4229843+	4238	1156	129818.1126	7.397
Gorai.006G056300.1	Chr06: 20515068-20521162-	6095	968	107633.1413	6.963

Table 4. Continued

GrRLPs Gene ID	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gorai.006G184100.1	Chr06: 44144569-44148609+	4041	1128	124913.2793	5.836
Gorai.007G226700.1	Chr07: 27315839-27317522+	1684	268	28929.38184	6.281
Gorai.007G328800.1	Chr07: 54926909-54930142-	3234	418	46303.17604	5.101
Gorai.007G281100.1	Chr07: 48214073-48219415+	5343	972	108873.963	7.524
Gorai.007G289600.1	Chr07: 49603046-49605335+	2290	627	69287.40524	4.784
Gorai.007G324500.1	Chr07: 54360255-54369269-	9015	495	56088.84164	7.798
Gorai.007G324800.1	Chr07: 54421943-54423160-	1218	382	42764.55094	5.161
Gorai.007G327300.1	Chr07: 54649996-54651585-	1590	357	40106.76204	5.588
Gorai.007G327600.1	Chr07: 54680184-54681428-	1245	414	46097.90444	5.739
Gorai.007G327700.1	Chr07: 54686251-54699995-	13745	837	92566.69674	5.934
Gorai.007G328000.1	Chr07: 54772423-54773544-	1122	373	41751.67324	5.226
Gorai.007G328200.1	Chr07: 54799120-54800259-	1140	379	42412.48054	5.266
Gorai.007G328500.1	Chr07: 54835735-54836963-	1229	382	42685.03154	5.116
Gorai.007G328600.1	Chr07: 54867351-54868547-	1197	398	44538.97024	5.252
Gorai.007G328900.1	Chr07: 54956716-54957954-	1239	412	45820.53084	5.229
Gorai.007G329100.1	Chr07: 55040507-55041751-	1245	390	43691.01834	6.149
Gorai.007G329300.1	Chr07: 55109059-55112715-	3657	336	37691.98634	4.927
Gorai.007G329400.1	Chr07: 55126400-55127560-	1161	370	41845.95644	7.594
Gorai.007G339000.1	Chr07: 56426052-56429093-	3042	651	72541.96074	5.035
Gorai.008G086100.1	Chr08: 18788697-18790732+	2036	267	28644.01214	5.539
Gorai.009G010400.1	Chr09: 838538-843236-	4699	799	88815.07434	5.269
Gorai.009G014000.1	Chr09: 1130692-1136363+	5672	968	109167.0953	6.093
Gorai.009G367500.1	Chr09: 49240580-49252448+	11869	926	102130.1512	5.854

Table 4. Continued

GrRLPs Gene ID	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gorai.009G367700.1	Chr09: 49273665-49276550+	2886	961	106701.5044	6.561
Gorai.009G367800.1	Chr09: 49294772-49297711+	2940	941	104768.2571	6.469
Gorai.009G444200.1	Chr09: 69537484-69540288+	2805	834	92873.80584	8.142
Gorai.009G444400.1	Chr09: 69551723-69554836+	3114	993	109839.3388	6.241
Gorai.009G447900.1	Chr09: 69897022-69900057-	3036	982	108904.4139	7.214
Gorai.009G448000.1	Chr09: 69905283-69908477-	3195	1064	118038.9649	6.56
Gorai.009G448100.1	Chr09: 69914318-69922836-	8519	1064	118455.6945	7.808
Gorai.009G448200.1	Chr09: 69924245-69927361-	3117	1038	114714.2295	6.452
Gorai.009G448500.1	Chr09: 69949953-69953044-	3092	1016	112729.4733	6.405
Gorai.009G448600.1	Chr09: 69959437-69972698-	13262	1060	117531.6708	5.669
Gorai.009G448900.1	Chr09: 69986285-69989718-	3434	1105	122829.9057	7.313
Gorai.009G449100.1	Chr09: 70036505-70039441-	2937	978	108933.9358	6.998
Gorai.010G071900.1	Chr10: 10092979-10094563+	1585	474	53576.03274	7.88
Gorai.010G160400.1	Chr10: 45634625-45635920-	1296	420	45457.37794	8.121
Gorai.010G203200.1	Chr10: 56860432-56863748+	3317	1026	113688.377	4.942
Gorai.010G228600.1	Chr10: 60029120-60032457-	3338	834	92214.47424	5.737
Gorai.010G228800.1	Chr10: 60060775-60064375-	3601	844	94583.21224	5.67
Gorai.011G038000.1	Chr11: 2856373-2857680+	1308	435	46884.44624	6.953
Gorai.011G138000.1	Chr11: 21177035-21181167+	4133	1055	118643.1321	6.654
Gorai.011G138100.1	Chr11: 21197670-21201792+	4123	1104	124698.0646	5.962
Gorai.011G138200.1	Chr11: 21267637-21271938+	4302	989	111187.9764	4.84
Gorai.011G138300.1	Chr11: 21310663-21311859+	1197	398	44297.58204	4.809
Gorai.011G257100.1	Chr11: 58626486-58632499+	6014	949	106085.2774	5.518

Table 4. Continued

GrRLPs Gene ID	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gorai.011G045500.1	Chr11: 3486083-3487788-	1706	414	44800.59584	5.008
Gorai.011G154100.1	Chr11: 26565941-26569827+	3887	804	90800.46824	7.827
Gorai.011G155400.1	Chr11: 27024077-27028553-	4477	1042	117476.5896	7.542
Gorai.012G117000.1	Chr12: 26955108-26958299+	3192	1043	116791.2484	8.115
Gorai.012G117200.1	Chr12: 27004616-27009241+	4626	1077	119905.409	6.421
Gorai.012G068800.1	Chr12: 10074267-10080206+	5940	755	85026.78774	6.276
Gorai.012G117100.1	Chr12: 26960606-26965495+	4890	488	54341.02354	9.885
Gorai.012G134600.1	Chr12: 30681552-30684034+	2483	719	81119.89364	5.139
Gorai.013G069300.1	Chr13: 8028092-8031208+	3117	700	78635.28154	5.443
Gorai.013G174000.1	Chr13: 46246487-46249822-	3336	1000	110967.5357	5.426
Gorai.013G236800.1	Chr13: 55462862-55469090-	6229	1128	126647.04	5.754
Gorai.013G237000.1	Chr13: 55508806-55513051-	4246	700	78157.10944	6.122
Gorai.013G237100.1	Chr13: 55514034-55520062-	6029	923	103718.4031	6.01
Gorai.006G087900.1	Chr06: 32287790-32293708+	5919	632	69371.56414	8.213
Gorai.011G029600.1	Chr11: 2174274-2179467+	5194	623	68238.42914	6.534

3.2.4 GhRLPs Biochemical Features and Chromosomal Distribution

Phylogenetic analysis has conducted that *G. hirsutum* ($4n = 4x = 52$; AADD genome) is a hybrid of two diploid species, *G. arboreum* ($2n = 2x = 26$; AA genome) and *G. raimondii* ($2n = 2x = 26$; DD genome). The A- subgenome has 57 RLPs on the 13 chromosomes. The B sub-genome has The D- subgenome has 89 RLPs on the 13 chromosomes. 5 RLPs, which are named as Gh_Sca006073G01 (scaffold6073: 676-3840+) that encodes 915 amino acids, Gh_Sca007858G02 (scaffold7858: 4481-5671-) that encodes 396 amino acids, Gh_Sca012881G01 (scaffold12881: 156-1679+) that encodes 507 amino acids, Gh_Sca024892G01 (scaffold24892: 442-1224-) that encodes 260 amino acids, Gh_Sca042904G01 (scaffold42904: 24-833-) that encodes 269 amino acids, are not well annotated (Fig 5; Fig 6).



Figure 5. Representative *G. hirsutum* A-subgenome RLPs chromosomal map



Figure 6. Representative *G. hirsutum* D-subgenome RLPs chromosomal map

On the A- subgenome in *Gossypium hirsutum*, chromosome 1 includes six RLPs which are Gh_A01G0363, Gh_A01G0383, Gh_A01G0384, Gh_A01G1976, Gh_A01G1977, and Gh_A01G2062; chromosome 2 includes seven RLPs which are Gh_A02G0194, Gh_A02G0195, Gh_A02G0215, Gh_A02G0704, Gh_A02G0705, Gh_A02G1220, and Gh_A02G1437; chromosome 3 includes three RLPs which are Gh_A03G0525, Gh_A03G0773, and Gh_A03G0841; chromosome 4 includes eight RLPs which are Gh_A04G0061, Gh_A04G0063, Gh_A04G0345, Gh_A04G0346,

Gh_A04G0347, Gh_A04G0799, Gh_A04G0802, and Gh_A04G0939; chromosome 5 includes six RLPs which are Gh_A05G0030, Gh_A05G0060, Gh_A05G3063, Gh_A05G3366, Gh_A05G3373, and Gh_A05G3868; chromosome 6 includes three RLPs which are Gh_A06G1562, Gh_A06G1649, and Gh_A06G2050; chromosome 7 includes two RLPs which are Gh_A07G1586 and Gh_A07G1749; chromosome 8 includes two RLPs which are Gh_A08G0171 and Gh_A08G1930; chromosome 9 includes two RLPs which are Gh_A09G0698 and Gh_A09G2492; chromosome 10 includes eight RLPs which are Gh_A10G0328, Gh_A10G0389, Gh_A10G1129, Gh_A10G1130, Gh_A10G1219, Gh_A10G1261, Gh_A10G1945, and Gh_A10G2219; chromosome 11 includes five RLPs which are Gh_A11G0629, Gh_A11G1919, Gh_A11G2279, Gh_A11G2283, and Gh_A11G2353; chromosome 12 includes two RLPs which are Gh_A12G0462 and Gh_A12G0722; chromosome 13 includes three RLPs which are Gh_A13G0521, Gh_A13G1285, and Gh_A13G2320 (Table 5).

On the D- subgenome in *Gossypium hirsutum*, chromosome 1 includes ten RLPs which are Gh_D01G0393, Gh_D01G0395, Gh_D01G0772, Gh_D01G2187, Gh_D01G2188, Gh_D01G2217, Gh_D01G2236, Gh_D01G2319, Gh_D01G2320, and Gh_D01G2332; chromosome 2 includes twelve RLPs which are Gh_D02G0234, Gh_D02G0236, Gh_D02G0237, Gh_D02G0257, Gh_D02G0258, Gh_D02G0259, Gh_D02G0277, Gh_D02G0280, Gh_D02G0281, Gh_D02G0282, Gh_D02G0749, and Gh_D02G1154; chromosome 3 includes two RLPs which are Gh_D03G0283 and Gh_D03G0505; chromosome 4 includes six RLPs which are Gh_D04G0212, Gh_D04G0581, Gh_D04G1292, Gh_D04G1293, Gh_D04G1294, and Gh_D04G1454;

chromosome 5 includes fourteen RLPs which are Gh_D05G0117, Gh_D05G1174, Gh_D05G3192, Gh_D05G3300, Gh_D05G3622, Gh_D05G3624, Gh_D05G3625, Gh_D05G3626, Gh_D05G3628, Gh_D05G3663, Gh_D05G3664, Gh_D05G3666, Gh_D05G3668, and Gh_D05G3670; chromosome 6 includes five RLPs which are Gh_D06G0603, Gh_D06G1447, Gh_D06G2016, Gh_D06G2019, and Gh_D06G2021; chromosome 7 includes three RLPs which are Gh_D07G0874, Gh_D07G1774, and Gh_D07G1953; chromosome 8 includes eight RLPs which are Gh_D08G0203, Gh_D08G0204, Gh_D08G0207, Gh_D08G0249, Gh_D08G0250, Gh_D08G0480, Gh_D08G0482, and Gh_D08G2296; chromosome 9 includes three RLPs which are Gh_D09G0711, Gh_D09G1576, and Gh_D09G2477; chromosome 10 includes eight RLPs which are Gh_D10G0253, Gh_D10G0334, Gh_D10G0404, Gh_D10G1221, Gh_D10G1222, Gh_D10G1277, Gh_D10G1381, and Gh_D10G1575; chromosome 11 includes twelve RLPs which are Gh_D11G0740, Gh_D11G2056, Gh_D11G2590, Gh_D11G2599, Gh_D11G2900, Gh_D11G2968, Gh_D11G2991, Gh_D11G2997, Gh_D11G2999, Gh_D11G3000, Gh_D11G3001, and Gh_D11G3383; chromosome 12 includes two RLPs which are Gh_D12G0464 and Gh_D12G0745; chromosome 13 includes four RLPs which are Gh_D13G1586, Gh_D13G2135, Gh_D13G2467, and Gh_D13G2468 (Table 5).

Table 5. The list of GhRLPs

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_A01G0363	A01:5401990..5403567+	1578	525	58354.55654	4.63
Gh_A01G0383	A01:5759260..5878720+	5946	1981	219745.361	5.5
Gh_A01G0384	A01:5915753..5916334+	582	193	21214.35274	5.7
Gh_A01G1976	A01:99753053..99756155+	2358	785	87603.43004	6.61
Gh_A01G1977	A01:99759965..99763330+	3366	1121	125111.9309	6.03
Gh_A01G2062	scaffold36_A01:4197..12063-	2553	850	94426.13774	5.25
Gh_A02G0194	A02:2168659..2170154+	1470	489	55260.93344	8.09
Gh_A02G0195	A02:2181427..2182776+	1350	449	50858.93234	8.52
Gh_A02G0215	A02:2467075..2469926-	1872	623	69052.12354	5.09
Gh_A02G0704	A02:11908278..11911160+	2694	897	98463.66704	5.78
Gh_A02G0705	A02:11960560..11963564+	2871	956	104977.2061	6.19
Gh_A02G1220	A02:72347933..72351061+	2298	765	83102.67264	4.98
Gh_A02G1437	A02:80136784..80137593-	810	269	28773.06894	5.03
Gh_A03G0525	A03:12648638..12651694-	3057	1018	112783.9254	5.41
Gh_A03G0773	A03:36618926..36621865+	2940	979	107172.0196	5.92
Gh_A03G0841	A03:46867450..46873605-	1875	624	68735.97324	5.36
Gh_A04G0061	A04:705709..708276+	2568	855	94628.08574	6.37
Gh_A04G0063	A04:852130..855243+	3114	1037	115405.9967	6.52
Gh_A04G0345	A04:8912460..8915216-	2757	918	101854.6729	6.23
Gh_A04G0346	A04:8950075..8952831-	2442	813	90625.64564	5.75
Gh_A04G0347	A04:9032391..9035477-	2760	919	102276.0995	6.65
Gh_A04G0799	A04:53385046..53387223+	2178	725	80677.58244	6.7
Gh_A04G0802	A04:53417795..53419129+	1335	444	49876.60644	6.1

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_A04G0939	A04:58397674..58399916+	2169	722	81602.60334	5.19
Gh_A05G0030	A05:523717..537521-	2538	845	94949.66944	4.84
Gh_A05G0060	A05:795299..797482+	2184	727	81595.49254	6.11
Gh_A05G3063	A05:78544669..78551342-	2085	694	78414.79494	5.91
Gh_A05G3366	A05:88176359..88182408-	2409	802	89807.28704	5.15
Gh_A05G3373	A05:88323612..88325810-	2199	732	80023.79854	5.2
Gh_A05G3868	scaffold1234_A05:57061..58530+	1470	489	53442.24444	7.03
Gh_A06G1562	A06:100080284..100083361+	3078	1025	113782.7974	5.25
Gh_A06G1649	A06:101534886..101539380-	1554	517	58354.52664	5.93
Gh_A06G2050	scaffold1356_A06:12288..13583-	1296	431	46935.07654	7.86
Gh_A07G1586	A07:61961163..61964249-	3087	1028	115098.4756	5.75
Gh_A07G1749	A07:71567508..71568410-	903	300	33852.19024	8.41
Gh_A08G0171	A08:1716135..1719173-	2394	797	87947.26244	5.09
Gh_A08G1930	A08:100118465..100121488+	3024	1007	111944.6535	5.17
Gh_A09G0698	A09:52482771..52487899+	1917	638	70488.57054	7.74
Gh_A09G2492	scaffold2332_A09:42002..49961-	1926	641	71302.20814	7.18
Gh_A10G0328	A10:3018046..3019353+	1308	435	46832.41414	6.59
Gh_A10G0389	A10:3832608..3834011-	1404	467	50645.46024	5.11
Gh_A10G1129	A10:57261653..57264212-	2028	675	75827.65494	8.09
Gh_A10G1130	A10:57322109..57326044-	3594	1197	133794.6981	5.68
Gh_A10G1219	A10:63910044..63912941+	2898	965	107938.7731	6.75
Gh_A10G1261	A10:65985947..65990289-	3489	1162	130902.6636	5.1
Gh_A10G1945	A10:96804761..96809115+	2001	666	75095.81154	4.72

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_A10G2219	scaffold2446_A10:128193..131614+	1791	596	65261.79444	6.23
Gh_A11G0629	A11:6071250..6072983-	1734	577	64339.87884	6.92
Gh_A11G1919	A11:51941824..51942630-	807	268	28900.29684	5.64
Gh_A11G2279	A11:78203228..78206338+	1818	605	67446.01874	4.96
Gh_A11G2283	A11:78505330..78509400+	2796	931	104085.0538	6.49
Gh_A11G2353	A11:80508971..80511343+	2250	749	83309.87154	5.14
Gh_A12G0462	A12:9911866..9914553+	2688	895	99134.49434	5.89
Gh_A12G0722	A12:33544853..33545656+	804	267	28546.79084	5.23
Gh_A13G0521	A13:12094627..12097443-	1626	541	60494.41674	5.22
Gh_A13G1285	A13:67310977..67313805-	2829	942	104248.0407	5.46
Gh_A13G2320	scaffold3701_A13:62957..82854-	2421	806	90176.69204	5.14
Gh_D01G0393	D01:4686244..4689216+	2973	990	110568.563	5.37
Gh_D01G0395	D01:4710630..4711238+	609	202	22282.77304	5.28
Gh_D01G0772	D01:11171287..11174092-	2592	863	95433.96284	7.11
Gh_D01G2187	D01:60805463..60808666+	3204	1067	118865.1709	5.99
Gh_D01G2188	D01:60816492..60819698+	3207	1068	119797.2212	7.36
Gh_D01G2217	D01:61033331..61036579-	3249	1082	119607.5234	6.35
Gh_D01G2236	D01:61302345..61305518+	3174	1057	117471.4661	6.51
Gh_D01G2319	scaffold3714_D01:7841..9562-	1722	573	64494.21224	7.68
Gh_D01G2320	scaffold3714_D01:22698..26156-	2235	744	82323.12914	5.76
Gh_D01G2332	scaffold3717_D01:64936..73592+	2520	839	93586.80064	7.11
Gh_D02G0234	D02:2757983..2759275+	1293	430	48438.18144	7.52
Gh_D02G0236	D02:2780046..2781380+	1335	444	50125.06264	8.48

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_D02G0237	D02:2787974..2791318+	3273	1090	123265.7761	7.06
Gh_D02G0257	D02:2942631..2946086+	3411	1136	127550.0619	6.84
Gh_D02G0258	D02:2962327..2965794+	3468	1155	130865.4446	7.14
Gh_D02G0259	D02:2979373..2983290+	3216	1071	120538.3493	5.9
Gh_D02G0277	D02:3219359..3311044-	1842	613	67162.49844	4.67
Gh_D02G0280	D02:3391502..3393436-	1935	644	71782.17354	5.78
Gh_D02G0281	D02:3410121..3412647-	2001	666	73944.46184	5.02
Gh_D02G0282	D02:3466152..3467675-	1524	507	56645.70754	5.18
Gh_D02G0749	D02:11228607..11231790+	2823	940	103573.4761	5.78
Gh_D02G1154	D02:34706378..34712492-	1875	624	68575.85624	5.36
Gh_D03G0283	D03:3146482..3147288+	807	268	28618.85714	5.03
Gh_D03G0505	D03:8674102..8678395-	3132	1043	112795.0845	4.95
Gh_D04G0212	D04:3057084..3059282+	2199	732	79973.77004	5.25
Gh_D04G0581	D04:10348872..10350492+	1419	472	53104.94304	5.74
Gh_D04G1292	D04:42420063..42423257+	3195	1064	118715.2433	7.88
Gh_D04G1293	D04:42425480..42428671+	3192	1063	118508.8099	7.93
Gh_D04G1294	D04:42475631..42479455+	3345	1114	123636.8342	6.31
Gh_D04G1454	D04:46245805..46248047+	4338	722	81703.71104	5.17
Gh_D05G0117	D05:1257420..1264155+	3435	1144	128030.9119	5.91
Gh_D05G1174	D05:10126811..10128280-	1470	489	53442.24444	7.03
Gh_D05G3192	D05:49336874..49339876-	3003	1000	111921.9183	5.46
Gh_D05G3300	D05:52598110..52600995+	2886	961	107005.7631	5.99
Gh_D05G3622	D05:60020481..60023303+	2823	940	104356.2922	6.68

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_D05G3624	D05:60049834..60050568+	735	244	27051.22844	8.92
Gh_D05G3625	D05:60089937..60093483+	2988	995	110487.9856	6.38
Gh_D05G3626	D05:60096043..60099986+	2439	812	89836.37314	5.88
Gh_D05G3628	D05:60130242..60131696+	1455	484	53487.40544	8.19
Gh_D05G3663	D05:60951364..60954780-	2928	975	107874.594	5.27
Gh_D05G3664	D05:60987411..60990062-	2418	805	89405.23484	7.09
Gh_D05G3666	D05:61045176..61049117-	3219	1072	118721.7516	6.42
Gh_D05G3668	D05:61202705..61218377-	3186	1061	118078.3525	7.13
Gh_D05G3670	D05:61264474..61267410-	2937	978	108833.772	6.53
Gh_D06G0603	D06:10255538..10261858+	2370	789	89824.95734	7.16
Gh_D06G1447	D06:47199485..47200780-	1296	431	46839.16614	7.81
Gh_D06G2016	D06:61862247..61867312-	3204	1067	119526.4389	5.42
Gh_D06G2019	D06:61906982..61908235-	1254	417	47176.01744	5.71
Gh_D06G2021	D06:61965637..62014348-	3888	1295	143230.8182	5.44
Gh_D07G0874	D07:11364649..11367705-	3057	1018	113427.732	5.6
Gh_D07G1774	D07:42294987..42298070-	3084	1027	114881.4411	6.04
Gh_D07G1953	D07:48762756..48765806-	3051	1016	112973.9154	7.31
Gh_D08G0203	D08:1859230..1862454-	3225	1074	119404.0492	5.04
Gh_D08G0204	D08:1905739..1908675-	2937	978	109376.8378	6.02
Gh_D08G0207	D08:1925382..1928360-	2979	992	109805.6391	5.26
Gh_D08G0249	D08:2373266..2376160-	2808	935	103130.4891	5.14
Gh_D08G0250	D08:2383497..2386392-	2532	843	93672.53304	5.05
Gh_D08G0480	D08:5270612..5276574+	2517	838	92879.05524	4.6

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_D08G0482	D08:5317716..5322404+	2718	905	101443.5412	5.46
Gh_D08G2296	D08:62587855..62590878+	3024	1007	111905.5319	5.01
Gh_D09G0711	D09:31061077..31066200+	1905	634	69663.85564	7.68
Gh_D09G1576	D09:43363264..43366650+	3387	1128	124962.2648	5.52
Gh_D09G2477	scaffold4365_D09:175641..179139+	3159	1052	117024.1279	7.02
Gh_D10G0253	D10:2155673..2159089+	1845	614	67312.28574	6.32
Gh_D10G0334	D10:2895243..2896556+	1314	437	47151.81454	6.57
Gh_D10G0404	D10:3675891..3678005-	1425	474	51447.21504	4.91
Gh_D10G1221	D10:21382334..21407058+	4458	1485	167539.4985	6.36
Gh_D10G1222	D10:21449744..21454103+	2184	727	81256.53634	5
Gh_D10G1277	D10:23230213..23233603-	3057	1018	113488.6748	6.53
Gh_D10G1381	D10:27413598..27417839-	2181	726	82102.18424	6.5
Gh_D10G1575	D10:42672154..42675081-	2928	975	106840.0078	5.16
Gh_D11G0740	D11:6367007..6368746-	1740	579	64469.93134	6.53
Gh_D11G2056	D11:28222802..28223608+	807	268	28970.47774	6.46
Gh_D11G2590	D11:53835125..53838872+	2703	900	100998.6444	5.82
Gh_D11G2599	D11:53981643..53982242-	600	199	21761.03174	5.7
Gh_D11G2900	D11:59199595..59201841-	1941	646	71630.28784	4.98
Gh_D11G2968	D11:60517557..60518699-	1143	380	42747.91794	5.69
Gh_D11G2991	D11:60990167..60990754-	588	195	21862.88964	4.65
Gh_D11G2997	D11:61158203..61161016-	2814	937	103926.9035	5.65
Gh_D11G2999	D11:61254716..61257799-	3084	1027	113841.6415	5.37
Gh_D11G3000	D11:61272288..61275236-	2949	982	109608.3582	5.6

Table 5. Continued

Gene ID GhRLPs	Chromosome location	Sequences length	A.a Length	Protein Mass	pI
Gh_D11G3001	D11:61298920..61301835-	2916	971	108809.1728	5.45
Gh_D11G3383	scaffold4497_D11:715..3618-	2904	967	108116.4722	6.17
Gh_D12G0464	D12:7615360..7620741+	2688	895	99289.64474	5.73
Gh_D12G0745	D12:18212855..18213658+	804	267	28644.01214	5.5
Gh_D13G1586	D13:48513932..48516892-	2961	986	109230.7104	5.34
Gh_D13G2135	D13:56971416..56977262-	2880	959	107851.088	5.55
Gh_D13G2467	scaffold4656_D13:23724..26018+	1359	452	51430.86924	5.77
Gh_D13G2468	scaffold4656_D13:28196..29974+	1497	498	55264.88354	5.22
Gh_Sca006073G01	scaffold6073:676..3840+	2748	915	101418.5425	6.37
Gh_Sca007858G02	scaffold7858:4481..5671-	1191	396	44300.55154	5.3
Gh_Sca012881G01	scaffold12881:156..1679+	1524	507	56576.52064	5.29
Gh_Sca024892G01	scaffold24892:442..1224-	783	260	29401.45204	5.02
Gh_Sca042904G01	scaffold42904:24..833-	810	269	28773.06894	5.03

3.3 Distribution of LRRs in Arabidopsis and Cotton

A RLP includes leucine-rich repeats motifs such as plant specific LxxLxxLxLxxNxLt/sgxIpxxLG and LxxLxLxxN/CxL residues and cysteine-rich LRR (LRRNT, and LRRCT). The number of LRRs in plant RLPs varies from 4 to 40 in plants depending on the length of peptides. All RLPs were divided into eight categories based on the number of LRRs the RLPs possessed in intervals of five (Table 6)

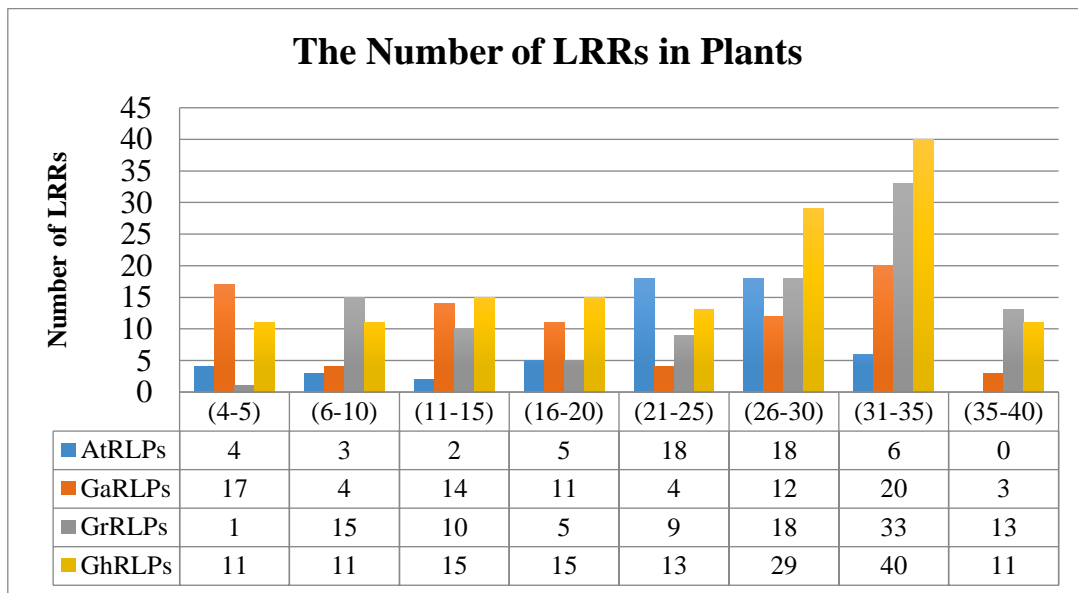


Table 6. Categorized LRRs in plants

The lowest number of LRRs containing RLPs in *Arabidopsis thaliana* was in the 11-15 intervals, whereas the highest number of LRRs was in the 21-25 and 26-30 intervals. The lowest number of LRRs containing RLPs in *Gossypium arboreum* was in the 35-40 intervals, whereas the highest number of LRRs was in the 31-35 intervals. The lowest number of LRRs containing RLPs in *Gossypium raimondii* was in the 4-5 intervals, whereas the highest number of LRRs was in the 31-35 intervals. The lowest

number of LRRs containing RLPs in *Gossypium hirsutum* was in the 4-5, 6-10, and 36-40 intervals, whereas the highest number of LRRs was in the 31-35 intervals.

3.4 Classifications and Phylogenetic Analysis of RLPs

3.4.1 OrthoMCL Analysis in Cotton

Most of the genes in tetraploid cotton have evolved from duplication events, which provide new gain-function or lost-function. However, these homologs may not be clustered with true homologs in other species. Subsequently, complicated domain structures of many proteins beclouds exactly clustering protein families and sharing similar domains, which involve in conservative motifs, in proteins cause of cluster mistakenly in many protein families. A third challenge in clustering orthologs result from uncompleted genome sequences (Li., L. et al. 2003).

The inflation value is a crucial parameter in MCL algorithm, which regulates the cluster granularity. A higher inflation value indicates higher cluster granularity. The final output from MCL analysis illustrates at least two species clusters containing sequences ID of orthologs (Li., L. et al. 2003).

Investigating the *G. arboreum* RLPs against the OrthoMCL database revealed that they clustered into 12 different orthologs. 35 GaRLPs belong to the OG5_130149 ortholog cluster, 20 GaRLPs belong to the OG5_134835 ortholog cluster, 15 GaRLPs belong to the OG5_164541 ortholog cluster, three GaRLPs belong to the OG5_164541 ortholog cluster, three GaRLPs belong to the OG5_128518 ortholog cluster, two GaRLPs belong to the OG5_212573 ortholog cluster, two GaRLPs belong to the OG5_244942 ortholog cluster, one GaRLP belongs to the OG5_136779 ortholog cluster,

one GaRLP belong to the OG5_189973 ortholog cluster, one GaRLP belongs to the OG_170335 ortholog cluster, one GaRLP belongs to the OG5_190449 ortholog cluster, one GaRLP belongs to the OG5_159729 ortholog cluster, and two GaRLPs did not cluster to any ortholog group.

Investigating the *G. raimondii* RLPs against the OrthoMCL database revealed that they clustered into nine different orthologs. 69 GrRLPs belong to the OG5_130149 ortholog cluster, 20 GrRLPs belong to the OG5_159698 ortholog cluster, four GrRLPs belong to the OG5_128518 ortholog cluster, three GrRLPs belong to the OG5_164541 ortholog cluster, two GrRLPs belong to the OG5_170335 ortholog cluster, two RLPs belong to the OG5_177950 ortholog cluster, two GrRLPs belong to the OG5_134835 ortholog cluster, one GrRLP belongs to the OG5_159729 ortholog cluster, one GrRLP belongs to the OG5_212573 ortholog cluster, and three of them did not cluster to any ortholog group.

Investigating the *G. hirsutum* RLPs against the OrthoMCL database revealed that they clustered into 14 different orthologs. 71 GhRLPs belong to the OG5_130149 ortholog cluster, 25 GhRLPs belong to the OG5_159698 ortholog cluster, 20 GhRLPs belong to the OG5_134835 ortholog cluster, seven GhRLPs belong to the OG5_164541 ortholog cluster, six GhRLPs belong to the OG5_128518 ortholog cluster, four GhRLPs belong to the OG5_177950 ortholog cluster, four GhRLPs belong to the OG5_170335 ortholog cluster, three GhRLPs belong to the OG5_244942 ortholog cluster, two GhRLPs belong to the OG5_190826 ortholog cluster, two GhRLPs belong to the OG5_189973 ortholog cluster, two GhRLPs belong to the OG5_188400 ortholog cluster,

two GhRLPs belong to the OG5_159729 ortholog cluster, one GhRLP belongs to the OG5_212573 ortholog cluster, one GhRLP belongs to the OG5_127136 ortholog cluster, and one of them did not cluster with any ortholog group (APPENDIX D).

3.4.2 Determination of Conserved Motifs in RLPs

The extracellular LRR-RLPs were divided into several domains named between A to G. An alterable tandem repeats indicated in the LRR domain C1 and linker domain C2. On the other hand, the LRR containing C3 domain and the linker domain D were conserved. For this reason, the C3 and D domains were preferred to perform the phylogenetic analysis (Fig 7) (Fritz-Laylin L.K., et al. 2005; Petre, B., et al 2014).



Figure 7. Conserved domain structure of RLPs (A-G)

3.4.3 Phylogenetic Trees for Four Species

3.4.3.1 Arabidopsis RLPs Phylogenetic Analysis

Arabidopsis RLPs were aligned via ClustalW algorithm and a rooted phylogenetic tree with branch length (UPGMA) was built. The tree was visualized by using the Interactive Tree of Life version 3.4.3 and the RLPs were divided into eight groups. Group I was marked in red and was further divided in two subgroups: dashed red line subgroup, which includes six RLPs descending from AtRLP6/7 and solid red line subgroup which includes nine RLPs descending from AtRLP52. Group II was marked in orange and includes thirteen RLPs descending from AtRLP54. Group III is in yellow and includes seven RLPs. Group IV is in green and includes AtRLP2 and AtRLP3, which are close paralogs. Light purple color represents group V and includes AtRLP44 and AtRLP57, which are close paralogs too. Group VI in blue includes eleven RLPs descending from AtRLP17. Group VII in purple includes AtRLP51 and AtRLP55, which are close paralogs and finally group VIII marked with light green includes AtRLP4, which encodes malectin-like RLP (Fig 8).

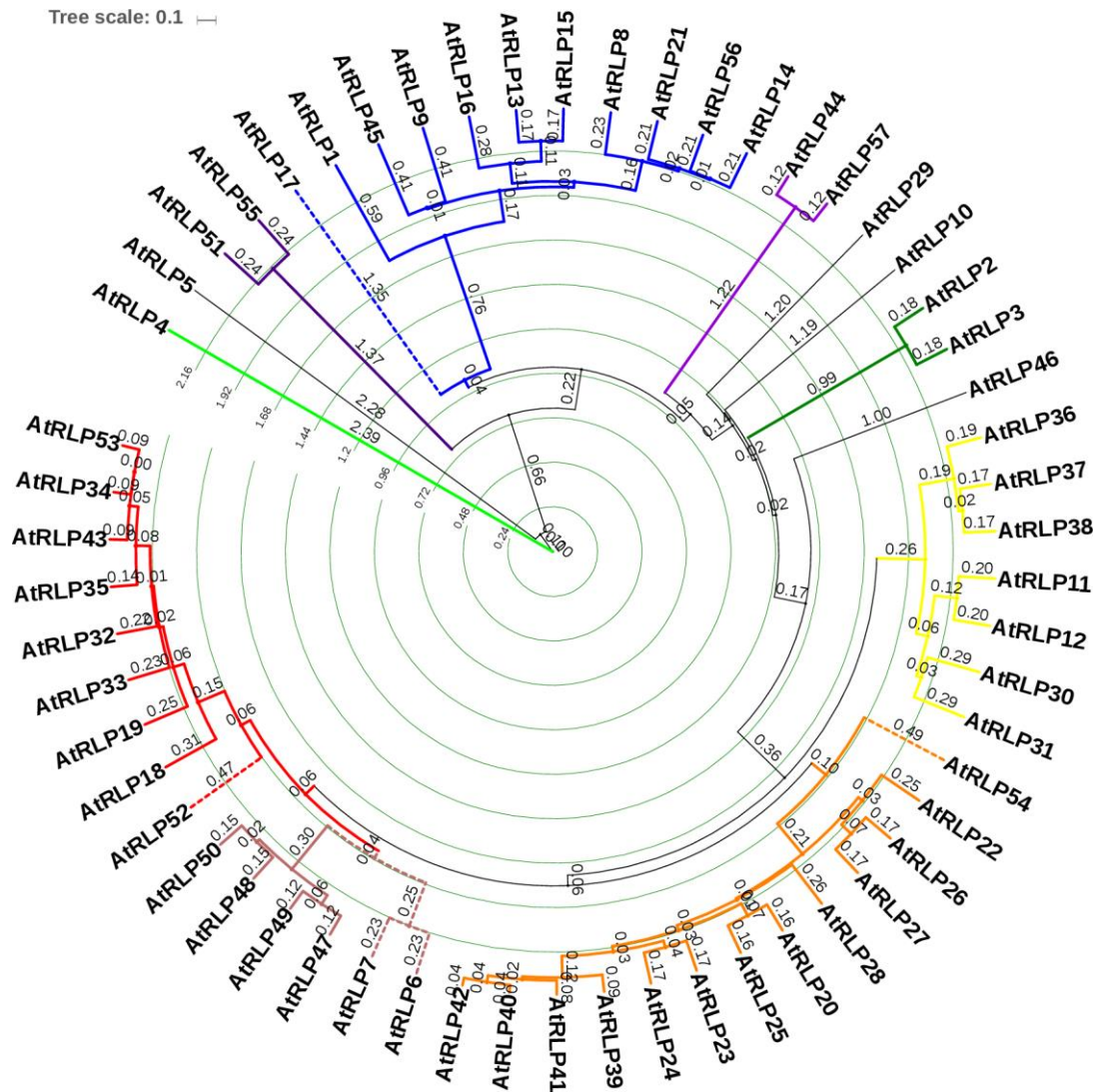


Figure 8. Arabidopsis Receptor-like protein tree

3.4.3.2 *G. arboreum* RLPs Phylogenetic Analysis

G. arboreum RLPs were aligned via ClustalW algorithm and a rooted phylogenetic tree with branch length (UPGMA) was built and analyzed as before. Group I in red is subdivided into two subgroups: dashed red lines subgroup including ten RLPs,

3.4.3.3 *G. raimondii* RLPs Phylogenetic Analysis

G. raimondii RLPs were aligned via ClustalW algorithm for multiple sequences alignment and a rooted phylogenetic tree with branch length (UPGMA) was built and analyzed as above. Group I indicated in red includes forty-seven RLPs descending from Gorai.002G039000.1 and Group II indicated in orange includes nine RLPs descending from Gorai.012G117000.1 and Gorai.012G117100.1. Group III in yellow includes twelve RLPs descending from Gorai.011G154100.1 and Gorai.011G155400.1. Group IV in green includes twenty-five RLPs descending from Gorai.013G069300.1 and Gorai.012G134600.1. Group V in blue includes three RLPs descending from Gorai.011G045500.1. Group VI in purple includes three RLPs descending from Gorai.007G226700.1. Group VII in light green is composed of two RLPs, Gorai.011G029600.1 and Gorai.006G087900.1, which encode malectin-like receptor like protein (Fig10).

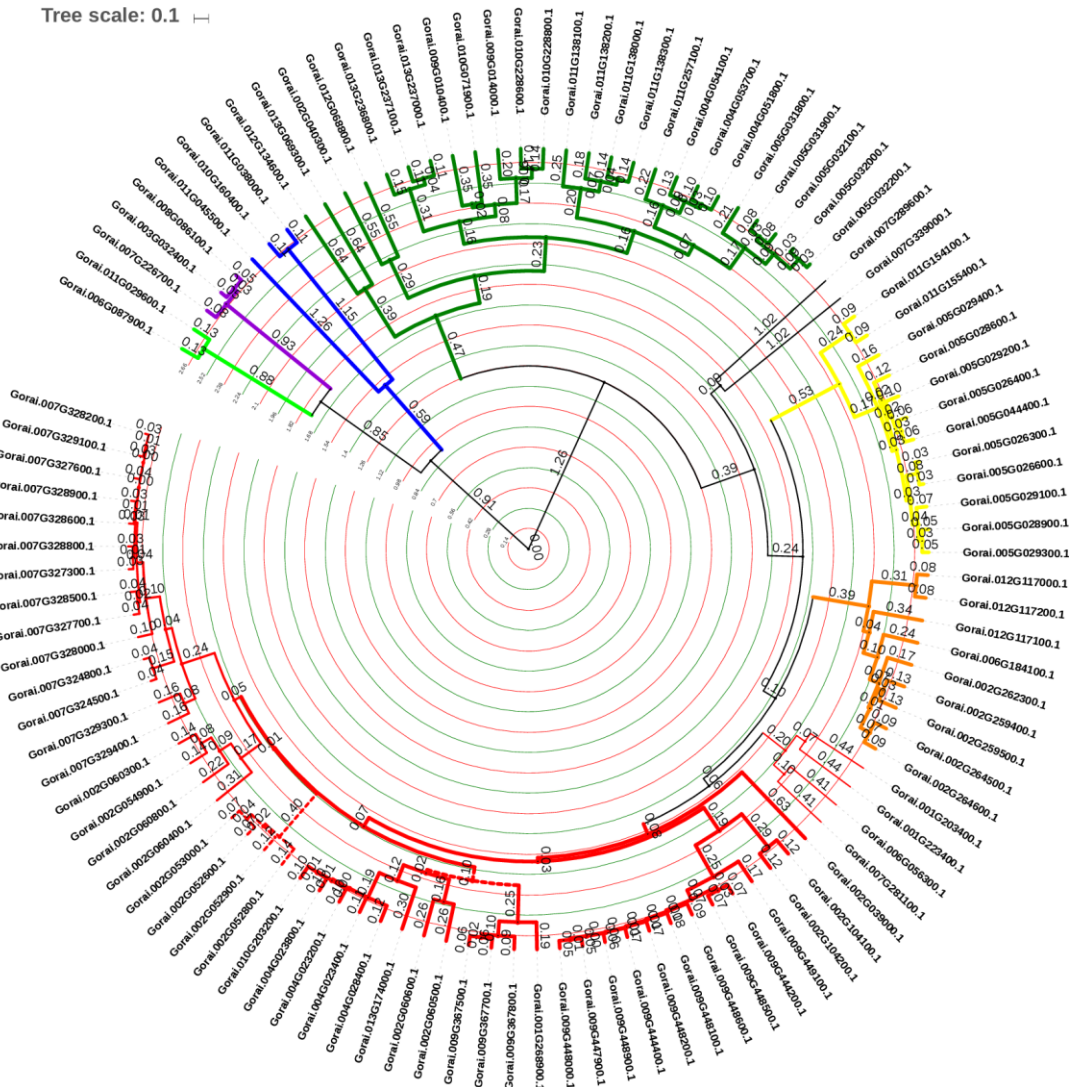
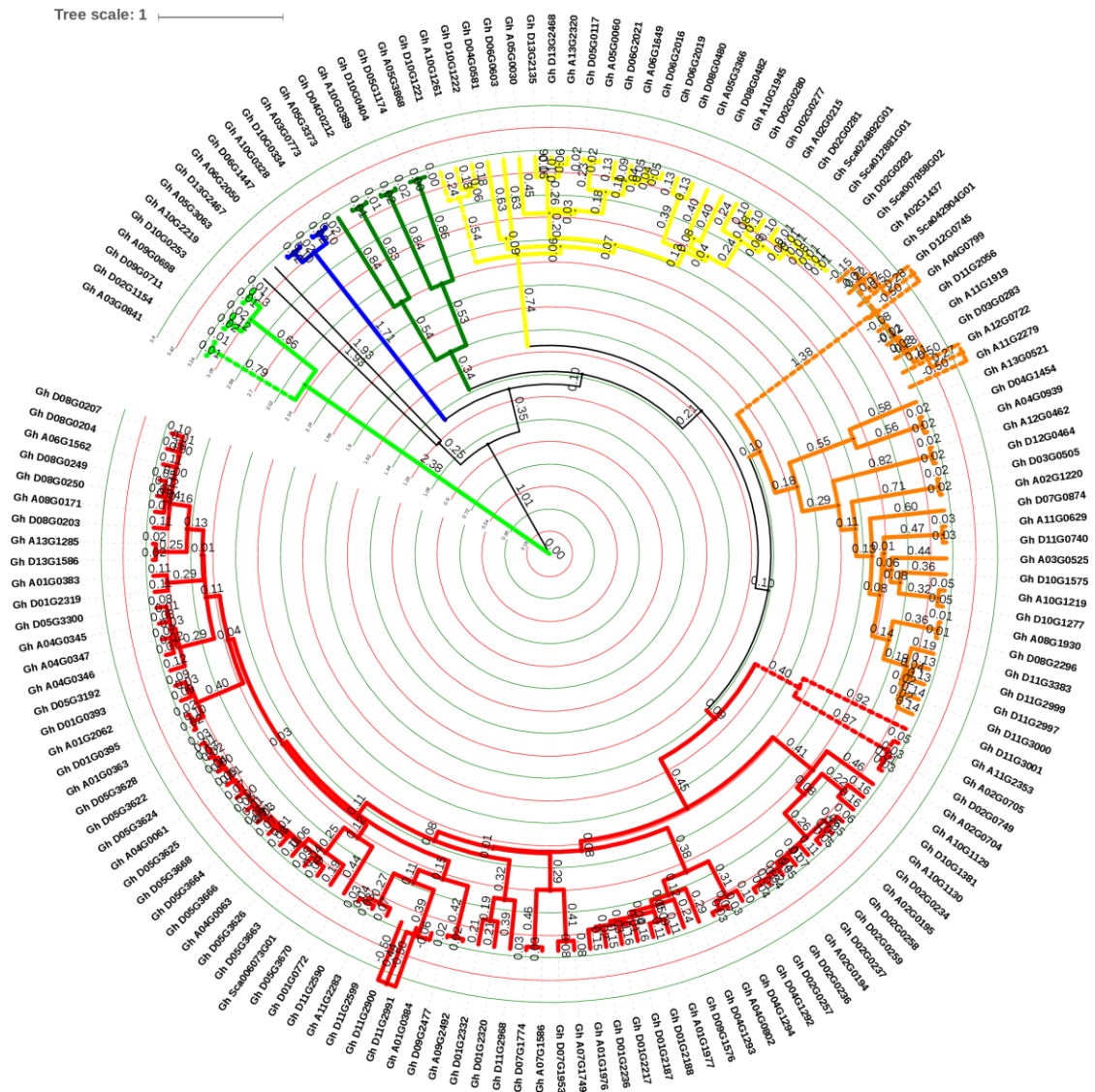


Figure 10. *Gossypium raimondii* Receptor-like protein tree

3.4.3.4 *G. hirsutum* RLPs Phylogenetic Analysis

G. hirsutum RLPs were also aligned via ClustalW and the rooted phylogenetic tree was visualized by using the Interactive Tree of Life version 3.4.3. Group I indicating red includes seventy-five RLPs descending from dashed-red –line.



Group II in orange includes thirty RLPs and group III in yellow includes twenty-seven 27 RLPs. Group IV in green includes seven RLPs descending from GhA03G07773. Group V in blue includes four RLPs, which are close homologs. Group VI in light green includes six malectin-like RLPs (Fig 11).

3.4.3.5 *Gossypium* spp. and *Arabidopsis* RLPs Phylogenetic Analysis

Plant RLPs were also aligned via ClustalW and the rooted phylogenetic tree was visualized by using the Interactive Tree of Life version 3.4.3 (Fig 8). The phylogenetic analysis was divided into 10 major clusters based on *Arabidopsis* RLPs distribution with sequences similarities. Group I indicating red includes two hundreds eight RLPs including 35 AtRLPs. Group II indicating orange includes one hundred twenty two RLPs descending from AtRLP2 and AtRLP3, which are close homolog. . Group III indicating yellow includes five RLPs descending from AtRLP46. . Group IV indicating green includes fifteen RLPs descending from AtRLP44 and AtRLP57, which are close homolog. . Group V indicating blue includes ten RLPs descending from AtRLP4. . Group VI indicating purple includes three RLPs descending from AtRLP10. Group VII indicating light purple includes three RLPs descending from AtRLP16. Group VIII indicating brown includes nine RLPs descending from AtRLP51 and AtRLP55, which are close homolog. Group IX indicating light green includes five RLPs descending from AtRLP29. Group X indicating gray includes five RLPs descending from AtRLP17 (Fig 12).

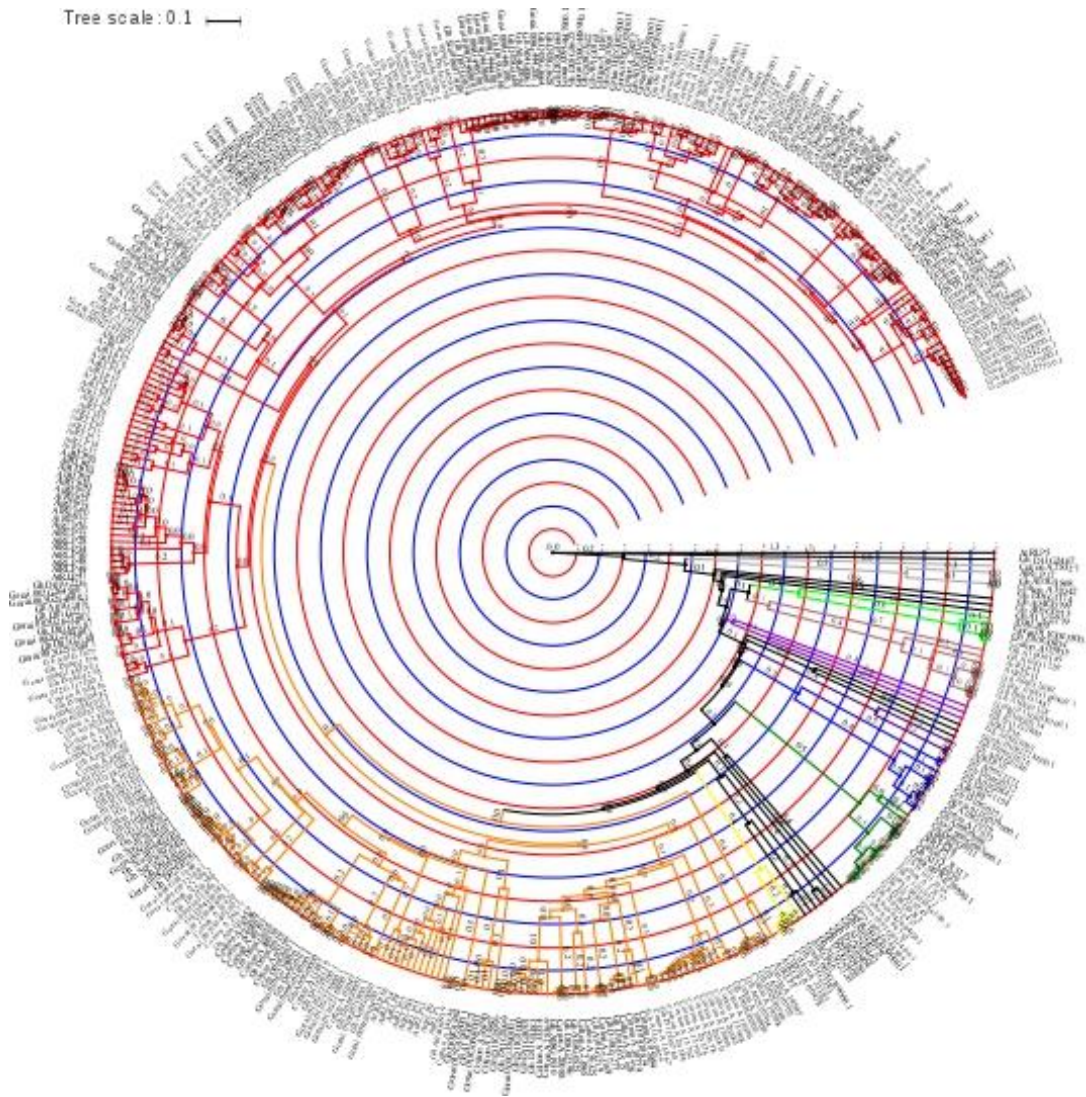


Figure 12. Distribution of all RLPs in phylogenetic tree

Most Arabidopsis RLPs are located in the red and orange clusters. Phylogenetic tree analysis with other plant LRR-RLPs suggested that cotton LRR-RLPs are likely evolved independently and clustered together.

3.5 RLPs Evolves Independently and Rapidly

Evolution processes through divergent alternations of genes, proteins, and phenotypic changes in time (Soskine M. and Tawfik D.S., 2010). Some species evolve convergently in distance relative taxa with identical selection process (Hoy, R. R., 2012). Because of the spread of adaptive phenotypic convergence in nature, some gene families might be useful to peruse evolution at the sequence level (Zhang, J. Z. & Kumar, S., 1997). Polyploidy plays a critical role in genome evolution. One of the critical questions in plant science and crop breeding is to understand how plants evolved through genome evolution mechanisms. Loss of functions, gain of functions, gene movement, and chromosomal structural changes cause diploidization after whole or partial genome duplication. *Gossypium hirsutum* is a hybrid of its antecedent A and D diploid progenitors. The polyploidization and hybridization of antecedent diploid genomes provide that cotton is a significant model organism for polyploid genome evolution studies (Rong, J., et al. 2010).

The diploid genomes of *G. arboreum* and *G. raimondii* include 86 and 107 RLPs. 20 RLPs in both are single-copy genes. They have mostly identical DNA sequences with exception of point mutations. 20 RLPs in *G. arboreum* genome are utterly different than *G. raimondii* RLPs, and 18 RLPs in *G. raimondii* genome are also species-specific genes, which are called “Singletons”. Furthermore, 46 RLPs in *G. arboreum* and 69 RLPs in *G. raimondii* create different numbers of cluster within each other. Each cluster includes at least one or more gene(s) with other species (Appendix C).

G. hirsutum is allotetraploid cotton and includes 151 RLPs. 16 RLPs in *G. hirsutum* genome are single-copy genes with *G. arboreum* RLPs. They have mostly identical DNA sequences with exception of point mutations. 38 RLPs in *G. hirsutum* genome are utterly different than *G. arboreum* RLPs, and 24 RLPs in *G. arboreum* genome are also species-specific genes, which are called “Singletons”. Furthermore, 97 RLPs in *G. hirsutum* and 46 RLPs in *G. arboreum* create different numbers of clusters within each other. Each cluster includes at least one or more gene(s) with other species (Appendix C).

G. hirsutum is allotetraploid cotton and includes 151 RLPs. 21 RLPs in *G. hirsutum* genome are single-copy genes with *G. raimondii* RLPs. They have mostly identical DNA sequences with exception of point mutations. 34 RLPs in *G. hirsutum* genome are utterly different from *G. raimondii* RLPs, and 19 RLPs in *G. raimondii* genome are also species-specific genes, which are called “Singletons”. Furthermore, 96 RLPs in *G. hirsutum* and 67 RLPs in *G. raimondii* create different numbers of clusters within each other. Each cluster includes at least one or more gene(s) with other species (Appendix C).

These findings suggested that *Gossypium ssp.* evolved convergently and divergently through evolutionary mechanisms such as specification, hybridization, and polyploidation (Appendix D).

3.6 Cloning Receptor-Like Proteins for VIGS

3.6.1 Determination of Candidate RLPs from GhRLPs

Candidate cotton RLPs were chosen based on homology to well-studied RLPs on Arabidopsis and by comparing them through phylogenetic tree analysis. At the amino acid level, blast analysis against the Gossypium protein database illustrated the presence of cotton Gh_A07G1586 (Gh_RLP23_1) as the closest ortholog to AtRLP23 with 33.7% identity and 47.4% similarity. Other orthologs for AtRLP23 include Gh_D07G1774 (Gh_RLP23_1) with 34.2% identity and 48.6% similarity, Gh_D07G1953 (GhRLP23_2) with 34.8% identity and 49.0% similarity, Gh_A11G2283 (GhRLP23_3A) with 33.4% identity and 48.4% similarity and Gh_D09G2477 (GhRLP23_3D) with 33.0% identity and 46.0% similarity. Gh_D01G0395 is the ortholog of AtRLP34 with 10.3 % identity and 13.8% similarity and Gh_D11G2968 is the ortholog of AtRLP27 with 18.6% identity and 26.9 similarity. Furthermore, Gh_D05G3192 is the ortholog of AtRLP7 (34.8% identity and 50.9% similarity), Gh_A03G0773 is the ortholog of AtGSO1 (35.9% identity and 48.1% similarity), Gh_D04G1293 is the ortholog of AtRLP6 (34 % identity and 49.3% similarity), Gh_A12G0722 is the ortholog of AtRLP44 (73.4% identity and 83.2% similarity) and Gh_D13G2467 is the ortholog of AtRLP9 (13.6% identity and 21.2% similarity) (Appendix E). The coding DNA sequences of candidate RLPs were retrieved from the Gossypium cDNA database to design primers for VIGS using Primer3web version 4.0.0 webpage (Primers are provided below Table 7).

Table 7. Used primers in this study

G. Hirsutum Gene IDs	Primer IDs	5'-Sequences-3'	Length (bp)	Amplicon (bp)
Gh_A07G1586-Gh_D07G1774	GhRLP23_1_F	CGGAATTCGGAGTCGATCGGGAACCTCA	28	280
Gh_A07G1586-Gh_D07G1774	GhRLP23_1_R	GGGGTACCCTGGACGAAACTGTTGGTGG	28	280
Gh_D07G1953	GhRLP23_2_F	CGGAATTCCTGGATTCTGAACGCTCTGC	28	336
Gh_D07G1953	GhRLP23_2_R	GGGGTACCTCCAACCGTATAGCCTTGT	28	336
Gh_D09G2477	GhRLP23_3 _D _F	CGGAATTCCTGCTTTCCCTAACCAGTCCC	28	300
Gh_D09G2477	GhRLP23_3 _D _R	GGGGTACCGAAACAGGTTGGAGCTGAGG	28	300
Gh_A11G2283	GhRLP23_3 _A _F	CGGAATTCCTGAGTTCCCAAGTTTCTACG T	31	262
Gh_A11G2283	GhRLP23_3 _A _R	GGGGTACCCGATTTCTCCCACAGCATG	28	262
Gh_D01G0395	Gh_D01G0395_F	CGGAATTCCTTTGGTAGCACTTGAGTCGC	28	313
Gh_D01G0395	Gh_D01G0395_R	GGGGTACCTCCACACCATAACCCATCA	28	313
Gh_D11G2968	Gh_D11G2968_F	CGGAATTCCTCAAAGTCATCGGAAACT	28	301
Gh_D11G2968	Gh_D11G2968_R	GGGGTACCAAAGCTCTCTATTTGGCCGT	28	301
Gh_D05G3192	Gh_D05G3192_F	CGGAATTCCTGGGAAGCATCTATCAGC	28	325
Gh_D05G3192	Gh_D05G3192_R	GGGGTACCCAAGTTGGTTGTCATTGAGGTT	30	325
Gh_A03G0773	Gh_A03G0773_F	CGGAATTCGCAGCCTCAAACAACATGCT	28	327
Gh_A03G0773	Gh_A03G0773_R	GGGGTACCACTTGGAAATGCTACCCGTC	28	327
Gh_D04G1293	Gh_D04G1293_F	CGGAATTCGCAATGCAGATGGTCAGGTT	28	305
Gh_D04G1293	Gh_D04G1293_R	GGGGTACCTGGCTTCTTGAGTTTCAGTGA	29	305
Gh_A12G0722	Gh_A12G0722_F	CGGAATTCCTTAGTCAACCTGGCAGTGC	28	250
Gh_A12G0722	Gh_A12G0722_R	GGGGTACCAGTAGCATTGAACCTTGGCA	28	250
Gh_D13G2467	Gh_D13G2467_F	CGGAATTCCTGGTGGGTTGCCCTTTCAAAT	28	308
Gh_D13G2467	Gh_D13G2467_R	GGGGTACCAGTGAGGGTTCCAATATCTCGA	30	308
GhCLA1	GhCLA1-F	GGAATTCACACAACATCGATGATTTAG	26	500
GhCLA1	GhCLA1-R	GGGGTACCATGATGAGTAGATTGCAC	26	500

3.6.2 Silencing of GhRLPs by Agrobacterium-Mediated VIGS

Virus Induced Gene Silencing (VIGS) approach is a very effective tool for the investigation of gene function in many crop species (Xiquan G. et al. 2011). We used GhCLA1 gene, which encodes a 1-deoxyxylulose 5-phosphate synthase required for chloroplast development to get photobleaching phenotype, thus can be used as a visual marker for VIGS efficiency (Mendel M.A., et al. 1996; Xiquan G. et al, 2011). Full-length Arabidopsis *RLP23* sequence was used as a query to blast them against the *Gossypium hirsutum* cDNA database to identify their cotton orthologs (Gh_A07G1586-Gh_D07G1774; GhRLP23_1, Gh_D07G1953; GhRLP23_2, Gh_A11G2283; GhRLP23_3A, Gh_D09G2477; GhRLP23_3D). Other candidates RLPs were chosen as ancestors based on clusters on GhRLPs phylogenetic tree. These ancestors corresponding homologs in Arabidopsis RLPs were used to rename cotton orthologs (Gh_D01G0395; AtRLP34, Gh_D11G2968; AtRLP27, Gh_D05G3192; AtRLP7, Gh_A03G0773; GSO1, Gh_D04G1293; AtRLP6, Gh_A12G0722; AtRLP44, and Gh_D13G2467; AtRLP9). The primers were designed to amplify conserved DNA fragments of the orthologs for inserting them into the VIGS vector (pTRV2). The mixed 1:1 ratio agrobacterium culture including both VIGS vectors, pTRV1 and pTRV2: Ghgene were inoculated into the emerging two cotyledons of cultivar FM9169B2F *Gossypium hirsutum* below 1 cm from the cotyledons with sterile insulin syringe. At least 15 plants were used for each construct. After GhCLA1 VIGSed plants showed photobleaching phenotype, a solution of 0.001% Tween containing 10^7 conidia/ml *F. oxysporum* spores was inoculated into the stems of cottons with sterile insulin syringe.

Infected plants were observed for 25 days post-inoculation to calculate disease index of population.

3.6.3 Phenotype Observation and Disease Severity

Unclear phenotype in a single mutant background of the AtRLP genes complicates to clarify their roles in biological process (Wang G., et al, 2008). Condition-specific manner or multiple functions for a gene may affect the obvious phenotype. Previous studies on AtRLP genes with specific elicitors suggested that gene expression levels, phenotype of a single mutant background and over expression of AtRLP genes can cause variations (Wu J., et al. 2016). We didn't observe any physiological phenotype after gene silencing of cotton seedlings. *Fusarium* inoculated plants were observed 25 days post inoculation with green fluorescent protein (GFP) and GhCLA1 as control groups to measure VIGS efficiency and disease severity (Appendix F) (Fig 13).

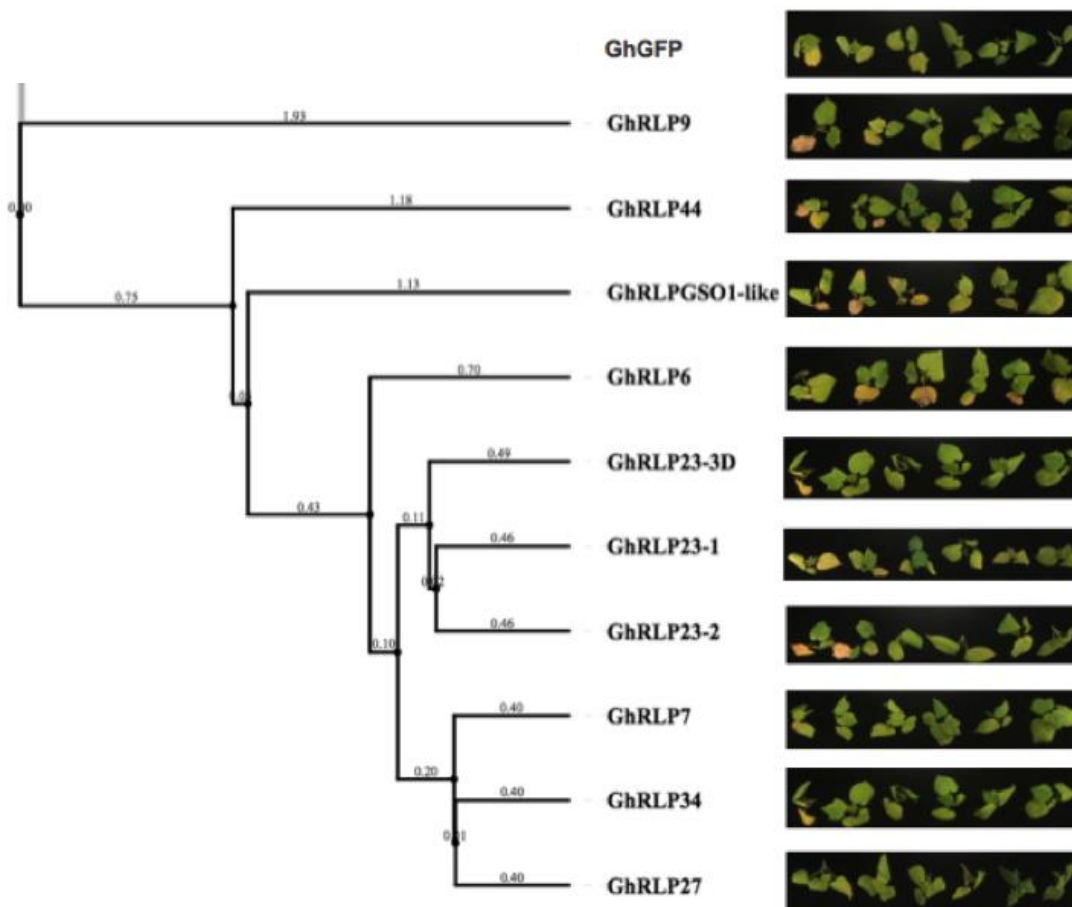


Figure 13. Phylogenetic tree and phenotypic observation of VIGSed plants

At least 15 plants were infected with *Fusarium oxysporum* below 1 cm cotyledons with sterile insulin syringe. VIGSed plants revealed three different phenotypes such as wilt, yellowish, and death cells. The disease symptom was divided into 4 levels (level 0-4) and calculated using the formula $DI = (\sum (n \times \text{number of seedlings at level } n)) / (4 \times \text{number of total seedlings} \times 100)$ (n refers the severity of the disease level of the plants). After the infected plant become obvious, disease index of

infected and VIGSed populations were scored based on true leaves phenotype containing cotyledons during 25 days post inoculation (Table 8).

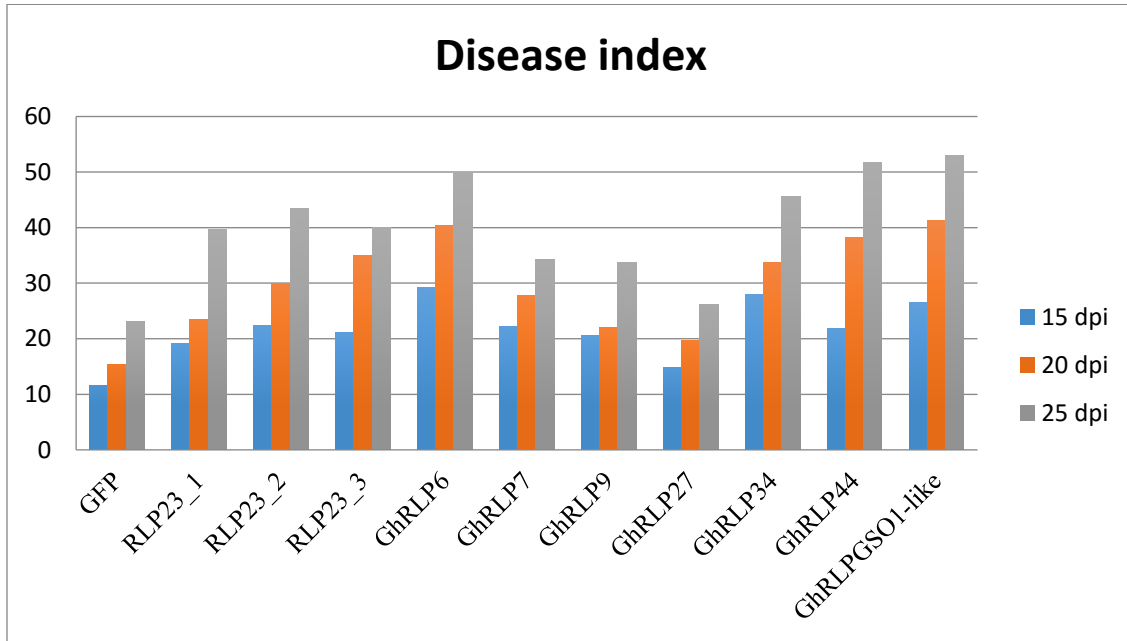


Table 8. Disease index of infected cottons

GhRLPGSO1-like, GhRLP44, GhRLP6, and GhRLP34 are more susceptible in comparison to GFP with 2.30, 2.24, 2.16, and 1.97 folds increase, respectively (Appendix F).

4. CONCLUSIONS

Cotton is a significant crop worldwide as with economic characteristics involving disease resistance, fiber production, and oil ingredient. The Arabidopsis genome has 56 of 57 e-LRR-RLPs and one malectin-like RLP. The protein sequences of AtRLPs were used as queries to blast against the *Gossypium* protein database at <https://www.cottongen.org/tools/blast/blast>. Blast analysis indicated that *G. arboreum*, *G. raimondii*, and *G. hirsutum* genome have 85, 105, and 145 eLRR-RLPs, respectively. The Arabidopsis genome has one malectin-like receptor-like protein, whereas *G. arboreum*, *G. raimondii*, and *G. hirsutum* have 1, 2 and 6, respectively (Table 2). LRR-RLPs appear to be evolved rapidly as the homology between cotton and Arabidopsis LRR-RLPs is often below 40%. The chromosomal location, full length, amino acid sequence and gene IDs were obtained at the *Gossypium* genome database in all cotton species. The protein mass and isoelectronic point of the whole RLPs were calculated by using IPC server (Lukasz P. Kozlowski, 2016) (Table 3, Table 4, and Table 5). To classify the evolutionary relationships of the receptor like proteins in all species a phylogenetic analysis was conducted using UPGMA.

Well-studied RLP23 in Arabidopsis were used to identify its orthologs in *G. hirsutum* genome database and some ancestors on GhRLPs phylogenetic clusters were used to set up VIGS construct which were transferred to *Agrobacterium*. *Agrobacterium* cultures including pTVR1 and pTVR2: CLA1 or pTVR2: GFP, or pTVR2: Gene were infiltrated into stems of cotton seedlings below 1 cm cotyledons with sterile insulin syringe. After

photobleaching phenotype observed, *Fusarium oxysporum* (10^7 spors.ml¹) was inoculated into stems. GhRLPGSO1-like, GhRLP44, GhRLP6, and GhRLP34 silenced plants showed more susceptibility than others. These results suggest that receptor-like proteins in plants might be valuable candidate genes for further studies to investigate the disease resistance potential and breeding approaches.

REFERENCES

- Albert I., et al. An RLP23–SOBIR1–BAK1 complex mediates NLP-triggered immunity (2015). *Nature Plants*, 1, 15140.
- Bej A., Sahoo B.R., Swain B., Basu M., Jayasankar P., Samanta M. LRRsearch: An asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors (2014). *Computers in Biology and Medicine*, 53, 164-170.
- Belfanti E., Silfverberg-Dilworth E., Tartarini S., et al. The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety (2004). *Proceedings of the National Academy of Sciences of the United States of America*. 101, 3, 886-890.
- Buhrow L.M., Clark S.M., and Loewen M.C. Identification of an attenuated Barley Stripe Mosaic Virus for the virus-induced gene silencing of pathogenesis-related wheat genes (2016). *Plant Methods*, 12, 1-12.
- Catanzariti A.M., Lim G.T.T. and Jones D.A. The tomato I-3 gene: a novel gene for resistance to Fusarium wilt disease (2015). *New Phytologist*, 207, 106–118.
- Chen C.W., Panzeri D., Yeh Y.H., et al. The Arabidopsis malectin-like leucine-rich repeat receptor-like kinase IOS1 associates with the pattern recognition receptors FLS2 and EFR and is critical for priming of pattern-triggered immunity (2014). *The Plant Cell*, 26, 3201-3219.

Chen J.Y, Huang J.Q., et al. Genome-wide analysis of the gene families of resistance gene analogues in cotton and their response to *Verticillium* wilt (2015). *BMC Plant Biology*, 15, 148.

Davis, R.M., Colyer, P.D., Rothrock, C.S., Kochman, J. K. *Fusarium* wilt of cotton: population diversity and implications for management (2006). *Plant Disease*, 90, 6, 692-703.

De Jonge R., Peter van Esse H., Maruthachalam K., et al. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing (2012). *Proceedings of the National Academy of Sciences of the United States of America*. 109, 13, 5110-5115.

Dixon M.S., Jones D.A., Keddie J.S., Thomas C.M., Harrison K., Jones J.D. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins (1996). *Cell*, 84, 451–459.

Dominique R. VIGS vectors for gene silencing: many targets, many tools (2004). *Annual Review Plant Biology*, 55, 495-519.

Edgar R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.

Fiers M., Ku K. L., and Liu C. M. CLE peptide ligands and their roles in establishing meristems (2007). *Curr Opin Plant Biol.*, 10, 39–43.

Finn R. D., Clements J. et al. HMMER web server: 2015 update (2015). *Nucleic Acids Res.*, 43, 30-38.

Finn R.D., Coghill P., Eberhardt R.Y., et al. The Pfam protein families database: towards a more sustainable future (2016). *Nucleic Acids Research*. 44, D279-D285.

Fritz-Laylin L.K., Krishnamurthy N., Tör M., Sjölander K.V., Jones J.D. Phylogenomic analysis of the receptor-like proteins of rice and arabidopsis (2005). *Plant Physiol.* 138, 611–623.

Fürst U., Hegenauer V., Kaiser B., Körner M., Welz M. and Albert M. Parasitic *Cuscuta* factor(s) and the detection by tomato initiates plant defense (2016). *Communicative & Integrative Biology*, 9, e1244590.

Gust A.A., and Felix G. Receptor like proteins associate with SOBIR1-type of adaptors to form bimolecular receptor kinases (2014). *Current Opinion in Plant Biology*, 21, 104-111.

Hok S., Danchin E.G.J., Allasia V., Panabières F., Attard A., Keller H. An arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease (2011). *Plant Cell Environ.* 34, 1944–1957.

Hoy R. R. Evolution Convergent evolution of hearing (2012). *Science*, 338, 894–895.

Hunter S. et al. InterPro: The integrative protein signature database (2009). *Nucleic Acids Research*, 37, D211–D215.

Jehle A.K., Lipshis M., Albert M., Fallahzadeh-Managhani V., Furst U., Mueller K., Felix G. The receptor-like protein ReMAX of Arabidopsis detects the microbe-associated molecular pattern eMAX from Xanthomonas (2013). *Plant Cell*, 25, 2330-2340.

Jiang Z., Ge S., Xing L., et al. *RLP1.1*, a novel wheat receptor-like protein gene, is involved in the defence response against *Puccinia striiformis* f. sp. *tritici* (2013). *Journal of Experimental Botany*. 64, 3735-3746.

Jones D.A., Jones J.D.G. The roles of leucine-rich repeat proteins in plant defences (1997). *Advances in Botanical Research*, 24, 89–167.

Joosten M.H., Vogelsang R., Cozijnsen T.J., Verberne M.C., De Wit P.J. The biotrophic fungus *Cladosporium fulvum* circumvents Cf-4-mediated resistance by producing unstable AVR4 elicitors (1997). *The Plant Cell*, 9, 367–379.

Katoh K., and Standley D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability (2013). *Molecular Biology and Evolution*, 30, 4, 772-780.

Kozłowski L.P. IPC – Isoelectric point calculator (2016). *Biology Direct*, 11,55.

Kruijt M., De Kock M.J., and De Wit P.J. Receptor-like proteins involved in plant disease resistance (2005). *Mol. Plant. Pathol.*, 6,85-97.

Larkan, N. J., Lydiate, D. J., Parkin, I. A. P., Nelson, M. N., Epp, D. J., Cowling, W. A., Rimmer, S. R. and Borhan, M. H. The Brassica napus blackleg resistance gene LepR3 encodes a receptor-like protein triggered by the Leptosphaeria maculans effector AVRLM1. *New Phytol.*, 197, 595–605.

Letunic I., Doerks T., and Bork P. SMART: recent updates, new developments and status in 2015 (2015). *Nucleic Acids Res.*, 43, 257–260.

Li F. et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution (2015). *Nat. Biotechnol.*, 33, 524–530.

Li L., Stoeckert C.R, Jr., and Roos D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes (2003). *Genome Res.* 13, 2178-2189.

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G, et al. Genome sequence of the cultivated cotton *Gossypium arboreum* (2014). *Nature Genetics*, 46, 567-572.

Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I., and Baulcombe, D.C. Virus-induced gene silencing in plants (2003). *Methods.* 30, 296–303.

Magalhães D.M., Scholte L.L. S., Silva N.V., Oliveira G.C., Zipfel C., Takita M.A. and De Souza A.A. LRR-RLK family from two Citrus species: genome-wide identification and evolutionary aspects (2016). *BMC Genomics*, 17, 623.

Marchler-Bauer A., Derbyshire M.K., Gonzales N.R., et al. CDD: NCBI’s conserved domain database (2015). *Nucleic Acids Research.* 43, D222-D226.

Mendel M.A., Feldmann K.A., Herrera-Estrella L., Rocha-Sosa M., Leon P. CLA1, a novel gene required for chloroplast development, is highly conserved in evolution (1996). *Plant Journal*, 9, 649-658.

Offord V., Coffey T.J., and Werling D. LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database (2010). *Developmental and Comparative Immunology*, 34, 1035-1041.

Panter S.N., Hammond-Kosack K.E., Harrison K., Jones J.D., Jones D.A. Developmental control of promoter activity is not responsible for mature onset of Cf-9B-mediated resistance to leaf mold in tomato (2002). *Molecular Plant Microbe Interact*, 15, 1099–1107.

Paterson A. H., Wendel J.F., Gundlach H., Guo H., Jenkins, J., Jin, D., et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibers (2012). *Nature*, 492, 423-427.

Petersen T.N., Brunak S., Heijine G.V., Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions (2011). *Nature Methods*, 8, 785-786.

Petre B., Hacquard S., Duplessis S., and Rouhier N. Genome analysis of poplar LRR-RLP gene clusters reveals RISP, a defense-related gene coding a candidate endogenous peptide elicitor (2014). *Front. Plant Sci.*, 5, 111.

Ron M. and Avni A. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato (2004). *Plant Cell*, 16, 1604–1615.

Rong J., Feltus F.A., Liu L., and Paterson A.H. Gene copy number evolution during tetraploid cotton radiation (2010). *Heredity*, 105, 463-472.

Ruiz M.T., Voinnet O., Baulcombe D.C. Initiation and maintenance of virus induced gene silencing (1988). *Plant Cell*, 10, 937-946.

Schallus T., Jaechk C., Fehér K., et al. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein *N*-glycosylation (2008). *Molecular Biology of the Cell*, 19, 3404-3414.

Schwechheimer C., Smith C., Bevan M.W. The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high-level expression (1998). *Plant Mol Biol.*, 36, 195–204.

Senthil-Kumar M., and Mysore K.S. New dimensions for VIGS in plant functional genomics (2012). *Trends in Plant Science*, 16,12, 656-665.

Sonnhammer E.L., von Heijne G., Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences (1998). *Proc Int Conf Intell Syst Mol Biol.*, 6, 175-82.

Soskine M. and Tawfik D.S. Mutational effects and the evolution of new protein functions (2010). *Nature Rev. Genet.* 11, 572–582.

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0 (2013). *Mol Biol Evol.*, 30,2725–2729.

Tor M., Lotze M.T., and Holton N. Receptor-mediated signalling in plants: molecular patterns and programmes (2009). *Journal of Experimental Botany*, 60, 3645-3654.

Uebler S., and Dresselhaus T. Identifying plant cell-surface receptors: combining 'classical' techniques with novel methods (2014). *Biochemical Society transactions*, 42, 2, 395-400.

Wang G, Ellendorff U, Kemp B, et al. A Genome-wide functional investigation into the roles of receptor-like proteins in Arabidopsis (2008). *Plant Physiology*. 147,2, 503-517.

Wang G., Fiers M., Ellendorff U., Wang Z., et al. The diverse roles of extracellular leucine-rich repeat-containing receptor-like proteins in plants (2010). *Crit. Rev. Plant Sci.*, 29, 285–299.

Wu J., Liu Z., Zhang Z., Lv Y., Yang N., Zhang G. et al. Transcriptional regulation of receptor-like protein genes by environmental stresses and hormones and their overexpression activities in *Arabidopsis thaliana* (2016). *J. Exp. Bot.*, 67, 3339–3351.

Xiao S., Brown S., Patrick E., Brearley C., and Turner J.G. Enhanced transcription of the *Arabidopsis* disease resistance genes RPW8.1 and RPW8.2 via a salicylic acid-dependent amplification Circuit is required for hypersensitive cell death (2003). *The Plant Cell*, 15, 33–45.

Xiquan G. et al. Silencing GhNDR1 and GhMKK2 compromises cotton resistance to *Verticillium* wilt (2011). *Plant Journal*, 66, 293–305.

Zhang H.B, Li Y., Wang B., Chee P.W. Recent advances in cotton genomics (2008). *International Journal of Plant Genomics*. 742304.

Zhang J.Z. and Kumar S. Detection of convergent and parallel evolution at the amino acid sequence level (1997). *Mol. Biol. Evol.*, 14, 527–536.

Zhang Z., Fradin E., De Jonge R., Van Esse H. P., Smit P., Liu C. M. et al. Optimized agroinfiltration and virus-induced gene silencing to study Ve1-mediated *Verticillium* resistance in tobacco (2012). *Molecular Plant Microbe Interact*, 26, 182–190.

Zipfel C. Plant pattern-recognition receptors (2014). *Trends in Immunology*, 35, 345–351.

APPENDIX A

OrthoMCL analysis in Cotton

Group	Group size in database	Accession in NCBI	Name	Definition	Taxon	Description
OG5_127136	141	atha NP_001117591	AT1G73066.1	LRR family protein	<i>A. thaliana</i>	Protein binding
OG5_128518	88	atha NP_001154317	AT1G07390.3	AtRLP1	<i>A. thaliana</i>	Protein binding
OG5_130149	58	atha NP_001154652	AT3G28890.1	AtRLP43	<i>A. thaliana</i>	Protein binding
OG5_134835	26	atha NP_181039	AT2G34930	LRR family protein	<i>A. thaliana</i>	Protein binding
OG5_136779	20	rcom 27985.m000845	N.A	LRR family protein	<i>R. communis</i>	serine-threonine
OG5_159698	7	rcom 27555.m000030	N.A	LRR family protein	<i>R. communis</i>	serine-threonine
OG5_159729	7	atha NP_181808	AT2G42800	AtRLP29	<i>A. thaliana</i>	Protein binding
OG5_164541	6	atha NP_190544	AT3G49750.1	AtRLP44	<i>A. thaliana</i>	Protein binding
OG5_170335	5	atha NP_174156	AT1G28340	AtRLP4	<i>A. thaliana</i>	Protein binding
OG5_177950	4	atha NP_193611	AT4G18760	AtRLP51	<i>A. thaliana</i>	Protein binding
OG5_188400	3	atha NP_176717	AT1G65380	AtRLP10	<i>A. thaliana</i>	Protein binding
OG5_189973	3	atha NP_178125	AT1G80080	AtRLP17/TMM	<i>A. thaliana</i>	Protein binding
OG5_190449	3	osat NP_001058121	N.A	Leucine-rich repeat	<i>O. sativa Japonica</i>	Os06g0627500
OG5_190826	3	atha NP_564237	AT1G25570	Leucine-rich repeat	<i>A. thaliana</i>	Protein binding
OG5_212573	2	atha NP_192331	AT4G04220	AtRLP46	<i>A. thaliana</i>	Protein binding
OG5_244942	2	rcom 30150.m000491	N.A	LRR family protein	<i>R. communis</i>	Protein binding

APPENDIX B

GhRLP6

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
GhRLP6 Clon1	GC AATGCAGATGGTCAGGTTATGGACTTGACTTGAGCARRAAGTATTCTGGAGCAATGATAGTCAAAATAGTCTTTCCGCTTCGGCATCTTCAGCAACTTAATTTGGCTTACACAGTTTCAR													
Consensus	GC AATGCAGATGGTCAGGTTATGGACTTGACTTGAGCARRAAGTATTCTGGAGCAATGATAGTCAAAATAGTCTTTCCGCTTCGGCATCTTCAGCAACTTAATTTGGCTTACACAGTTTCAR													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
GhRLP6 Clon2	ATCCAGTTTCCTTCAGGTTTGAARATTTGGCAATTTGAGGATCTTAATTTGTCAATGACGGCTCAC TGGCAATTTCCGGTGGAGTTTCATACATGACAAAGTTGGTTACTCTCGATTTGTCA													
Consensus	ATCCAGTTTCCTTCAGGTTTGAARATTTGGCAATTTGAGGATCTTAATTTGTCAATGACGGCTCAC TGGCAATTTCCGGTGGAGTTTCATACATGACAAAGTTGGTTACTCTCGATTTGTCA													
	261	270	280	290	300	305								
GhRLP6 Clon1	AAATCATGGCTTCTTGATCTGGATCCTGAACTCAGAAAGCCA													
Consensus	AAATCATGGCTTCTTGATCTGGATCCTGAACTCAGAAAGCCA													

GhRLP7

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
GhRLP7 Clon2	AATTCCTGGGAGGCATCTACAGCTTCGATCTCGCTCAAAATTCAC TCCAGGACCACTCCAGCTTTACACAGTCAATGCTTATTTCTTAATCTCAGACATGAGTTGGTTGGAGAAATACCTCC													
Consensus	AATTCCTGGGAGGCATCTACAGCTTCGATCTCGCTCAAAATTCAC TCCAGGACCACTCCAGCTTTACACAGTCAATGCTTATTTCTTAATCTCAGACATGAGTTGGTTGGAGAAATACCTCC													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
GhRLP7 Clon2	TACTATTGCAATTCAGTTCTCTGTTGACTTGACTTATCCAGAAATAGTTGGTGGGATGTTCCAGACTGTCTTGGCAATTTAGAGCTCCACAGTTATGGATTTGAGACGAAATACCTTTTC													
Consensus	TACTATTGCAATTCAGTTCTCTGTTGACTTGACTTATCCAGAAATAGTTGGTGGGATGTTCCAGACTGTCTTGGCAATTTAGAGCTCCACAGTTATGGATTTGAGACGAAATACCTTTTC													
	261	270	280	290	300	310	320	330						
GhRLP7 Clon2	GGAAARATTCACACACATACGTCAGATATAGTTCATTGAGAACCTTACCTCAATGACACCACTTG													
Consensus	GGAAARATTCACACACATACGTCAGATATAGTTCATTGAGAACCTTACCTCAATGACACCACTTG													

GhRLP9

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
GhRLP9 Clon3	CGTGGGTCGCTTCAAAATCTCAGTCTTGTATTTATATGATGGATACGCGATGGAGGACCAATAGGGTTACCCGCTTTGAAARACTTGGAGAACTACATTTAGGGTGTGGTAAATGAAATGT													
Consensus	CGTGGGTCGCTTCAAAATCTCAGTCTTGTATTTATATGATGGATACGCGATGGAGGACCAATAGGGTTACCCGCTTTGAAARACTTGGAGAACTACATTTAGGGTGTGGTAAATGAAATGT													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
GhRLP9 Clon3	TACCTTCACCTCAATTTGGATATATCCCGTCATTGAGACTCTTCTTAGAAACTTGAGCTTGAGGGGACCAATATCCAAAGATGGCAAAATTTACAAATTTAAAGAAATGACCTTACG													
Consensus	TACCTTCACCTCAATTTGGATATATCCCGTCATTGAGACTCTTCTTAGAAACTTGAGCTTGAGGGGACCAATATCCAAAGATGGCAAAATTTACAAATTTAAAGAAATGACCTTACG													
	261	270	280	290	300	308								
GhRLP9 Clon3	GATTTATCCGACACTTCCACATTTATTCGAGATTTGAAACCTCAGT													
Consensus	GATTTATCCGACACTTCCACATTTATTCGAGATTTGAAACCTCAGT													

GhRLP23_1

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
GhRLP23_1 Clon4	GAGTCGATCGGAACTTCAGTCTTGGATATTTGGACATTAGCCACTGCCATTTTTGGGAGCTACCGTACTCACTTGGTGGCTTACTCACTCAAAATCTTGATTTTTCTATACAACTTTT													
Consensus	GAGTCGATCGGAACTTCAGTCTTGGATATTTGGACATTAGCCACTGCCATTTTTGGGAGCTACCGTACTCACTTGGTGGCTTACTCACTCAAAATCTTGATTTTTCTATACAACTTTT													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
GhRLP23_1 Clon4	CCCAACTATTCCTTCTCAATGGACACTTAATCACTGCATACCTTTGGATTTGCTGATATAAATTTGGGGCAGATCCCTTCTTCTTATCCAACTTACTCAGCTTTTTTATCTGTCCTTGC													
Consensus	CCCAACTATTCCTTCTCAATGGACACTTAATCACTGCATACCTTTGGATTTGCTGATATAAATTTGGGGCAGATCCCTTCTTCTTATCCAACTTACTCAGCTTTTTTATCTGTCCTTGC													
	261	270	280	285										
GhRLP23_1 Clon4	CACCAACAGTTTCGTCAGGGAAAC													
Consensus	CACCAACAGTTTCGTCAGGGAAAC													

GhRNP23_2

1 10 20 30 40 50 60 70 80 90 100 110 120 130
GhRNP23_2
Clon5
CTGGATCTGACGCTCGCTTTGTTGCGGTTGAGGATGCTTCATTCTTARTCGGTCAGCTCTTATGATCTTTAGCTTATCCGARRGTAGTTCATGAACTTGAGGACAGGTCGGTGGGATT
Consensus
CTGGATCTGACGCTCGCTTTGTTGCGGTTGAGGATGCTTCATTCTTARTCGGTCAGCTCTTATGATCTTTAGCTTATCCGARRGTAGTTCATGAACTTGAGGACAGGTCGGTGGGATT
131 140 150 160 170 180 190 200 210 220 230 240 250 260
GhRNP23_2
Clon5
GCTGCTCTGGGATGGARTGAGTGCATGATACACTGGTCACGTTATGGACTTATCTCAGCAGTAGTTCCCTTATGGCTCATCACTCCACACAGCCTTTCCAGCTCCATCACTTAGAG
Consensus
GCTGCTCTGGGATGGARTGAGTGCATGATACACTGGTCACGTTATGGACTTATCTCAGCAGTAGTTCCCTTATGGCTCATCACTCCACACAGCCTTTCCAGCTCCATCACTTAGAG
261 270 280 290 300 310 320 330 341
GhRNP23_2
Clon5
GCTCACCTTTCCGACATGACTTCATCGCTGAAATCCCTTCGCARTTGGTACCTTACAGGCTATCAGGTTGGA
Consensus
GCTCACCTTTCCGACATGACTTCATCGCTGAAATCCCTTCGCARTTGGTACCTTACAGGCTATCAGGTTGGA

GhRNP23_3

1 10 20 30 40 50 60 70 80 90 100 110 120 130
GhRNP23_3
Clon6
TGCTTTCCCTAACCGCTCCCAATACACCACTAGTCTTCCARTGTTAGTGAGCTGGGGTGGGTCGTTGATCTTAGACGGTCCCAACCTTCTTAACTCACCAGACCACTGGCATATCTAGACCT
Consensus
TGCTTTCCCTAACCGCTCCCAATACACCACTAGTCTTCCARTGTTAGTGAGCTGGGGTGGGTCGTTGATCTTAGACGGTCCCAACCTTCTTAACTCACCAGACCACTGGCATATCTAGACCT
131 140 150 160 170 180 190 200 210 220 230 240 250 260
GhRNP23_3
Clon6
TTCTCAAATGATATACATGTTGAGATGCCCAATGGATATGGAAATGAGCTCGTACTCTATTTCTATTGGACCTTCTTTAACTCTCAGCGGATTCATCACTCCCACTCTCCCTTGG
Consensus
TTCTCAAATGATATACATGTTGAGATGCCCAATGGATATGGAAATGAGCTCGTACTCTATTTCTATTGGACCTTCTTTAACTCTCAGCGGATTCATCACTCCCACTCTCCCTTGG
261 270 280 290 300
GhRNP23_3
Clon6
TCTAATATCTCTTTTGGACCTCAGCTCCACCTGTTTC
Consensus
TCTAATATCTCTTTTGGACCTCAGCTCCACCTGTTTC

GhRNP27

1 10 20 30 40 50 60 70 80 90 100 110 120 130
GhRNP27
Clon7
TCCAAAGTGCATCGAARCTGAGTCTCTCTCACAGCAGTTGATCTGGGAAATATATTTTCATGGTCAATACCAAGAAATTTGCTAARAGTTGATGTTGAGAGTTTACGCATCACCAACAT
Consensus
TCCAAAGTGCATCGAARCTGAGTCTCTCTCACAGCAGTTGATCTGGGAAATATATTTTCATGGTCAATACCAAGAAATTTGCTAARAGTTGATGTTGAGAGTTTACGCATCACCAACAT
131 140 150 160 170 180 190 200 210 220 230 240 250 260
GhRNP27
Clon7
GATATGAGGGGCTACTACACGATCTTGGGTAATGCAAGGTTAAATCTCTCGATGTTGGAAACACCAATTTGATGACACTTTCCCAAGTTGGTGGAAATTTGGATCAGCTGCAAGTTCTTA
Consensus
GATATGAGGGGCTACTACACGATCTTGGGTAATGCAAGGTTAAATCTCTCGATGTTGGAAACACCAATTTGATGACACTTTCCCAAGTTGGTGGAAATTTGGATCAGCTGCAAGTTCTTA
261 270 280 290 300 305
GhRNP27
Clon7
TCTTGGGTCGATAGATTTTACGGCCAAATAGAGGCTTTGAT
Consensus
TCTTGGGTCGATAGATTTTACGGCCAAATAGAGGCTTTGAT

GhRNP34

1 10 20 30 40 50 60 70 80 90 100 110 120 130
GhRNP34
Clon8
TTTGGTACACTTGAGTCGTAGATCTTCTTGCACAGCAGTTGTTGGGAGATTCCTCGGAGTGGACTGGTTGATTTCTTGAAGTGTAAACCTGTCCAGAGACCACTCGTTGGGCTCATACCT
Consensus
TTTGGTACACTTGAGTCGTAGATCTTCTTGCACAGCAGTTGTTGGGAGATTCCTCGGAGTGGACTGGTTGATTTCTTGAAGTGTAAACCTGTCCAGAGACCACTCGTTGGGCTCATACCT
131 140 150 160 170 180 190 200 210 220 230 240 250 260
GhRNP34
Clon8
CAGGAAACCAATTCACACCTTTCTCACGATCTCTACGACGCAATATAGGACTGTGGGTTCCCGTTTCAAAAGCTGCGGCCGTAGTGAACACCACTCAGCAATTTTGTGAGAGGAGTTG
Consensus
CAGGAAACCAATTCACACCTTTCTCACGATCTCTACGACGCAATATAGGACTGTGGGTTCCCGTTTCAAAAGCTGCGGCCGTAGTGAACACCACTCAGCAATTTTGTGAGAGGAGTTG
261 270 280 290 300 310 315
GhRNP34
Clon8
ATTCTGCATTTGGATTGGATGGAAATTTGATGATGGTTATGGGTTGGAGA
Consensus
ATTCTGCATTTGGATTGGATGGAAATTTGATGATGGTTATGGGTTGGAGA

GhRNP44

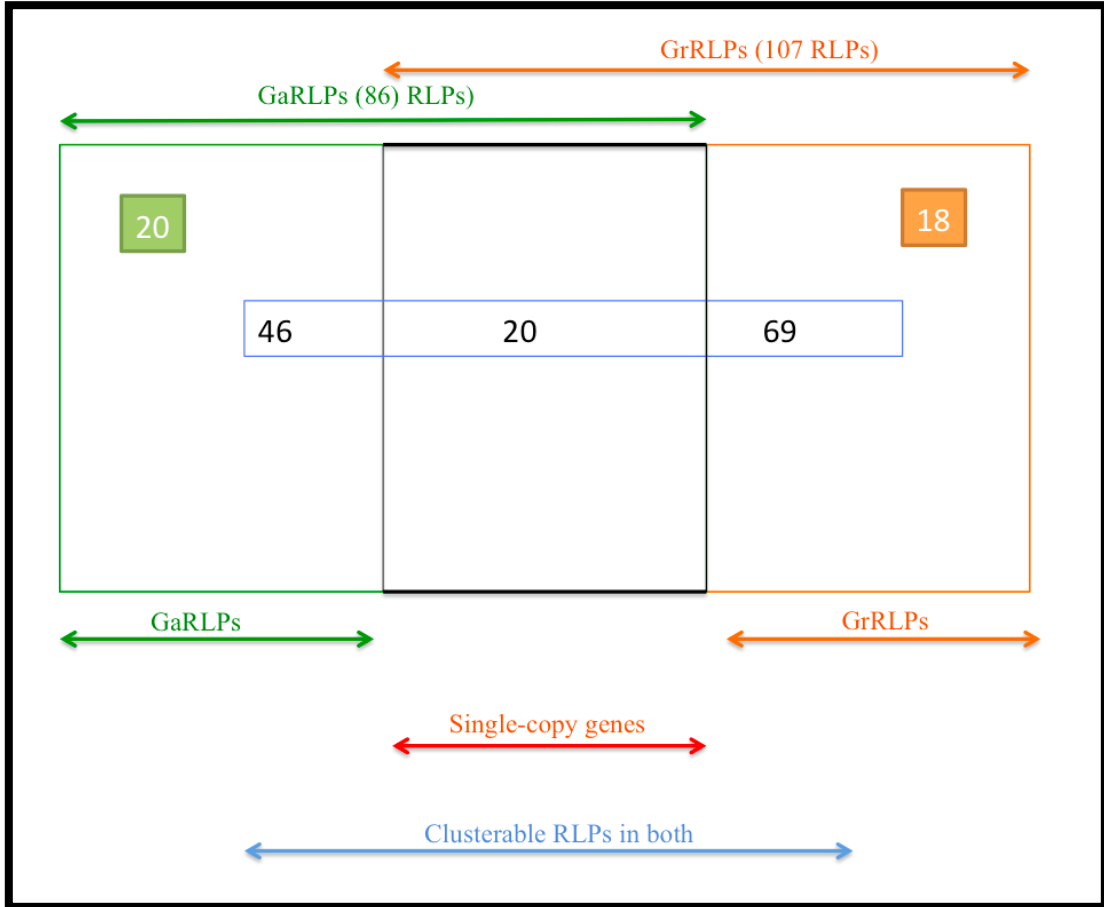
1 10 20 30 40 50 60 70 80 90 100 110 120 130
GhRNP44
Clon9
TTTATGTCACCTGGCAGTCTCATCTCTCATCAACCGCTCGAGGAGAGATCCACCCGAGCTTACCTTGTGCGCTTATCTAARCGTATTGATCTCATGACATTTACTTACCGGTGAGTTCT
Consensus
TTTATGTCACCTGGCAGTCTCATCTCTCATCAACCGCTCGAGGAGAGATCCACCCGAGCTTACCTTGTGCGCTTATCTAARCGTATTGATCTCATGACATTTACTTACCGGTGAGTTCT
131 140 150 160 170 180 190 200 210 220 230 240 250
GhRNP44
Clon9
CAGAAATAGGCTTCTAGCAGCTTGTGCGCATTTGATGTTCTTACACACAGCTTTCAGGGCCAAATACCACTTCTTTAGGGACAGGAGTGGGGCTTTCAGAGGTTCAATGCTACT
Consensus
CAGAAATAGGCTTCTAGCAGCTTGTGCGCATTTGATGTTCTTACACACAGCTTTCAGGGCCAAATACCACTTCTTTAGGGACAGGAGTGGGGCTTTCAGAGGTTCAATGCTACT

GhRLPGS01-like

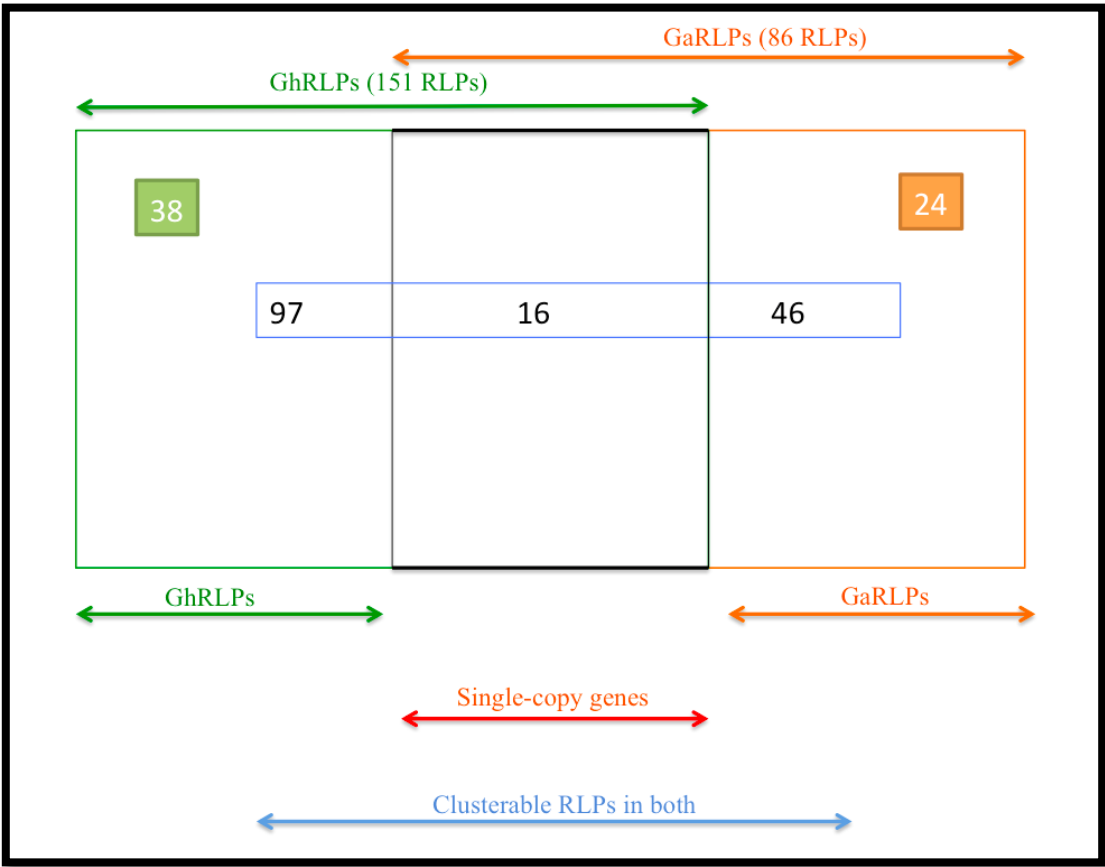
	1	10	20	30	40	50	60	70	80	90	100	110	120	130
GhRLPGS01-like	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
Clon10	ATTTGCGAGCCTCAACACACATGCTTGATGGGGAAATCCCTGCATCAATTGGGAGCTTAAATCGCTACAAATCTGAACTTGGCCAAATACAGCCTTTCAGGATCCATCCCTCCAGAGATAAGTGGCCT													
Consensus	AaTTcGCGAGCCTCAACACACATGCTTGATGGGGAAATCCCTGCATCAATTGGGAGCTTAAATCGCTACAAATCTGAACTTGGCCAAATACAGCCTTTCAGGATCCATCCCTCCAGAGATAAGTGGCCT													
	131	140	150	160	170	180	190	200	210	220	230	240	250	260
GhRLPGS01-like	----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----													
Clon10	ATCCAAATTTGCAGTACTTGAAATTTATTAGGAAATAGGCTGAATGGAGAGATCCATGGGAGCTCAACCAATTTGGTTCAGCTCCAGAGCTTGACTTATCCAGTACACACCTATCAGGAAACATAAACCTG													
Consensus	ATCCAAATTTGCAGTACTTGAAATTTATTAGGAAATAGGCTGAATGGAGAGATCCATGGGAGCTCAACCAATTTGGTTCAGCTCCAGAGCTTGACTTATCCAGTACACACCTATCAGGAAACATAAACCTG													
	261	270	280	290	300	310	320	330	340					
GhRLPGS01-like	----- ----- ----- ----- ----- ----- ----- ----- -----													
Clon10	CTTAAATCCATTTGCAGACCTTCAGTTGTAGTTTTCTCCACACATGCTTTGACGGGTAGCATCCAGT													
Consensus	CTTAAATCCATTTGCAGACCTTCAGTTGTAGTTTTCTCCACACATGCTTTGACGGGTAGCATCCAGT													

APPENDIX C

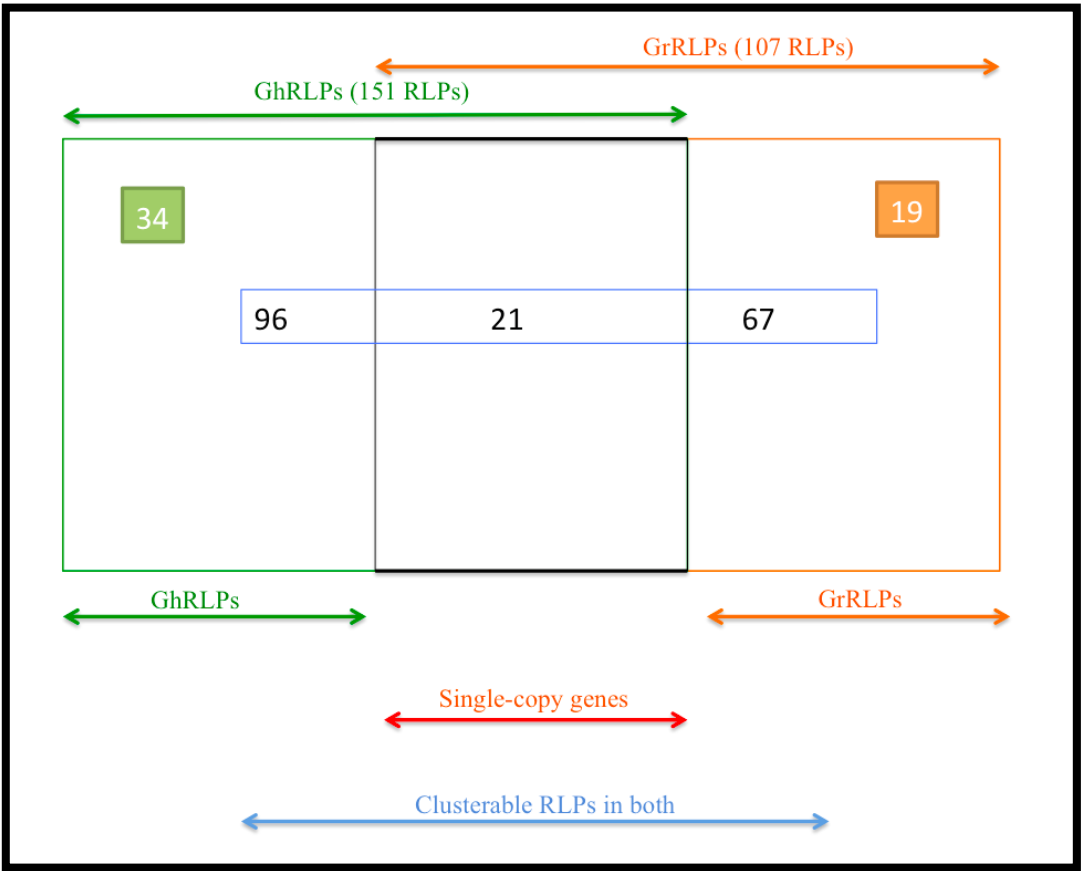
COMPARISON BETWEEN GaRLPs and GrRLPs



COMPARISON BETWEEN GhRLPs and GaRLPs

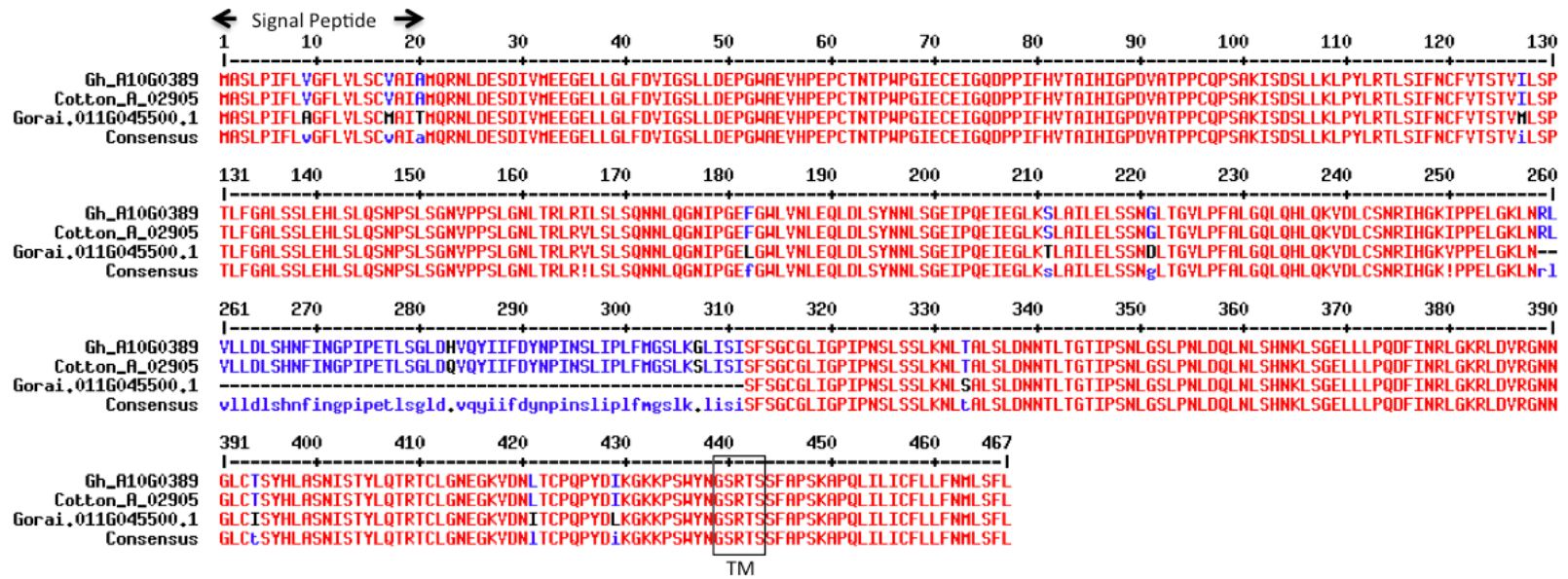


COMPARISON BETWEEN GhRLPs and GrRLPs



APPENDIX D

An Example of the 3 single-copy RLPs in cotton species



APPENDIX E

```

# Aligned_sequences: 2
# 1: AtRLP23
# 2: Gh_D07G1953
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1058
# Identity:   368/1058 (34.8%)
# Similarity: 518/1058 (49.0%)
# Gaps:      210/1058 (19.8%)
# Score: 1404.0
#
AtRLP23          1  -----MSKALLHLHFLSLFLLCCVCHSSIFTLNFHFTGIVACRPHQI      42
                  :|.|||. | .||. ||..|| ||:           :.:|...:
Gh_D07G1953     1  MGQLLQHFLKCLLFL-ILSHFLTQCV--SSV-----LRSCLDSE      38

AtRLP23          43  QAFTKFTNEFDTRGCNNSDT-----FNGVWCDNS      71
                  .|...:..|. :...:..||           :.:|...:
Gh_D07G1953     39  SALLRLKDGFIILNRSASYDTLAYPKVSSWKLEGQVGGDCCSWDGI ECDN      88

AtRLP23          72  TGAVAVLQLRKC-LSGTLKSNSSSLFGFHQLRYVDLQNNLTSSSLPSGFG      120
                  ||...|. |... |. :.:||:| ||...||. :.:|...:|...:|...:|
Gh_D07G1953     89  TGHVIGLDLSSSFLYGSINSNNSLFQLHHLRRLNLSDDNFRSEIPSAIG      138

AtRLP23          121  NLKRLEGLFLSSNGFLGQVPSSFSNLTMLAQLDLSYNKLTGSFPLVRGLR      170
                  ||...|. |...:|...:|...:|...:|...:|...:|...:|...:|...:|
Gh_D07G1953     139  NLRLSRLDLSSSGFTGQIPYEVLLLSNLVILDLSRNRLELRNPSLKSLS      188

AtRLP23          171  K----LIVLDLSYNHFSGTLNPN-----SLF      193
                  : |.:|||...:|. :| |. |           .:|
Gh_D07G1953     189  ERLNLKILDLDMVNVSSTI-PQSLANLSALTYLSLDACELHGSFPIEVF      237

AtRLP23          194  ELHQLRYLNLA FNFS-SSLPSKFGNLHRLLENLILSSNGFSGQVPSTISN      242
                  :||...|. :|...:|...:|...:|...:|...:|...:|...:|...:|...:
Gh_D07G1953     238  QLHHLRRLSLSLSDNDFNGSEIPSAIANLRLSRLDLSSSGFTGQIPYEV LQ      287

AtRLP23          243  LTRLTKLYLDQNKLTSSFP---LVQNLTNLYELDLSYNKFFGVIPSSLL      288
                  |:..:|. |...|. |...:|...:|...:|...:|...:|...:|...:|...:
Gh_D07G1953     288  LSNLVRDLDSGNPLELRNPSLKSLSERLLNLKILDLDMVNVSSTIPQSLA      337

AtRLP23          289  TLPFLAHLALRENNLAGSVEVSNS-----      312
                  .|...:|...:|...:|...:|...:|...:
Gh_D07G1953     338  NLSALTYLSLEDCELHGTFP TSLANLTQLTYLSLAGNEFSPATLPWLTQL      387

AtRLP23          313  -----STSSRLEI-----      320
                  ||:|...:
Gh_D07G1953     388  TKLTALNLDSTNSYGEVLSYLKNLTKLTYLTLTRNQFSERIPSWFGNLTG      437

AtRLP23          321  ---MYLGSNHFEQGILEPISKLINLKHLDLSFLNTSYPIDLKLFS LKSL      367
                  :. ||||. |...|. :...:|...:|...:|...:|...:|...:|...:|...:
Gh_D07G1953     438  LNTLALGSNEFWGSI PKSI FTLKNLVDLDLYGNHLSGTYKLESFLNLKKL      487

AtRLP23          368  RSLDLSGNSISSASLSSDSYIPLTL---EMLTLRHCDINEFPN ILKTLKE      414
                  :.:|...|. |...:|...:|...:|...:|...:|...:|...:|...:
Gh_D07G1953     488  KTLQLSSNQF---SLLSTTVINVTVPKFTLLTLASCNLSKFP SLLSSQDK      534

AtRLP23          415  LVYIDISNNRMKGKIP EWLWLSL--PLLQSVTLGNNYFTGFQGS AEILVNS      462
                  |...:|...:|...:|...:|...:|...:|...:|...:|...:|...:
Gh_D07G1953     535  LKLLDLGGNKIHGCI PKWIWLSGQTLQVLDLSENFLTGF DQTTVPQWI      584

```

AtRLP23	463	SVLLLLYLDSDNNFEGALPDLPLSIKGFVANSFTSEIPLSICNRSSLAAI	512
Gh_D07G1953	585	NLRELDLGSNKLQGS LPIPPASIYNYFISNNLLKGEISSIICNLTSIAVL	634
AtRLP23	513	DLSYNNFTGP IPPCLRN----LELVYLRNNNLEGSIPDALCDGASLRTLD	558
Gh_D07G1953	635	DLSNNSFSGMLPPCLGNLSKLSVLDLQNNNFSGPIPRACEKGSALRMID	684
AtRLP23	559	VSHNRLTGKLP RSFVNCSSLKFLSVINNRIEDTFPFWLKALPNLQVLTLR	608
Gh_D07G1953	685	LSQNQLNGRIPRSLVNCNMLGFLNLGNNQIEDTSPSWLGRPELRILILR	734
AtRLP23	609	SNRFYGFISPPHQGPLGFPELRIFEISDNKFTGSLPPNYFVNWKASRTM	658
Gh_D07G1953	735	HNGFHGAIGEPKSNE--FPKLRILDLSFNKLTGCLRSRHFQRWKAM----	778
AtRLP23	659	NQDGGLYMVYEEKLFDEGGYTDALDLQYK-----GLHMEQA	696
Gh_D07G1953	779	-----KVV DIGKLRYLEA-DISFKAGERSVWHDFSYSMTMTKA	815
AtRLP23	697	KALTSY-----AAIDFSGNRLEGQIPESIGLLKALIAVNISNNAFTGH	739
Gh_D07G1953	816	GVETKYERIQDILVAIDLSSNKFDGCIPEDIQMLKALQFLNLSNNFLSGP	865
AtRLP23	740	IPLSMANLENLES LDMSRNQLSGTIPNGLGSISFLAYINVSHNQLTGEIP	789
Gh_D07G1953	866	IPSSLANLSNLQALDLSRNKLSGEIQELVQLTFLGFFNVSHNQLTGPPI	915
AtRLP23	790	QGTQITGQSKSSFEGNAGLCGLPLKESCF--GTGAPPMYHQKQEDKEEEE	837
Gh_D07G1953	916	QGKQFGTFENNSFEDNLGLCGNPLSKKCYPEGLSPPPPSLSKKEDGEDSW	965
AtRLP23	838	EEEEEEVLNGRAVAIGYGSGLLLGLAIAQVIASYPKPEWLVKIIG---L	884
Gh_D07G1953	966	LQ-----FGWKAIMLGYSGVVNGLVLG YLFNPMKHKLFVKYFGRKMQ	1008
AtRLP23	885	NKRRKR-- 890	
Gh_D07G1953	1009	NRRRGRMN 1016	

```

# Aligned_sequences: 2
# 1: AtRLP23
# 2: Gh_D09G2477
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1083
# Identity: 357/1083 (33.0%)
# Similarity: 498/1083 (46.0%)
# Gaps: 223/1083 (20.6%)
# Score: 1344.0
#
#
#=====
AtRLP23      1 -----MSKALLHLHF---      10
                .:||||.|
Gh_D09G2477  1 MGRARRGKTTSYIREKIDQQVYFGSKCMKVMHYDAGLSTKALLCLEFSCI      50

AtRLP23     11 -----LSLFLCCVCHSSIFTLNFHFTGIVA-----CRPHQIQ      43
                |:|.||. |.....|.|.: |.....
Gh_D09G2477  51 LIDKEDEIMGSSLALTLV-----SLLICLFQFTRCVSSVQFLCRSDERL      95

AtRLP23     44 AFTKFTNEF--DTRGCNNSDTF-----NGVWCDNSTGAV      75
                |...|.... |.....|.: :|:|:|:|.|
Gh_D09G2477  96 ALLLFKESLAVDKEASANPFAYPKADGWKFQGVDCCYWDGIECDHNTGHV      145

AtRLP23     76 AVLQL-RKCLSGTLKSNSSLFGFHQLRYVDLQNNNLTSSSLPSGFGNLKR      124
                ..|.| .|||.|:|.|||.....|:|.|:|.|||..|
Gh_D09G2477  146 TALDLSSSCLYGSINSTSSLFHLHLRKLNLADNDFNSSLPSRLGNLSM      195

AtRLP23    125 LEGLFLSSNGFLGQVPSSFS-----      144
                |...|:|:|.||:|...|
Gh_D09G2477  196 LTYLNLSTSSFSGQIPIEISWLSRSLDLSNTMGLEFGMFSHGGLKLER      245

AtRLP23    145 -----NLTMLAQLDLSYNKLTGS      162
                ||:|:|:|:|:|:|:|
Gh_D09G2477  246 PDFKSLIQNLTSKHLHLRHVAISSVPVNIIVNLSSLSSLDLSYCGLLGK      295

AtRLP23    163 FPL-VRGLRKLIVLDLSYN-HFSGTLNPNSSLFELHQLRYLNLAFFNFSS      210
                .|. :...:|:|:|:|. |.||| |.....|:|:|:|:|:|
Gh_D09G2477  296 LPTSIFHLPNIQFLDVSNNLHLSGTL---PATLSCRRLKFLSIRFTSLSG      342

AtRLP23    211 SLPSKFGNLHRL-----NLISSNGFSGQV      236
                .|:|:|:|:|:|. |. |:|:|:|:|:|
Gh_D09G2477  343 VLPASIGNLHSLLELLDVLTACKFRGPLPSSLGNLTNLTELALLNNSFSGDI      392

AtRLP23    237 PSTISNLRLTKLYLDQNKLT-SSFP-LVQNLNLYELDLSYNKFFGVIP      284
                ||:|:|:|:|:|. |. |...|:|:|:|. |:|:|:|:|
Gh_D09G2477  393 PSSLSNLTRMGFLSLGINNFNPSSIPSWFANLQLHELHLPFCEGITGPIP      442

AtRLP23    285 SLLTLPLFLAHLALRENNLAGSVEVSNSSTSSRLEIMYLGSNHFEGQILE      334
                |...|. |. |:|:|. |. |.....|:|:|:|:|:|
Gh_D09G2477  443 SFFANLTLAVLDLKRNLQTLTGRFPVGITNL-TQLESLSLGSNMMDGALPD      491

AtRLP23    335 PISKLINLKHLDLSFLNNTSYPIDLKLFSKSLRSLDLSGNSISSASLSS      384
                .|. |. |:|:|. |:|:|. |:|:|. |:|:|. |:|:|. |:|:|. |:|:|. |
Gh_D09G2477  492 SIFGLENLQILEIYSNRLSGIVEMDQFVRLKYL SVLYLSSNNLTLSLTS      541

AtRLP23    385 ---DSYIPLTLEMLTLRHCDINEFPNILKTLKELVYIDISNRMKGKPIE      431
                .:|:|:|. | |. |:|:|. |:|:|. |:|:|. |:|:|. |:|:|. |:|:|. |
Gh_D09G2477  542 PNTTSLPMPFSE-LGLGSCNLRFPNFLT HQNQLAYLDLSQNDIHGEMPK      590

AtRLP23    432 WLW--SLPLLQSVTLGNNYFTGFQGSAEILVNSSVLLLYLDSNNFEGALP      479
                |:| |...|.....|.....|:|:|:|. |. |. |:|:|. |:|:|. |
Gh_D09G2477  591 WIWEMSFDTFLLDLDFNSLTGFHQSPDLLPWSNILFLDLSNLFQGSPP      640

```

AtRLP23	480	DLPLSIKGFVSNSTSEIPLSICNRSSLAAIDLSYNNFTGPIPPCL--	527
Gh_D09G2477	641	IPSSSIMVYLASNNSFTGEIPQLFCSLGSVRVLDLSRNNLGGIIPQCLSK	690
AtRLP23	528	--RNLELVYLRNNNLEGSIPDALCDGASLRTLDSVSHNRLTGKLPFSVNC	575
Gh_D09G2477	691	ASKSLSVLNLNVNNFHGPVPQAWMNGSKLKMINGKKNLTKGLPRSMARC	740
AtRLP23	576	SSLKFLSVINNRIEDTFPFWLKALPNLQVLTLSNRFGPIPPHQQPLG	625
Gh_D09G2477	741	RMLEFLDIGNNQIRDTFPFWLESLEPKLKILILRSNRFHGEIKSREFNSV-	789
AtRLP23	626	FPFLRIFEISDNKFTGSLPPNYFVNWKASSRTMNQDGGLYMVYEEKLF--	673
Gh_D09G2477	790	FPKLRVIDISNNGFIGSLPSSYLESWIAMKRFHVE----HLSYMQSSFYD	835
AtRLP23	674	-----DEGGYGYTDALDLQYKGLHMEQAKALTSYAAIDFSGNRLE	713
Gh_D09G2477	836	HMFAGLTIIPDEYNYSM-----LTNKGIKMEYTKILEVFMAVDLSCNKFS	881
AtRLP23	714	GQIPESIGLLKALIAVNISNAFTGHIPLSMANLENLESLSMRNQLSGT	763
Gh_D09G2477	882	GEIPESIGNLKGLELLNLSNNILVGQIPTVIGTLTNLEALDLSHNQLFGR	931
AtRLP23	764	IPNGLGISFLAYINVSHNQLTGEIPQGTQITGQSKSSFEGNAGLCGLPL	813
Gh_D09G2477	932	IPWQLRQLNFLEVFVSYNHLSGPIPQGRQFGTFPNSSSFDGNLGLCGNPL	981
AtRLP23	814	KESCFGTGAPPMYHQKQEDKEEEEEEEEEEEVLNGRAVAIGYGSGLLLG	863
Gh_D09G2477	982	SKKCEDLEASP-----PPSSSTFEQNHGVGLEWRGVLLGFGTGFLFG	1023
AtRLP23	864	LAIQVIASYPKPEWL-----VKIIGLNKRRKR* 891	
Gh_D09G2477	1024	TGLGWVVVTRNLKWFAITFRIVGRWPKH---- 1052	

```

# Aligned_sequences: 2
# 1: AtRLP23
# 2: Gh_A07G1586
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1061
# Identity:      358/1061 (33.7%)
# Similarity:   503/1061 (47.4%)
# Gaps:         204/1061 (19.2%)
# Score: 1361.0
#
#
#=====
AtRLP23          1 MSKALLHLHLFSLFLLL-CCVCHSSIFTLNHFHFTGIVACRPHQIQAFTKFT      49
                |..| .|...|:|.|| .|| |:..|         |..|...|...|
Gh_A07G1586     1 MGPA-FHLILALLLLFSLVC--SVKPL-----CHPDERSALLHFK      38

AtRLP23          50 NEF-----DTRGCNNSD-----TFNGVWCDNSTGAVAVLQ      79
                ..| .|:..|:|:|:|...|...| .:|...|...|...|...|
Gh_A07G1586     39 QSFIIKDSASSSPDAYPKTESWNVEDPSVDCSSWNGVECDNITGHVVVLE      88

AtRLP23          80 LRKC-LSGTLKSNSSLFGFHQLRYVDLQNNLTSSSLPSGFGNLKRLEGL     128
                |... |...|:|:|:|...|:|:~|:|:|:|:|:|:|:|:|:|...|...|
Gh_A07G1586     89 LNSSYLYGSINSNTTLFRLPHLQRLSLADNVFINSEIPSGINNLSSLTYL     138

AtRLP23          129 FLSSNGFLGQVPSSFNLTMLAQLDLSYN-----                157
                .|...|:|:|...|...|...|...|...|...|...|...|...|
Gh_A07G1586     139 DLSFSNFSGQVPLEILELAKLELLDLSGNPLKLRKPLRSLHNLTLNRQ     188

AtRLP23          158 -----                157

Gh_A07G1586     189 LYLTDVTLSSVPKMLTNFYSLTVLILSNCDLRGKFPPTAVFELPNLELLS     238

AtRLP23          158 -----KLTGSFPLVR-----GLRKLIVLDLS      178
                :|:|:|...|:|:~                ...|...|:|:|
Gh_A07G1586     239 LESNQELSGSLPDVQENHSLSKLSLANTRFSGQLPESIGNFKSLEYLDIS     288

AtRLP23          179 YNHFSGTLNPNSSLFELHQLRYLNLAFFNNFSSSLPSKFGNLHRLENLILS     228
                :|...|...| .|...|:|:|:|...|:|:~|:|:|:|...|...|:|:~|...|
Gh_A07G1586     289 HCHFFGKL--PYSLGGLTQLKYLDIFSNNFQQPIPSIGHLNQLHTLDLS     336

AtRLP23          229 SNGFSGQVPSTISNLRLTKLYLDQNK-LTSSFPLVQNLNLYELDLSYN     277
                .|...|:|:|:|:|:|:|...|...|...|...|...|...|...|...|...|
Gh_A07G1586     337 DNKFSGQIPSSLSNLTQLFYLSLATNSFVQGNLSWIGTQTNLTYLDLSKT     386

AtRLP23          278 KFFGVIPSSLLTLPFLAHLALRENNLAGSVEVSNSSTSSRLEIMYLGSNH     327
                ...|...|...|...|...|...|...|...|...|...|...|...|...|:~ .|:~
Gh_A07G1586     387 NLTGQIPSPQLNLTQITWLYLDNGLDQIIPPWIGSLTKLTEIKF-QDNY     435

AtRLP23          328 FEGQILEPISKLINLKHLDLSFLNNTSY---PIDLKFSSSLKSLRSLDLSG     374
                ..|...|...| .|:~|:|:|:|:|:~|:~|:~|...|...|:|...|...|
Gh_A07G1586     436 LGGPIPESI---FNLQNLELLYLHTNHLNGTLKQLQSFLELKTTLRQLSG     482

AtRLP23          375 NSISSASLSSDSYIPLTLEMLTLRHCDINEFPNILKTLKELVYIDISNNR     424
                |...|...|...|...|...|:~|...|...|:~|:~|:~|...|...|...|:~|:~
Gh_A07G1586     483 NYLSLLTNVSNVAPPKLLGLASCNLSEFPSPFLRSQDELEILELAENK     532

AtRLP23          425 MKGKIPFWLWLS--PLLQSVTLGNNYFTG--FQGSABILVNSSVLLLYLD     470
                :~|:~|...|:~|:~|...|...|...|...|...|...|...|...|...|:~
Gh_A07G1586     533 IHGQIPNFWFVGKQNLQYLNLFNSLTGYIFQKLPVLPWSNLKDFNLE     582

AtRLP23          471 SNNFEGALPDLPLSIKGFVANSFTSEIPLSICNRSSLAIDLSYNNFT     520
                ||...|:~|...|...|:~|:~|:~|...|...|...|...|...|...|...|
Gh_A07G1586     583 SNMLQGSPLPHPPSIRSRYKVSNNMLSGEIEPMFCNLPVSLVLDLSNNMNT     632

```

AtRLP23	521	GPIPPCLRN----LELVYLRNNNLEGSIPDALCDGASLRTLDVSHNRLTG	566
Gh_A07G1586	633	GTLPCLANLTNSLEVLSQLSNHFNSAIPPTYTKNCRLLTMDLSQNLQGG	682
AtRLP23	567	KLPRSFVNCSSLKFLSVINNRIEDTFPFWLKALPNLQVLTLSNRFGPI	616
Gh_A07G1586	683	KIPRSLAHCTQLEELILGNLINDSFPHWLGGPKLVILRSNRLHGM	732
AtRLP23	617	SPPHQGPLGFPELRIFEISDNKFTGSLPPNYFVNWKASS-RTMNQDGGLY	665
Gh_A07G1586	733	GKP-QTKSDFSKLQVIDLSNNHLRGKLPDYNFVWVWAMKVHSTNNSFSY	781
AtRLP23	666	MV-----YEEKLFDEGGYGYTDALDLQYKGLHMEQAKALTSYAAIDFS	708
Gh_A07G1586	782	MFANTSFQNRENVVYDY--YNYT--VTLAMKGRDLEYENVPDSISAIIDL	827
AtRLP23	709	GNRLEGQIPESIGLLKALIAVNISNNAFTGHIPLSMANLENLESLEDM	758
Gh_A07G1586	828	SNELEGEIPEAIGELKLRMLNLSNNKLSGRIPLSLGELSNLESLEL	877
AtRLP23	759	QLSGTIPNGLSISFLAYINVSHNQLTGEIPQGTQITGQSKSSFEGNAGL	808
Gh_A07G1586	878	KLWGKIPQQLSKLNFLVFNVSYNKLEGAVPQGAQFNTFNNDSEYEGNSGL	927
AtRLP23	809	CGLPLKESCFGTGAPPMYHQQEDKEEEEEEEEEEEV--NGRAVAIGY	856
Gh_A07G1586	928	CGYPLTETCGNPEVPASTHLGW----DEAEDEGMWSVIKFGWKIVLTGY	972
AtRLP23	857	GSGLLLGLAIAQVIASYPKPEWLVKI IGLNKRKR-----	890
Gh_A07G1586	973	AGGLILGMSLGNFNFAWKYGWFRKVLG----KCVVSNWKGSNWYGF	1017
AtRLP23	891	-----	890
Gh_A07G1586	1018	ISVWKKVSWNY	1028

```

# Aligned_sequences: 2
# 1: AtRLP23
# 2: Gh_A11G2283
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 990
# Identity:      331/990 (33.4%)
# Similarity:   479/990 (48.4%)
# Gaps:         159/990 (16.1%)
# Score: 1257.0
#
#
#=====
AtRLP23      1 MSKALLHLHFLSLFLLCCVCHSSIFTLNFHFTG---IVACRPHQIQAFTK      47
                                           :||. |.|:.  .|. |...:|...
Gh_A11G2283  1 -----MFTA-FSFSSEQPAVLCHSDERLALLQ      26

AtRLP23     48 FTNEF-----DTRGCNNSD--TFNGVWCDNSTGAVAVL      78
...| |.....| :||:|.||..||..|
Gh_A11G2283  27 LKDSFIIDKQALAAGFCAYPKVDSWDSQVSDCCSWDGIETITGVVIGL      76

AtRLP23     79 QL-RKCLSGTLKSNSSLFGFHQLRYVDLQNNLTSSSLPSGFGNLKRLEG    127
.| ..||:|.||:|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  77 DLSSSCLYGSINSTSSLFRLHLQKLNLANNHFNYSLIPYALGNLSMLTY    126

AtRLP23    128 LFLSSNGFLGQVPSSFSNLTMLAQLDLSYNKLTGSFFLVRGLRKLIVLDL    177
|.|||:|.||:|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  127 LNLSSSVFSGQIPSEISKLYRLSSLDFSNNWDTNP-----SQRLLVLE-    169

AtRLP23    178 SYNHFSGTLNPN--SSLFELHQLRYLNLA FNNSLPSKFGNLHRENL      225
.|: |...:|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  170 -----KPDVKSLIQNLTNLKYLNLSYVVMASPIPSVLANLSSLTSL    210

AtRLP23    226 ILSSNGFSGQVPSTISNLTRLTLYLD-----QNKL-----    256
.|...|...|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  211 YLESCGLQGMFPLAIFRLPNLETIWLSHNLDLTGYLLEFNFSNKLKLLAL    260

AtRLP23    257 -TSSF-----PLVQNLTNLYELDLSYNKFFGVIPSSLLTLPFLAHLALRE    300
.:|| ..||:|.||:|...|...|...|...|...|...|...|...|
Gh_A11G2283  261 WNTSFGELPASIEINLSSLEFLGLGHCNFGSVPSTLGNLPNLKFLDLAT    310

AtRLP23    301 NNLAGSVEVSNSSSTSRLEIMYLGSNHFEGQILEPISKLINLKHLDL--S    348
|:..||...:|.||:|...|...|...|...|...|...|...|...|
Gh_A11G2283  311 NSFTGSVPPTLGNL-TKLDTLHDNYFTGFIPSELTNLTQLTALNLLGN    359

AtRLP23    349 FLNTSYP-----IDLKLFSSLKSLRSLDLSGNS      376
.|:|. | .....|...|...|...|...|...|...|...|
Gh_A11G2283  360 MLHGSVPSSISRLEKLFKFFDCDDNRLGGILEMDAFLELKDLYLFLSLNN    409

AtRLP23    377 --ISSASLSSDSYIPLTLEMLTLRHCDINEFPNLIKTLKELVYIDISNNR    424
:|.||:|.||:|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  410 FYLVSPDNSNATSPQAQLVDIGLRHCHLREFPYFLRNQHRQLLDLSSNN    459

AtRLP23    425 MKGKIPWEL--WSLPLLQSVTLGNNYFTGFQGSAEILVNSSVLLLYLDSN    472
:|:|:|:|: |...|...|...|...|...|...|...|...|...|
Gh_A11G2283  460 IKGQIPQWMSKVSVEITLLFLDLSNNSLIGFDDFPLVLPWSKLQYKLDLDFN    509

AtRLP23    473 NFEALPDLPLSIKGFVASNSFTSEIPLSICNRSSLAIDLSYNNFTGP      522
...|:|...|...|...|...|...|...|...|...|...|...|
Gh_A11G2283  510 ILRGSFPVPLSTVFYSISNNSLNGEIQLLCNLSSLSILDVSYNNMSGG      559

AtRLP23    523 IPPCLRN---LELVYLRNNNLEGSIPDALCDGASLRTLDVSHNRLTGKL      568
||.||. | |...:|:|:|...|...|...|...|...|...|...|
Gh_A11G2283  560 IPVCLSNFSKSLLVLKVRSNQLDGPIPSGWATGNRLKMLDLSKNKLQEKI    609

```


AtRLP23	569	PRSFVNCSSLKFLSVINNRIEDTFPPFWLKALPNLQVLTLLRSNRFYGPISP	618
Gh_A11G2283	610	PKSLMECKMLEYLDLGNQIRDAFPPSWLGSPELNLILSSNAFYGRMEN	659
AtRLP23	619	PHQGPLGFPELRIFEISDNKFTGSLPPNYFVNWKASSRTMNQDG----GL	664
Gh_A11G2283	660	PKLNLI VFPKLR IIDL SHNRFNGTLPWGYFERWISMK---NLDGKN SPPK	706
AtRLP23	665	YMVYE-----EKLFDDEGGYGYTDALDLQYKGLHMEQAKALTSYAAIDFSG	709
Gh_A11G2283	707	YMLLESLDMRINLMHVPRDYDY--SMTITNKGMEMKYPKIIIRTLVAIDFSN	754
AtRLP23	710	NRLEGQIPESIGLLKALIAVNI SNNAFTGHIPLSMANLENLESLDMSRNQ	759
Gh_A11G2283	755	NRFDEI PKLIGK LKELHLLNF SNNNLVGGIPVAIAKLTNLESLDLSENK	804
AtRLP23	760	LSGTIPNGLGS-ISFLAYINVSHNQLTGEIPQGTQITGQSKSSFEGNAGL	808
Gh_A11G2283	805	LVGRIPMELSTQLTFLSFLNVSHNRLTGRI PGGGFET FQSSSFDGNLGL	854
AtRLP23	809	CGLPLKESC---FGTGAPPMYHQKQEDKEEEEEEEEEEEVNLGRAVAIG	855
Gh_A11G2283	855	CGKPLLKECSSNSGSLPPPSL-----TSSGEFGLDWKVVLFG	891
AtRLP23	856	YGSGLLGLAIAQVIASYKPEWLVKI IGL--NKRKR--- 890	
Gh_A11G2283	892	YCGS LFGVVI GHVVIK KPDWFANTFSKFPTRRRRRHVL 931	

```

# Aligned_sequences: 2
# 1: AtRLP23
# 2: Gh_D07G1774
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1065
# Identity:      364/1065 (34.2%)
# Similarity:   518/1065 (48.6%)
# Gaps:         213/1065 (20.0%)
# Score: 1346.5
#
#
#=====
AtRLP23          1 MSKA--LLHLHFLSLFLLCC----VCH----SSI--FTLNFHFTGIVACR      38
|. . | | : . | . . | . | . | . | . | : | | | : | . . : | . . . . . : . .
Gh_D07G1774     1 MGSAFHLILLALLLFLVCSVEPLCHPDRSALLHFKQSFIIKHSASSS      50

AtRLP23          39 PHQIQAFTKFTNEFDTRGCNNSD-----TFNGVWCDNSTGAVAVLQLRK      82
| | | : | . | | . . | . . | . . | . . | . . | . . | . . | . . | . . | . .
Gh_D07G1774     51 PH---AYPK-----TEYWNVEDPSVDCCSWNGVECDNITGHVVVLELNS      91

AtRLP23          83 C-LSGTLKSNSSLFGFHQLRYVDLQNNNLTSSSLPSGFGNLRLEGLFSL      131
. | . | : : | | : : | | . . . . | : : . . | . . . : | . : | | . . | . . . | . |
Gh_D07G1774     92 SYLYGSINSNTTLFRLRHLQRLSLADNVFINSEIPSGINNLNLTLYLDLS      141

AtRLP23          132 SNGFLGQVP-----SSFSNLTMLAQLDL      154
. : . . | | | | | . . . . . | . . . . . | . . . . . | . . . . . | . . . . .
Gh_D07G1774     142 FSNFSGQVPLEILELSKLELLELSGNSLKLWKPGLRSLLNLTNLRQLYL      191

AtRLP23          155 -----S      155
|
Gh_D07G1774     192 ADVTISSVPMNLANFYSLTALILSNCDLRGKFP TAVFELPNLECLSLES      241

AtRLP23          156 YNKLTGSFPLVRGLRKLIVLDLSYNHFSGTLNPNS-----      190
. . : | | : | . | . | . . . : | : . | . . . | | | | | . | . |
Gh_D07G1774     242 NQELSGSLPDVQENQSLILRLANTRFSGQL-PESIGNFKSLEYLDISHC      290

AtRLP23          191 -----SLFELHQRLRYLNLAFFNNFSSSLPSKFGNLHRLLENLILSSNGF      232
| | . . | | | : | | : : | | : : | | . . | | . . | : | : | . . | . | . |
Gh_D07G1774     291 HFFGKLPYSLGGLTQLKYLDFSYNNFQPIPSSIGHNLNQLHTLDLSDNKF      340

AtRLP23          233 SGQVPSTISNLTRLTKLYLDQNKLTSSFPLVQ-NL-----TNLYELDLS      275
| | | : | | : : | | | : | . . | . | . | . | . | . | . | . | . | . | . | . |
Gh_D07G1774     341 SGQIPSSLSNLTQLFYLSL----ATNSF--VQGNLSWFGTQTNLTLYLDLS      384

AtRLP23          276 YNKFFGVIPSSLLTLPFLAHLALRENNLAGSVE--VSNSTSSRLEIMYL      323
. . . . | . | . | . | . . . . | . . | . . | . . | . . | . . | . . | . . | . .
Gh_D07G1774     385 KANLTGQIPPSLQNLTKITWLYLYGNLDGQIPPWIGSLTKLTRIKFQ--      432

AtRLP23          324 GSNHFEGQILEPISKLINLKHLDLDFLNTSY---PIDLKLFSLLKSLRSL      370
. | . . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
Gh_D07G1774     433 -DNDLGGPIPESI---FNLQNLLELLYLHTNNLNGTLKQLQSFLELKALTRL      478

AtRLP23          371 DLSGNSISSASLSSDSYIPLTLEMLTLRHCDINEFPNLIKTLKELVYIDI      420
. | | | . : | . . . . | . . | . . | : : | . . | . . | : : | | . : | . : | . . :
Gh_D07G1774     479 QLSGNYLSLLTNVSNVAPPKLLGLASCNLSEFPFSLRSQDELEVLEL      528

AtRLP23          421 SNNRMKGKIPWLWSL--PLLQSVTLGNNYFTG--FQGSAEILVNSSVLL      466
. | : : . | : | . | . : . | . . | . . | | | . . . . | . | : . .
Gh_D07G1774     529 AENKIHGQIPNWFVGWGVKQNLQYLNLFNSLTGYIFQKLPVLPWSNLKD      578

AtRLP23          467 LYLDNNFEGALPDLPLSIKGFVGNVSNFTSEIPLSICNRSSLAIDLSY      516
. . : | | . . : | | . | . | : : . | : : | . . . | | . . . | | . . | | . .
Gh_D07G1774     579 FNLESNMLQGSPLPHPPPSIRS YKVSNNMLSGEIEPFCNLPVLDLNS      628

```

AtRLP23	517	NNFTGPIPPCLR-----LELVYLRNNNLEGSIPDALCDGASLRTLVDVSHN	562
Gh_D07G1774	629	. .: . :::. :: :..:: .	678
AtRLP23	563	RLTGKLPFSFVNCSSLKFLSVINNRIEDTFPFWLKALPNLQVLTTLRSNRF	612
Gh_D07G1774	679	: . : ... :.. :..: : :	728
AtRLP23	613	YGPISPPHQGPLGFPELRIFEISDNKFTGSLPPNYFVNWKASSRTMNQDG	662
Gh_D07G1774	729	: : :..: : :	773
AtRLP23	663	----GLYMVYEEKLFDEGGYGYTD----ALDLQYKGLHMEQAKALTSYAA	704
Gh_D07G1774	774	.. : :.. ... :.. : . : :..:.. :	822
AtRLP23	705	IDFSGNRLEGQIPESIGLLKALIAVNISNNAFTGHIPLSMANLENLES	754
Gh_D07G1774	823	. . : ... : ... : ... : ... : ... : ...	872
AtRLP23	755	MSRNQLSGTIPNGLGSISFLAYINVSHNQLTGEIPQGTQITGQSKSSFEG	804
Gh_D07G1774	873	:: ... : : . : :	922
AtRLP23	805	LSRNKLWGKIPPPQLSKLNFLVFNVSYNKLEGAVPQGAQFNTFNNDSEYEG	852
Gh_D07G1774	923	NAGLCGLPLKESCFTGAPPYHQQEDKEEEEEEEEEEEVL--NGRAV	967
AtRLP23	853	: ... : : . : ... : ...	890
Gh_D07G1774	968	NSGLCGYPLTETCGNPEVPASTHLGW-----DEAEDEGMSVVKFGWKIV	1012
AtRLP23	891	AIGYGSGLLLGLAIAQVIASYPKPEWLVKIIGLNKRRKR-----	
Gh_D07G1774	1013 : :..: ... : ... : .	
AtRLP23	891	----- 890	
Gh_D07G1774	1013	GFSWISVWKKVPWNY 1027	

```

# Aligned_sequences: 2
# 1: AtRLP34
# 2: Gh_D01G0395
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 903
# Identity:      93/903 (10.3%)
# Similarity:   125/903 (13.8%)
# Gaps:         710/903 (78.6%)
# Score: 413.0
#
#
#=====

```

AtRLP34	1	MVGIESHRKTESWGNNSDCCNWEGVTCNAKSGEVIELNLSCSSLHGRFHS	50
Gh_D01G0395	1	-----	0
AtRLP34	51	NSSIRNLHFLTTLDRSHNDFEGQITSSIIENLSHLTSLDLSYNRFSGQILN	100
Gh_D01G0395	1	-----	0
AtRLP34	101	SIGNLSRLTSLDLSFNQFSGQIPSSIGNLSHLTFLGLSGNRFFGQIPSSI	150
Gh_D01G0395	1	-----	0
AtRLP34	151	GNSHLTFLGLSGNRFFGQFPSSIGGLSNLTNLHLSYNKYSGQIPSSIGN	200
Gh_D01G0395	1	-----	0
AtRLP34	201	LSQLIVLYLSVNNFYGEIPSSFGNQLTRLDVSNKLGGNFPNVLLNLT	250
Gh_D01G0395	1	-----	0
AtRLP34	251	GLSVVSLSNKFTGTLPNITSLSNLMAFYASDNAFTGTFPFSLFIIPSL	300
Gh_D01G0395	1	-----	0
AtRLP34	301	TYLGLSGNQLKGTLEFGNISSPSNLQYLNIGSNNFIGPIPSSISKLINLQ	350
Gh_D01G0395	1	-----	0
AtRLP34	351	ELGISHLNTQCRPVDFSIFSHLKSLLDLRLSYLTTTIDLNDILPYFKTL	400
Gh_D01G0395	1	-----	0
AtRLP34	401	RSLDLSGNLVSATNKSSVSSDPPSQSIQSLYLSGCGITDFPEILRTQHEL	450
Gh_D01G0395	1	-----	0
AtRLP34	451	GFLDVSNKIKGQVPGWLWTLPLNLFYLNLSNNTFIGFQRPTKPEPSMAYL	500
Gh_D01G0395	1	-----	0
AtRLP34	501	LGSNNFTGKIPSFICELRSLYTLDLSDNNFSGSIPRCMENLKSNLSELN	550
Gh_D01G0395	1	-----	0
AtRLP34	551	LRQNNLSGGFPEHIFESLRSLDVGHNLVGLKLPRLRFFSNLEVLNVESN	600
Gh_D01G0395	1	-----	0
AtRLP34	601	RINDMFPFWLSSLQKLQVLVLRSNFAFHGPINQALFPKLRIIDISHNHFN	650
Gh_D01G0395	1	-----	0

AtRLP34	651	SLPTEYFVEWSRMSSSLGTYEDGSNVNYLGSGYYQDSMVLNMKGVESELVR	700
Gh_D01G0395	1	-----MKGEDIVLER	10
AtRLP34	701	ILTIYTAVDFSGNKFEIGEIPKSIKGLKELHVLNLSNNAFTGHIPSSIGNL	750
Gh_D01G0395	11	ILTIFFAAIDMSSNKFEGTIPETVGNLISLQVLNFSHNHLTGHIPSSIGNL	60
AtRLP34	751	TALESLDVSNKLYGEIPQEIGNLSLLSYMNFSHNQLTGLVPGGQQLTQ	800
Gh_D01G0395	61	VALESLDLSCNELVGEIPSELTGLNFLEVLNLSENQLVGLIPQGKQFNTF	110
AtRLP34	801	RCSSFEGNLGLFGSSLEEVCRDIHTPASHQQFETPQTEEEDEDL---IS	846
Gh_D01G0395	111	LNDSYAGNIGLCGFPVSKSCGRSEPPAIF-----DEEEVDSAFGLD	152
AtRLP34	847	WIAAAIGFGPGIAFGLMFGYILVSY-KPEWFMNPFGRNN----RRRKRHT	891
Gh_D01G0395	153	WKFVMMGYGCGLVFGFSAGYIMMTIRKPKWLVGMIQRAGNRVLRFRFKYR	202
AtRLP34	892	TTH	894
Gh_D01G0395	203	---	202

```

# Aligned_sequences: 2
# 1: AtRLP27
# 2: Gh_D11G2968
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 828
# Identity:   154/828 (18.6%)
# Similarity: 223/828 (26.9%)
# Gaps:      468/828 (56.5%)
# Score: 676.5
#
#
#=====

```

AtRLP27	1	MLFFIKVFMKTIILSVLLLFFIFASSFTLVVGLAGCRPDQIQALTQFKNEF	50
Gh_D11G2968	1	-----	0
AtRLP27	51	DSSDCNQTDYFNGVQCDNKTGVVTKLQLPSGCLHGSMKPNSSFLGQLHLR	100
Gh_D11G2968	1	-----	0
AtRLP27	101	YLNLSNNNFTSASLPSGFGNLRLEVLVYSSNGFLGQVPSSFSNLSQLNI	150
Gh_D11G2968	1	-----	0
AtRLP27	151	LDLSHNELTGSFPFVQNLTKLSILVLSYNHFSGTIPSSLLTLPFLSSLDL	200
Gh_D11G2968	1	-----	0
AtRLP27	201	RENYLTGSIEAPNSSSTSSRLEFMYLGNNHFEQILEPISKLINLKHLDLS	250
Gh_D11G2968	1	-----	0
AtRLP27	251	FLKTSYPIDLNLFSFVKSLVRLVLSGNSLLATSITSDSKIPLNLENLVLL	300
Gh_D11G2968	1	-----	0
AtRLP27	301	SCGLIEFPPTILKNLTKLEHIDLNNKIKGKVPWFWNLPRLRRVNLFNNL	350
Gh_D11G2968	1	-----	0
AtRLP27	351	FTDLEGSEEVLVNSSVRLLDLAYNHFRGPPFKPPLSINLLSAWNNNSFTGN	400
Gh_D11G2968	1	-----	0
AtRLP27	401	IPLTCNRSSLAILDLSY--NLTGPPIPRCLSDFQESLIVVNLRKNLEG	448
Gh_D11G2968	1	-----MDITLDGNNLEGTIPKICIGNLSSSLTAVDLGNNNFHG	37
AtRLP27	449	SLPDIFSDGALLRTLVDVGYNQLTGKLPSSLNCSMLRFVSDHNKIKDTF	498
Gh_D11G2968	38	QIPENFAKDCMLRSLRINNELEGLSPRSLGNCKGLNLLDVGNNLNDF	87
AtRLP27	499	PFWLKALPDLQALTLRSNKFHGPIPPDRGPLAFPKLRILEISDNNFTGS	548
Gh_D11G2968	88	PSWLGNDQLQVLIILRSNRFYQIESFDI-TVSLTKLRIIDLSHNNFSGY	136
AtRLP27	549	LPPNYFVNWEASSLQMNEDGRIYMGDYNPPYI-----YE	583
Gh_D11G2968	137	LPTLFFEHMHAI-----RDEYGKQVDPHYMREVKIMKDAHKFGVNYA	178
AtRLP27	584	DTVDLQYKGLFMEQGKVLTSYATIDFSGNKLEGQIPESIGLLKALIALNL	633
Gh_D11G2968	179	YGLSVTVKGLERKFESLLSIWTVIDLSSNQFRGEIPKILGELHLLIVLNL	228

AtRLP27	634	SNNAFTGHIPLSLANVTELESLDLSRNQLSGTIPNGLKTLFSLAYISVAH	683
Gh_D11G2968	229	SHNCLIGPIPSSLGNLSELESLDLSSNKLDGRIPTELKNLGFLEVLNLSQ	278
AtRLP27	684	NQLIGEIPQGTQITGQSKSSFEGNAGLCGLPLQGSCFA-PPTPQPKEEDE	732
Gh_D11G2968	279	NNLKGPQPQKQFDFTNDSYIGNLNLYGLPLSKNCGTDEETPAKFDRDD	328
AtRLP27	733	DEEVLNWK-AVVIGYWPGLLLGLIMAHVI-ASFKPKWLKIVGPEKRKED	780
Gh_D11G2968	329	DGDELNWKFSILMGYGCGLVLGMSMAYIVFTTGKPWWLIRIV--ERVRQR	376
AtRLP27	781	NPVRLFMTLDSRWDSFNKKNVEQKSDM	808
Gh_D11G2968	377	FAKR-----	380

```

# Aligned_sequences: 2
# 1: AtRLP7
# 2: Gh_D05G3192
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1084
# Identity:      377/1084 (34.8%)
# Similarity:   552/1084 (50.9%)
# Gaps:        149/1084 (13.7%)
# Score: 1547.5
#
#
#=====
AtRLP7          1 MELYVFFIYVVGKTLHKKIQNTNLLLLFGKLCFFTNIKVFI FSEKMSFLI    50
Gh_D05G3192    1 -----                                0

AtRLP7          51 RSIC---FLILIPSFLITFVSA-----TQHLCHSDQKDALLDFKN    87
   ..|   :|.|:..||.:|:..:   ..|||:..:|..:|.||:|
Gh_D05G3192    1 -MGCWSMWLYLVLFLLFFLFSFFNSPLLSLSLSLPPSHLCRPEQSSSLRFRKS    49

AtRLP7          88 EFGM-----VDSKSWVNKSDCCSWDGITCDAKSGNVIGLDLSS    125
   ...:   ..:||||:..:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    50 NLSISLDKTYPFQNIITFKTKSWDEGTDCCCKWDGVMCDHKEGNVIGLDLSW    99

AtRLP7          126 IFLYGQLKSNSLFLKLRHLRDLNLANNFNNSPIPAEFDKLTGLERLDLS    175
   ..|.|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    100 SGLVGSLSQNSALFSLQNLQWLNLAGNDFDSEIPYEF SKLKS LTYLNLNS    149

AtRLP7          176 QSSLSGQIPINLLQLTKLVSLDLSSDFFGDESFHYLSIDKSFLLP LARN    225
   ..|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    150 GIGLTGFVPPPEISLSELVSLDLGLN-----HLLFRSHDF-NMLVHN    190

AtRLP7          226 LRNLRELDMSYVKISSEIPEEFSNIR-SLRSLNLNGCNLFGEFPSSILLI    274
   |..|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    191 LTKLENMILDMDLSLVVYPYSFLNLTVMKHLSLFACHLQ GKFP AEILQF    240

AtRLP7          275 PNLQSIDLGNNPNLRGNLPVFHENNSLLKLTILYTSFSGAIPDSISSLKN    324
   |.:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    241 PYLENIILQFNVLDLIGHPPEKNWSSPLRVLDV SQTRFSKELPDSIGNLKH    290

AtRLP7          325 LTSLTLSVSYFSGKIPFSLGNLSHLSHLSLSSNNLIGEIPSSIGNLNQLT    374
   |.:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    291 LKTLNLRGCDFMGSIPSTLGNLTKITFLDISGNMFQQQIPDVFGNLNDLS    340

AtRLP7          375 NFYVGGNKLSGNLPATLSNLTKLNTISLSSNQFTGSLPPSISQLSKL KFF    424
   :....|.||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    341 SMDFSHNNFSGFLPSSTFNLTGLTFMDFSSNFLQGTLPNNVRGLPYLREL    390

AtRLP7          425 FADDNPFFIGAILSP LKIPSLTRIHL SYNQLN-----DLV    459
   ..|.|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    391 HLNANLLNGRVPGWLFSLPSLNWLDLNSNKLNGPIDPMQEPNLVQYVDLS    440

AtRLP7          460 GIE-----NIFMLPNLETFYIYHYNYT-KVRPLDLNVFSSLKQLGTLY    501
   ..|   :|.|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    441 ENEIQGPPIPGSFFDLVNLIDLYISSNNLSGNIKSSSELVKLRNLSL DLSH    490

AtRLP7          502 ISRIPISTTN--ITSDFPSNLEYLSLRSCNITDFPEFIRKGRNLQILDLS    549
   .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    491 NSLLSLTSCNNGSNSTLPMVIEF-HFSCSNMQRFP SFLNASKYLQVLDLS    539

AtRLP7          550 NNIKIGQVPDWL---WRMPTLNSVDLSNNSLSGFHVSVKASPESQLTSVD    596
   ||:|.|:|:|:|.   |.   :|:|:|:|:|:|:|:|:|:|:|:|:|
Gh_D05G3192    540 NNQIHGSITKWEAEGWE--SLITLNLNLYMNFMTG----IEQVPGKHL SLD    583

```


AtRLP7	597	LSSNAFQGPLFLPSKSLRYFSGSNNTGKI PRSICGLSSLEILDLSNNN	646
Gh_D05G3192	584	LRNSLQGPALPALQSLLYFLISDNELVGEIPPTICNSSSLVVLDSLKKN	633
AtRLP7	647	LNGSLPWCLETLMSSLSDDLRLNNSLSGSLPEIFMNATKLRSLDVSHNRM	696
Gh_D05G3192	634	LGGIVPDCLGN-FRRLTVMDLRTNYFCGKIPTTYVRYSSLRTLNLNDNQL	682
AtRLP7	697	EGKLPGLTGCSSLEVLNVGSRINDMFPFELNSLQKLQVLVLSNKFHG	746
Gh_D05G3192	683	EGSIPQSLTNCYSLEVLNLDLGNNDINDTFPYWLGTLPLNLQVLVLRNRFHG	732
AtRLP7	747	TLHNVGCVWFQPLQIIVSHNDFFGILPSDYFMNWTAMSSKKNIE-	795
Gh_D05G3192	733	DTQNFNGS-FTFSSLRMIDLQNEFTGHIPPELFEN--LKSMKDIQVDK	778
AtRLP7	796	--PEYIQNPSVYGSSLGYYTSLVLMKGVSMEMERVLTIYTAIDLQGNL	843
Gh_D05G3192	779	SGPKYM-----GEYYYQDSVIVTMKGSDFKLERILTSFTVIDFSSNH	821
AtRLP7	844	HGKIPDSIGLLKELRILNMSNGFTGHIPSSLANLKNLESLDISQNNISG	893
Gh_D05G3192	822	KESIPKELGELNSVIVLNFHNSLAGNIPPSLGKMAALESDDLSSNKLQG	871
AtRLP7	894	EIPPELGTLSLAWINVSHNLVGSIPQGTQFQRQKCSSYEGNPGLNGPS	943
Gh_D05G3192	872	RIPVQLTDLTFLGALNLSNNLEIGHIPLANHFDTFSNDSFDGNSGLCGFP	921
AtRLP7	944	LENVCGHIKESTPTQTEPLETKEEEEEESFSWIAAGLGFAPGVVFGLAMG	993
Gh_D05G3192	922	LSKKCGNDQE----PESPPSTVADESETSI-WKIAAMGYGSGLVGLSMG	966
AtRLP7	994	YIV-VSYKHQWFMKTFGRSKQONTRR----- 1019	
Gh_D05G3192	967	YIVFTTGRPRWLVKMIKRNPKRRRIRRNGRRKN 1000	

```

# Aligned_sequences: 2
# 1: GS01
# 2: Gh_A03G0773
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1278
# Identity:      459/1278 (35.9%)
# Similarity:   615/1278 (48.1%)
# Gaps:         328/1278 (25.7%)
# Score: 2035.0
#
#
#=====

```

GS01	1	-----MQPLVLLLLF---ILCFSGLGQPGIINNDLQTLLEVKKSLVTPNQ	42
		:...	
Gh_A03G0773	1	MDSTQISRFSLFITFFSCILAFS-FGD-NSSSTDAHWLLRIKSKLV----	44
GS01	43	EDDP---LRQWNSDNINYCSWTGVTCDNTGLFRVIALNLTGLGLTGSISP	89
	:	
Gh_A03G0773	45	--DPAGVLESW-SLRAHICSWNGLTCSHDEAF-VIALNLSASTLSGSIPT	90
GS01	90	WFGRFDNLIHLDLSSNNLVGPIPTALSNTLSLESFLFSNQLTGEIPSQL	139
		...:	
Gh_A03G0773	91	ELWNLVSLQTLDSLNLITGSIPPEIGRLRLRTRLLLFANNLSGKVPTEI	140
GS01	140	GSLVNIRSLRIGDNELVGDIPETLGNLVNLQMLALASCRLTGPIPSQLGR	189
		...:	
Gh_A03G0773	141	GLLKKLQVLRIGDNMLAAQIPRSIGNLTELRLVGLAYSQNGSIPAEIGN	190
GS01	190	LVRVQSLILQDNYLEGPIPAELGNCSDLTVFTAENMLNGTIPAEIGRLE	239
		...:	
Gh_A03G0773	191	LKNLVFLDLQHNSLSGFIPAEIHNCEELQNFASNNMLDGEIPASIGKLK	240
GS01	240	NLEILNLANNSLTGEIPSQLGEMSQLQYLSLMANQLQGLIPKSLADLGNL	289
		:	
Gh_A03G0773	241	SLQILNLANNLSGSIPEISGLSNLQYLLGNRLNGEIPWELNHLVQL	290
GS01	290	QTLDLSANNLTGEIPEEFWNMSQLLDLVLANNHLSGSLPKSICSNNNTNLE	339
		
Gh_A03G0773	291	QKLDLSSNNLSGTINLLNIHLQNLQVVFSSNNALTGSIIPSNFCLKNSNLQ	340
GS01	340	QLVLSGTQLSGEIPVELSKCQSLKQLDLSNNLAGSIPPEALFELVELTDL	389
		
Gh_A03G0773	341	QLFLSQNKLSGGFPLGLLNCSSLQQLDLSNNLEGEIPPTIDKLENLTDL	390
GS01	390	YLHNNTEGLTSPSISNLTNLQWLVLHNNLEGLPKPEISALRKLEVLFL	439
		
Gh_A03G0773	391	LLNNNSFNGSLPPEIGNMSNLENLYLFDNMITGSIPEIGKQLSLSIYL	440
GS01	440	YENRFSGEIPQEIGNCTSLKMIDMFGNHFEGEIPPSIGRLKELNLLHLRQ	489
		:	
Gh_A03G0773	441	YDNQMSGSIPELTCASLTADFFGNHFGSIPETIGKLDLVLQLRQ	490
GS01	490	NELVGGPLPASLGNCHQLNILDADNQLSGSIPSSFGFLKGLEQLMLYNNS	539
		:	
Gh_A03G0773	491	NDSLGPPIPPSLGYCKLQQLALADNKLKSGSLPATFRFLSQTITLYNNS	540
GS01	540	LQGNLPDSLISLRNLRINLSHNRNLNGTIHPLCGSSSYLSFDVTNNGFED	589
		.:	
Gh_A03G0773	541	FEGLPLDSLISLRNLRKIINFHNKFSGTILSLTGSNSLTALDNLNNSFSG	590
GS01	590	EIPLELGNLDRRLRLGKNQLTGKIPWTLGKIRELSLLDMSSNALTGTI	639
		
Gh_A03G0773	591	SIPSRLAISTNLGRRLRANLLNGSIPSELGRLKELIFLDLDFNNLTGEV	640

GSO1	640	PLQLVLCKKLTHIDLNNNFLSGPIPPWLGKLSQLGELKLSNQFVESLPT	689
Gh_A03G0773	641	PSQLSTCQKLOHLLLNNDNHFTGKMPSWLGSLODLGELHLSNNFQGOVPA	690
GSO1	690	ELFNCTKLLVLSLDGNSLNGSIPQEIGNLGAENVLNLDKNQFSGSLPQAM	739
Gh_A03G0773	691	ELGNCSRLKLSLHTNNLSGQIPREIGQLTSLNVLNLQRNNLSGFIPPTI	740
GSO1	740	GKLSKLYELRLSRNSLTGEIPVEIGQLQDLQSALDLSYNNFTGDIPSTIG	789
Gh_A03G0773	741	QQCKKLYELRLSENWLSGPIPSEIGKLTQLVILDLSRNLSGEIPSSLG	790
GSO1	790	TLSKLETLDLSDHNLQTEVPGSVGDMKSLGYLNVSFNNLGGKLLKQFSRW	839
Gh_A03G0773	791	NLLKLERLNLNLSFNRLGGELPSSLGQLTSLVTLNLSNNHLRGQLPSTFSGF	840
GSO1	840	PADSFNGTGLCGSPLSRC--NRVRSNNKQQGLSARSVVIISAISALTAI	887
Gh_A03G0773	841	PPTSFSGNDKLCGPPLTSCMDSAGHENNI---LSSSTAVVCIIVAIVFTST	887
GSO1	888	GLMILVIALFFKQRHDFFKVGHGSTAYTSSSSSSQATHKPLFRNGASKS	937
Gh_A03G0773	888	VICLVMIYIMLRMWCNMRKVL-----ISNSSEG-----GGNGIE	921
GSO1	938	DIRWEDIMEATHNLSEEFMIGSGGSKVYKAELENGETVAVKKILWK---	984
Gh_A03G0773	922	QIKRED-----QERWMYGDEKCRK-----GE-----YWRVMS	948
GSO1	985	-----DDLMSNK---SFSREVKTIGRIRHRHLVKLMGYCSSKSEG	1021
Gh_A03G0773	949	SMALVPSQNHDDHISSSCIFQVKMDTKKLG-----	979
GSO1	1022	LNLLIYEYMKNGSIWDWLHEDKPVLEKKKKLLDWEARLRIAVGLAQGVEY	1071
Gh_A03G0773	980	-----	979
GSO1	1072	LHHDCVPPIVHRDIKSSNVLLDSNMEAHLGDFGLAKVLTENCDTNTDSNT	1121
Gh_A03G0773	980	-----	979
GSO1	1122	WFACSYGYIAPEYAYSLKATEKSDVYSMGIVLMEIVTGKMPTDSVFGAEM	1171
Gh_A03G0773	980	-----	979
GSO1	1172	DMVRWVETHLEVAGSARDKLIDPKLKPLLPFEEDAACQVLEIALQCTKTS	1221
Gh_A03G0773	980	-----	979
GSO1	1222	PQERPSSRQACDSSLHVYNNRTAGYKKL	1249
Gh_A03G0773	980	-----	979

```

# Aligned_sequences: 2
# 1: AtRLP6
# 2: Gh_D04G1293
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1121
# Identity:      381/1121 (34.0%)
# Similarity:   553/1121 (49.3%)
# Gaps:        185/1121 (16.5%)
# Score: 1461.5
#
#
#=====
AtRLP6          1 MTGLYSSMSFFLRTIVL---LFSTSSFCNTF---ASLTQDSCHPDQRDAL      44
                :|...|  ||..|...|  |.....|..|:..|
Gh_D04G1293    1 -----MRNSALVSWLFFYSYVATVFRKAVSVSGCQSDQQELL      39

AtRLP6          45 LEFKNEFKIWIYPNGFLDIDGVLMDVTSYPKTKSWTKNSDCCYWDGITCDT      94
                |...|.                :...|..|...|...|...|...|...|...|
Gh_D04G1293    40 LGLKNS-----LNSSSSEKLLKWNQSRDCCPFDGITCNA      73

AtRLP6          95 KSGKVTGLDLSCSCLHGRLEPNSSLFRLQHLQSVNLAYNNFTNSPIPAEF     144
                .|:|.|||||...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    74 -DGQVIGLDLSKELISGAIDKSNLFRRLRHLQQLNLAYNSF-KSKFPSPGF     121

AtRLP6          145 SKFMRLERLNLSRSSFSGHISIKLLQLTNLVSLDLSSSPYPSPSSLSIEK     194
                .....|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    122 ENLANLRYLNLSNAGFTGQIPVEISYMTKLVTLDLKSWLLDLGSLKCLKK     171

AtRLP6          195 P-----LFL-----HLLALNFMNIRELDMSSVD      217
                |                |:|                .|:.....|:..|...
Gh_D04G1293    172 PNLVMLVQNLTRLQNLVLDGINISADRKNWSQALSSSLPDLQVLSMSGCH     221

AtRLP6          218 ISSAI-----PIEFSYMWSLRSITLKGCNLL      243
                :|..|                |...:.....|...|...|...|
Gh_D04G1293    222 LSGPINPSLAKLKSLSIIRLDINNLFGPFKFFAEFQNLTSLHLGGNNLS     271

AtRLP6          244 GRFPNSVLLIPNLESISLDHNLNLEGS LPNFLRNNSLLKLSIYNTSFSGT     293
                |...:..|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    272 GSVPEILQAPKLQTLDLDFNKLLQGSFLWFPPNASLQSLLVGDTNFGGW     321

AtRLP6          294 IPNSISNLKHLTSLKLLQSAFSGRIPSSLRSLSHLSNLVLSENNVEVGEIP     343
                :|...|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    322 LPESISNLGQLTRIELTNCKFNGPLPKTLEKLTQLVYLDVDFSSNNFSGSVP     371

AtRLP6          344 SSVSNLKQLTLFDVSDNNLNGNFP---SLLNLNQLRYIDICSNHFTGF     389
                |...|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    372 -SFTTLKALTYLNLAGNQLNGSILSTNWSSLLNLVSL---DLARNFSFGT     417

AtRLP6          390 LPPTISQLSNLEFFSACDNSFTGSIPLSFLNIS-SLTTGLSYNQLNDTT     438
                :|...|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    418 VPPTLFQSQSLRIIYLPQNQFTGGFSEVKGESLLEAIDLNRNRLQGPF     467

AtRLP6          439 NIKNISLHNLQRLRLDNNNFKASQVDLDVFLSLKRLVSLALSIGIPLSTT     488
                .: .:.....|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    468 PM-FVFEIQGLCELSLSCNKF-SGLITLSAFHKLKN----LSVLDLSYN     510

AtRLP6          489 NITSDSEFSS-----HLEYLELSEGCNIEFPEFIRNQRLSSIDLNS     530
                |:..|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    511 NLSVDSSEFINLPLPPFLPSIAKLLKSSCNLTKFPDFLKNLSILDHLDLSN     560

AtRLP6          531 NNIKGVQPNWLWRPELSTVDLSNNSLIGFNGSLKALSGSKIVMLDLSSN     580
                |...|...|...|...|...|...|...|...|...|...|...|...|...|
Gh_D04G1293    561 NRIHGKIPSWIWKTHLSYLNLSLNFLEAFERTSHIT--SPLVVIDLHGN     608

```


AtRLP9	623	NYLQGVIPSWFGGFFFAYLFLSNNLEGLPSTLFSKPTFKILDLSGNKF	672
Gh_D13G2467	453	-----	452
AtRLP9	673	SGNLPSHFTGMDMSLLYLNDNEFSGTIPSTLIKDVLVLDLRNNKLSGTIP	722
Gh_D13G2467	453	-----	452
AtRLP9	723	HFVKNEFILSLLLRGNTLTGHIPTDLCGLRSIRILDANNRLKGSIPTCL	772
Gh_D13G2467	453	-----	452
AtRLP9	773	NNVSFGRRLNYEVNGDKLPFEINDDEEFVYSRLLVLPQYSPDYTGVLN	822
Gh_D13G2467	453	-----	452
AtRLP9	823	FNVEFASKSRYSYTYQESFNFMFGLDLSSNELSGDIPKELGDLQRIRALN	872
Gh_D13G2467	453	-----	452
AtRLP9	873	LSHNSLSGLIPQSFSNLTDIESIDLFSNLLRGPQPDLKLDYMFVFNVS	922
Gh_D13G2467	453	-----	452
AtRLP9	923	YNNLSGSIPSHGKFSTLDETFIGNLLLCGSAINRSCDDNSTTEFLESDD	972
Gh_D13G2467	453	-----	452
AtRLP9	973	QSGDEETTIDMEIFYWLSLAATYGVWITFIVFLCFDSPWRRVWFHFVDAF	1022
Gh_D13G2467	453	-----	452
AtRLP9	1023	ISLFKCV 1029	
Gh_D13G2467	453	-----	452

APPENDIX F

GhGFP



GhRLP23_1



GhRLP23_2



GhRLP23_3d



GhRLP6



GhRLP44



GhRLPGSO1-like



GhRLP34



GhRLP9



GhRLP7



GhRLP27

