**Original Research Article**

# Insilico Analysis of Rpp1 Locus-Derived NB-ARC-LRR Domains Reveals Insights into Rust Resistance in Soybean


## Abstract

Soybean rust poses substantial economic challenges to soybean producers, leading to a decrease in seed size, weight, oil content, and overall crop yield. This study aimed to provide a comprehensive characterization of the rust resistance locus Rpp1 and its associated genes in comparison to other leucine-rich repeat (LRR)-containing genes within the soybean genome. Our investigation revealed distinct features of three specific genes (Glyma18G281700, Glyma18G281600, and Glyma18G281500) located within the Rpp1 locus on chromosome 18. Phylogenetic and domain analyses validated the presence of the ULP1 domain in the genes of the Rpp1 locus, while motif analysis unveiled unique patterns exclusive to these genes. Notably, these genes exhibited exclusive isoleucine and glutamic acid residues at positions 836 and 904, accompanied by an uncommon expansion of 11 glutamic acid residues at position 861. Expression data indicated that two of these genes, Glyma18G281700 and Glyma18G281600, displayed significantly higher expression levels during leaf trifoliate stages I and II, suggesting their active involvement under adverse conditions. Furthermore, our protein structure analysis revealed that Glyma18G281500, one of the three genes within the Rpp1 locus, exhibited notable distinctions from the other two genes and displayed a more stable protein structure. This observation may signify a distinct functional role for Glyma18G281500 compared to the other two genes. In summary, our findings suggest that the genes within the Rpp1 locus may contribute to rust resistance, particularly against specific rust races, owing to their distinctive characteristics associated with the ULP1 domain.



**Keywords:** Rpp1, Rust resistance in soybean, ULP1 domain




## 1. INTRODUCTION

Soybean (*Glycine max*) holds a prominent position as a globally significant crop, serving as a vital source of both protein and oil for human and animal consumption. However, soybean plants continually face formidable challenges from a spectrum of pathogens encompassing bacteria, fungi, and viruses, which pose substantial threats to crop yield and quality (Li et al. 2022). Among these adversaries, soybean rust, incited by the biotrophic fungus *Phakopsorapachyrhizi*, stands out as one of the most deleterious foliar diseases afflicting soybean (Hu et al. 2020). Over the years, extensive screening efforts involving at least 200 germplasm accessions and breeding lines have led to the identification of seven resistance loci denoted as Rpp (signifying resistance to *P. pachyrhizi*). These loci include Rpp1 (Pedley et al. 2019), Rpp2 (Bromfield & Hartwig, 1980), Rpp3 (Bromfield & Melching, 1982), Rpp4 (Hartwig, 1986), Rpp5 (Garcia et al. 2008), Rpp6 (Li et al. 2016), and Rpp7 (Childs et al. 2018). Nevertheless, it is crucial to acknowledge that none of these resistance genes exhibit efficacy against all currently known soybean rust pathotypes (Childs et al. 2018).

In response to the pervasive onslaught of these pathogens, plants have evolved a sophisticated immune system centered around the recognition of pathogen-derived molecules through specific receptors. Notably, a prominent category of receptors that play a pivotal role in soybean's defense against diseases are the leucine-rich repeat-containing (LRR) domain proteins. These proteins are characterized by the presence of nucleotide-binding site (NBS) and leucine-rich repeat (LRR) domains, complemented by variable amino- and carboxy-terminal domains, all of which collectively facilitate a range of critical protein-protein interactions and ligand binding events (McHale et al. 2006). Their significance in plant defense mechanisms becomes apparent as they serve as the vanguard, recognizing and responding to pathogen-associated molecular patterns (PAMPs) or effector molecules released by invading pathogens (Zhang et al. 2022). Notably, a majority of disease resistance genes (R genes) identified in plants thus far are comprised of nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins (Zhang et al. 2022). Furthermore, LRR domain-containing proteins have also been implicated in reinforcing soybean's defense against fungal pathogens, as evidenced by previous studies (Zhou et al. 2016; Si et al. 2021).

The objective of the present investigation was to conduct a comparative analysis of the Rpp1 locus with other NB-ARC-LRR domain-containing genes in soybean. This comprehensive analysis aimed to elucidate the variations and specificities associated with the Rpp1 locus, shedding light on its unique role within the intricate framework of soybean's defense mechanisms against the formidable soybean rust pathogen.

## 2. MATERIALS AND METHODS

### *Isolation and comparison of NBS-LRR containing genes*

In pursuit of identifying the nucleotide-binding site leucine-rich repeat (NBS-LRR) domain, a systematic search was conducted across a soybean genome in UNIPROT database (https://www.uniprot.org/) (supplementary file 1). To gain insights into the evolutionary relationships among various leucine-rich repeat (LRR) domain-containing proteins, including those within the Rpp1 locus, a rigorous multiple sequence alignment was executed. This alignment was carried out using ClustalW, into the BIOEDIT software (version 7.2). The precise conserved regions of amino acid sequences were compared among all LRR domain-containing proteins.

### *Phylogenetic Analysis*

To construct a phylogenetic tree, Neighbor Joining analysis was employed. This analysis was carried out using the Simple phylogeny tool (https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/). Subsequently, the resultant phylogenetic tree was refined and visually enhanced using the Interactive Tree of Life (ITOL) platform (https://itol.embl.de/upload.cgi), providing a clear depiction of the relationships and distinctions among the analysed proteins.

### *Identification of Protein Domains*

To ascertain the presence of both single and multiple domains within the protein sequences under investigation, a dedicated analysis was conducted using the batch-CD tool for leucine-rich repeat (LRR) domain-containing proteins, including those within the Rpp1 locus, which is accessible through the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Further, TB-tool was employed for the visualization of domain structures present within the analyzed protein sequences. This visualization facilitated a comprehensive understanding of the arrangement and distribution of domains, contributing to the elucidation of potential functional roles.

### *Motif Analysis*

To uncover conserved motifs within the protein sequences, a motif search was diligently carried out for leucine-rich repeat (LRR) domain-containing proteins, including those within the Rpp1 locus. This search was performed using the MEME

tool (http://meme.nbcr.net/meme/). By identifying conserved motifs, this analysis shed light on potential functional elements shared among the studied proteins.

***Expression data analysis***

Expression data pertinent to the various stages and tissues associated with the studied gene sequences were procured from the Legume Mine database via the Legume Information System (LIS). These data resources provided valuable insights into the expression patterns of the analyzed genes across different developmental stages and tissues. To effectively visualize the expression patterns and relationships among the studied genes, the acquired expression data were transformed into a heatmap format. This transformation was accomplished using the Clustvis online tool (https://biit.cs.ut.ee/clustvis/), which enabled the creation of a graphical representation facilitating the interpretation of gene expression profiles.

**Protein structure analysis**

The Phyre2 automatic fold recognition server was utilized to predict the structure and/or function of three genes (Glyma18G281700, Glyma18G281600, and Glyma18G281500) with a modeling confidence of 100.0%. These predictions were generated using the single highest scoring template (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Furthermore, additional structure-based analyses were conducted using Phyre2, including relative conservation analysis based on Jensen-Shannon divergence, evaluation of clashes between atoms in specific residues, Ramachandran analysis, and mutational sensitivity assessment using the SuSPect tool to predict potential missense mutations in the protein sequences.

## 3. **RESULT AND DISCUSSION**

Asian soybean rust (ASR) is a highly destructive ailment impacting soybean production on a global scale (Godoy et al. 2016). Genes that confer resistance against *Phakopsorapachyrhizi* (Rpps) have been harnessed to safeguard soybeans from this disease. Seven resistance to P. *pachyrhizi* (Rpp) loci have been identified and situated on genetic linkage maps of soybeans (Garcia et al. 2011; King et al. 2007; Lemos et al. 2011; Hartwig et al. 1986; Yu et al. 2015; Monteros et al. 2007; Kendrick et al. 2011; Childs et al. 2018). Due to the considerable variation in *P.pachyrhizi* races/isolates, relying on a single

Rpp locus/gene proves insufficient to offer lasting and comprehensive resistance against ASR disease.

The NBS-LRR protein, encoded by the NBS-LRR gene in plants, plays a critical role in enabling plants to counteract pathogen invasions (Tan et al. 2012). There have been numerous extensive reviews since the first NBS-LRR-encoding genes were cloned from plants in 1994 (Jones et al. 1997; Ellis et al. 2000; Meyers et al. 2003; Belkhadiret al. 2004; Jones et al. 2004). This gene is notably abundant across various plant species such as Arabidopsis thaliana and soybean, encompassing 149 and 319 NBS-LRR genes family, respectively (Wang et al. 2023). The NBS-LRR protein is also referred to as NB-ARC-LRR (nucleotide binding adaptors composed of Apaf-1, R protein, and CED-4) protein due to the presence of nucleotide binding and ATPase domains (NB-ARC), which function as signal transduction domains within a multifaceted protein superfamily (Wang et al. 2023).

The current study was undertaken to comprehensively characterize the Rpp1 locus in soybeans, with the primary objective of elucidating the domain specificity underlying rust resistance. In a prior investigation (Ratnaparkheet al. 2020), the Rpp1 locus was identified as harboring three distinct NB-ARC-LRR containing genes: Glyma18G281700, Glyma18G281600, and Glyma18G281500. Subsequently, a detailed analysis of these three genes was conducted, coupled with a comparative examination against other LRR domain-containing genes within the soybean genome.

To elucidate the genetic complexities within this genomic region, we commenced a multiple sequence alignment encompassing 21 distinct genes encoding leucine-rich repeat (LRR) domains. This systematic analysis unveiled significant disparities in conserved segments when juxtaposed with genes localized within the Rpp1 genomic region, as illustrated in Figure 1. Notably, within this gene set, exclusive to the Rpp1 locus were the occurrences of isoleucine and glutamic acid residues at positions 836 and 904, respectively. Furthermore, a distinctive feature emerged, with an expansion of 11 glutamic acid residues observed at position 861, setting it apart from all other genes under examination (Figure 1). This distinct pattern suggests its singularity and implies a potential involvement in defensive mechanisms. Phylogenetic analysis was executed to investigate the evolutionary relationships between the genes associated with rust resistance (Rpp1) and other LRR-containing genes in soybean. Our findings unveiled the presence of four distinct clades, with the Rpp1 genes (Glyma18G281700, Glyma18G281600, and Glyma18G281500) forming a distinct clade of their own, highlighting their unique functional roles among the diverse LRR genes (Figure 2).In addition to phylogenetic insights, we conducted domain analysis to probe the roles of various domains in soybean rust resistance. Our examination unveiled several

domains present in the 21 genes, including the NB-ARC superfamily, RX-CC_like, LRR superfamily, Rx N, PLN03210 superfamily, ULP1 superfamily, Retrotrans_gag superfamily, and PLN00113 superfamily, as depicted in Figure 3. Of particular interest, the ULP1 domain was identified within the Rpp1 locus genes, suggesting its potential association with the defense against rust disease. Notably, Pedley *et al*. (2019) corroborated our findings by reporting that each candidate gene within the Rpp1 locus encodes an N-terminal ubiquitin-like protease 1 (ULP1) domain. Their research further demonstrated that co-silencing of these Rpp1 candidates resulted in the abrogation of the immune response in the Rpp1-resistant soybean accession PI 200492, underscoring the pivotal role of Rpp1 as a ULP1-NBS-LRR protein in orchestrating the immune response.Motif analysis was also performed, revealing distinct patterns among the three genes within the Rpp1 locus, as depicted in Figure 4. There was absence of several motif was observed in Rpp1 locus containing genes.

Within LLR proteins, several highly conserved and meticulously ordered motifs are recognized, including the P-loop, kinase-2, and Gly-Leu-Pro-Leu motifs (Tan et al. 2012). The LRRs (leucine-rich repeats) constitute structurally adaptable domains specializing in protein-protein interactions, with the capacity to evolve diverse binding specificities. These LRRs are subject to diversifying selection (Ellis et al. 2000; Jones et al. 2006), particularly at the level of the projected solvent-exposed residues. Intriguingly, in this region, the NBS-LRR proteins not only lack uniformity but exhibit significantly greater diversity than anticipated through random genetic drift. This suggests the presence of selective pressures that promote the development of novel pathogen-specificities, catering to the recognition of distinct pathogen Avr proteins. The evolution of this gene family is also influenced by processes such as clustering of NBS-LRR genes through tandem duplications and ectopic duplications followed by local rearrangements and gene conversion (Wang et al. 2023).

Furthermore, we delved into the functional aspects of these genes by analyzing their expression data across various tissues and developmental stages. Among the 21 genes studied, Glyma18G281700 and Glyma18G281600 exhibited relatively higher expression levels during leaf trifoliate Stage-I and stage II, implying their active involvement during stress conditions (Figure-5).This study also aligns with recent research by Si *et al*. (2021), which delved into the expression patterns of genes encoding LRR-RLKs in oomycetes. Their findings indicate a dual pattern of gene expression, with active transcription during vegetative growth stages and host infection. This suggests the pivotal roles played by LRR-RLKs in various facets of the oomycete life cycle, encompassing growth, development, and interactions with host organisms. This aspect of our research contributes valuable insights into the functional significance of LRR-RLKs in oomycetes across different phases of their life cycle and their potential implications in host-pathogen interactions.

Oomycetes exhibit a diverse array of LRR-RLKs (Leucine-Rich Repeat Receptor-Like Kinases), with each species demonstrating a unique complement of these receptor proteins. What's intriguing is that these LRR-RLKs in oomycetes share a common structural pattern with their counterparts in extensively researched plant LRR-RLKs. This structural similarity suggests potential functional parallels between oomycetes and plants in terms of signal transduction and cellular communication mechanisms (Dievartet al. 2011; Soanes and Talbot, 2010). Further investigation into these shared characteristics could provide valuable insights into the roles these receptors play in both oomycete biology and plant-pathogen interactions. In our study, various parameters related to protein structure were also analyzed to comprehensively understand the functional diversity among three Rpp1-containing genes, namely Glyma18G281700, Glyma18G281600, and Glyma18G281500 (Figure 6). Among these three genes, Glyma18G281500 exhibited notable differences in terms of sequence compression (Figure 6A) and three-dimensional protein folding (Figure 6B). Additionally, Glyma18G281500 displayed a relatively lower number of clashes between atoms in different residues compared to the other two genes. However, non-significant differences were observed among all three genes in terms of relative conservation analysis, Ramachandran analysis, and mutational sensitivity.This comprehensive protein structure analysis suggests that Glyma18G281500 differs significantly from that of the other two genes and more stable in terms of atomic interactions. these differences do not necessarily translate into significant variations in terms of overall protein conservation, geometric conformations, or sensitivity to mutations across the three genes.

In summary, our study integrates phylogenetic analysis, domain analysis, motif analysis, and expression data to elucidate the distinct behavior of genes within the Rpp1 locus, particularly due to the presence of the ULP1 domain. Additianlly, among 3 Rpp1 containing genes Glyma18G281500 differs significantly from other two genes and also found with more stable protein structure, may suggest its distinct role over two other genes. While our findings suggest a potential contribution of these genes to rust resistance, it is crucial to acknowledge the possibility of race-specific responses. Further investigations in this realm hold promise for advancing our understanding of rust resistance mechanisms in soybeans.

## DISCLAIMER (ARTIFICIAL INTELLIGENCE)

The author certifies that no generative AI technologies, including but not limited to large language models (such as Copilot, etc.) or text-to-image generators, were used in the writing or editing of this manuscript.

## COMPETING INTERESTS

The authors of this manuscript declare that they have no financial or non-financial conflicts of interest that could influence their research or the interpretation of their results.

## REFERENCES


1. Belkhadir Y, Subramaniam R, Dangl JL. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr Opin Plant Biol. 2004;7:391-9.
2. Bhor TJ, Chimote VP, Deshmukh MP. Molecular tagging of Asiatic soybean rust resistance in exotic genotype EC 241780 reveals complementation of two genes. Plant Breeding. 2015;134:70-7.
3. Bromfield KR, Melching JS. Sources of specific resistance to soybean rust. Phytopathology. 1982.
4. Bromfield KR, Melching JS, Kingsolver CH. Virulence and aggressiveness of Phakopsorapachyrhizi isolates causing soybean rust. Phytopathology. 1980;70(1):17-21.
5. Bromfield KR, Hartwig EE. Resistance to soybean rust and mode of inheritance. Crop Sci. 1980;20:254-5.
6. Chakraborty N, Curley J, Frederick R, Hyten D, Nelson R, Hartman G, Diers B. Mapping and confirmation of a new allele at Rpp1 from soybean PI 594538A conferring RB lesion-type resistance to soybean rust. Crop Sci. 2009;49:783-90.
7. Cheng Q, Li N, Dong L, Zhang D, Fan S, Jiang L, et al. GmLecRK3a, a pattern recognition receptor from soybean, is an efficient receptor for Soybean mosaic virus. Mol Plant Pathol. 2017;18(7):961-74.
8. Childs SP, King ZR, Walker DR, Harris DK, Pedley KF, Buck JW, Boerma HR, Li Z. Discovery of a seventh Rpp soybean rust resistance locus in soybean accession PI 605823. Theor Appl Genet. 2018;131:27-41.
9. Diévart A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, Périn C. Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. BMC Evol Biol. 2011;11:1-9.
10. Ellis J, Dodds P, Pryor T. Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol. 2000;3:278-84.

11. Garcia A, Calvo ES, de Souza Kiihl RA, Harada A, Hiromoto DM, Vieira LGE. Molecular mapping of soybean rust (Phakopsorapachyrhizi) resistance genes: Discovery of a novel locus and alleles. Theor Appl Genet. 2008;117:545-53.

12. Garcia A, Calvo ÉS, de Souza Kiihl RA, Souto ERD. Evidence of a susceptible allele inverting the dominance of rust resistance in soybean. Crop Sci. 2011;51:32-40.

13. Godoy CV, et al. Ferrugem-asiática da soja no Brasil: passado, presente e futuro. PesqAgropec Bras. 2016;51:407-21.

14. Hartwig EE. Identification of a fourth major gene conferring resistance to soybean rust. Crop Sci. 1986;26:1135-6.

15. Hu D, Chen ZY, Zhang C, Ganiger M. Reduction of Phakopsorapachyrhizi infection on soybean through host- and spray-induced gene silencing. Mol Plant Pathol. 2020;21(6):794-807.

16. Hyten DL, Hartman GL, Nelson RL, Frederick RD, Concibido VC, Narvel JM, Cregan PB. Map location of the locus that confers resistance to soybean rust in soybean. Crop Sci. 2007;47:837-40.

17. Jones DA, Jones JDG. The role of leucine-rich repeat proteins in plant defenses. Adv Bot Res. 1997;24:90-167.

18. Jones DA, Takemoto D. Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. Curr Opin Immunol. 2004;16:48-62. doi: 10.1016/j.coi.2003.11.016.

19. Jones JD, Dangl JL. The plant immune system. Nature. 2006;444:323-9.

20. Kendrick MD, Harris DK, Ha BK, Hyten DL, Cregan PB, Frederick RD, Boerma HR, Pedley KF. Identification of a second Asian soybean rust resistance gene in Hyuuga soybean. Phytopathology. 2011;101:535-43.

21. King ZR, Childs SP, Harris DK, Pedley KF, Buck JW, Boerma HR, Li Z. A new soybean rust resistance allele from PI 423972 at the Rpp4 locus. Mol Breed. 2017;37:62.

22. Lemos NG, Braccini ADL, Abdelnoor RV, de Oliveira MCN, Suenaga K, Yamanaka N. Characterization of genes Rpp2, Rpp4, and Rpp5 for resistance to soybean rust. Euphytica. 2011;182:53-64.

23. Li S, Smith JR, Ray JD, Frederick RD. Identification of a new soybean rust resistance gene in PI 567102B. Theor Appl Genet. 2012;125:133-42.

24. Li W, Chakrabarty PK, Yu Y. Identification and characterization of a leucine-rich repeat receptor-like kinase gene that confers race-specific resistance to Pseudomonas syringaepv. glycinea in soybean. Mol Plant Pathol. 2016;17(8):1254-65.

25. Li Y, Shi C, Wei D, Gu X, Wang Y, Sun L, et al. Soybean continuous cropping affects yield by changing soil chemical properties and microbial community richness. Front Microbiol. 2022;13:1083736.

26. Liu M, Li S, Swaminathan S, Sahu BB, Leandro LF, Cardinal AJ, Bhattacharyya MK, Song Q, Walker DR, Cianzio SR. Identification of a soybean rust resistance gene in PI 567104B. Theor Appl Genet. 2016;129:863-77.

27. McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. Genome Biol. 2006;7:1-11.

28. Meyer JD, Silva DC, Yang C, Pedley KF, Zhang C, van de Mortel M, Hill JH, Shoemaker RC, Abdelnoor RV, Whitham SA. Identification and analyses of candidate genes for Rpp4-mediated resistance to Asian soybean rust in soybean. Plant Physiol. 2009;150:295-307.

29. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell. 2003;15:809-34. doi: 10.1105/tpc.009308.

30. Monteros MJ, Missaoui AM, Phillips DV, Walker DR, Boerma HR. Mapping and confirmation of the 'Hyuuga' red-brown lesion resistance gene for Asian soybean rust. Crop Sci. 2007;47:829-34.

31. Pedley KF, Pandey AK, Ruck A, Lincoln LM, Whitham SA, Graham MA. Rpp1 encodes a ULP1-NBS-LRR protein that controls immunity to Phakopsorapachyrhizi in soybean. Mol Plant Microbe Interact. 2019;32(1):120-33.

32. Ratnaparkhe MB, Marmat N, Kumawat G, Shivakumar M, Kamble VG, Nataraj V, et al. Whole genome re-sequencing of soybean accession EC241780 providing genomic landscape of candidate genes involved in rust resistance. Curr Genomics. 2020;21(7):504-11.

33. Savary S, et al. The global burden of pathogens and pests on major food crops. Nat EcolEvol. 2019;3:430-9.

34. Si J, Pei Y, Shen D, Ji P, Xu R, Xue X, et al. Phytophthora sojae leucine-rich repeat receptor-like kinases: diverse and essential roles in development and pathogenicity. iScience. 2021;24(7).

35. Silva DC, Yamanaka N, Brogin RL, Arias CA, Nepomuceno AL, Di-Mauro AO, et al. Molecular mapping of two loci that confer resistance to Asian rust in soybean. Theor Appl Genet. 2008;117:57-63.

36. Soanes DM, Talbot NJ. Comparative genome analysis reveals an absence of leucine-rich repeat pattern-recognition receptor proteins in the kingdom Fungi. PLoS One. 2010;5(9)

37. Tan S, Wu S. Genome-wide analysis of nucleotide-binding site disease resistance genes in Brachypodiumdistachyon. Comp Funct Genomics. 2012.

38. Wang X, Xu Y, Fan H, Cui N, Meng X, He J, et al. Research Progress of Plant Nucleotide-Binding Leucine-Rich Repeat Protein. Horticulturae. 2023;9(1):122.

39. Yu N, Kim M, King ZR, Harris DK, Buck JW, Li Z, Diers BW. Fine mapping of the Asian soybean rust resistance gene Rpp2 from soybean PI 230970. Theor Appl Genet. 2015;128:387-96.

40. Zhang S, Li C, Si J, Han Z, Chen D. Action mechanisms of effectors in plant-pathogen interaction. Int J Mol Sci. 2022;23(12):6758.

41. Zhang S, Xu R, Luo X, Jiang Z, Shu H, Zhang Y. A Lysin motif-containing protein GmLecRK2 positively regulates resistance of soybean to Phytophthora sojae. J Exp Bot. 2018;69(5):1185-97.

42. Zhou F, Guo Y, Qiu LJ. Genome-wide identification and evolutionary analysis of leucine-rich repeat receptor-like kinase genes in soybean. BMC Plant Biol. 2016;16:1-13.

**Figures:**

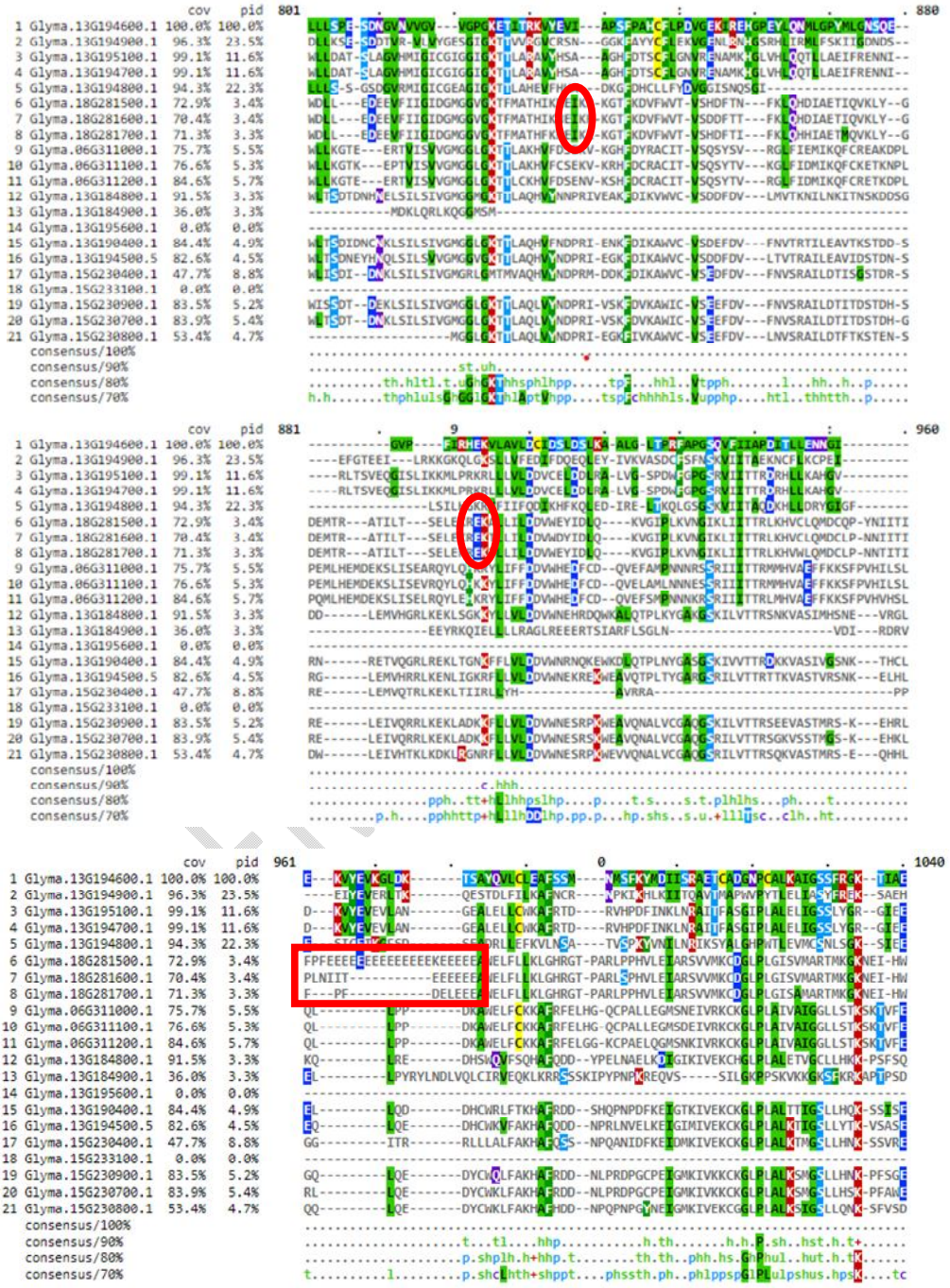

**Figure 1: Depicting conserved region in different NB-ARC- LRR containing genes in Soybean.** Multiple sequence alignment was executed by using ClustalW into the BIOEDIT software (version 7.2) for 21 different LRR containing genes including Rpp1 locus underlined genes. Conserved regions of amino acid sequences

were highlighted among all LRR domain-containing proteins. Rpp1 locus genes (Glyma18G281700, Glyma18G281600, and Glyma18G281500) contained exclusive isoleucine and glutamic acid residues at positions 836 and 904, along with an unusual expansion of 11 glutamic acid residues at position 861.

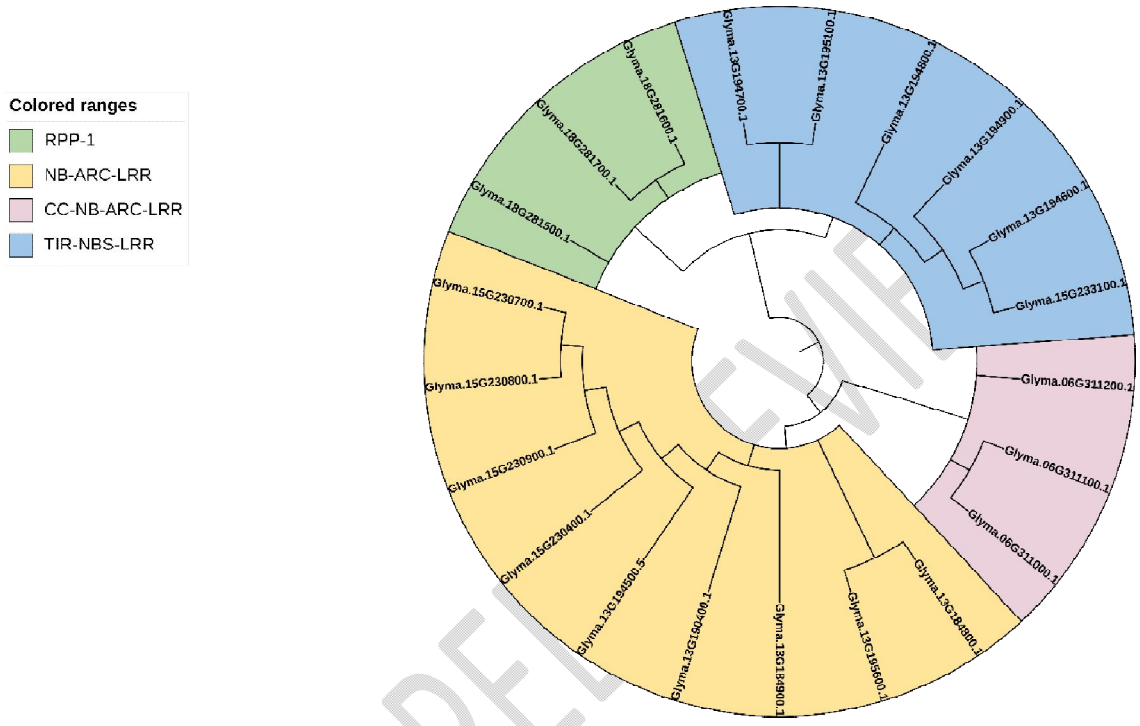


**Figure 2: Phylogenetic tree of NB-ARC- LRR containing domain protein family in Glycine Max.** Phylogenetic tree was constructed for 21 different LRR containing genes including Rpp1 locus underlined genes by using Simple phylogeny, a tool provided by EMBL-EBI (https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny) and a life tree was built using 1000 bootstrap tests on iTOL (Interactive tree of life) (https://itol.embl.de/). Phylogenetic tree showed four distinct clades, with the Rpp1 genes (Glyma18G281700, Glyma18G281600, and Glyma18G281500) forming a distinct clade of their own.

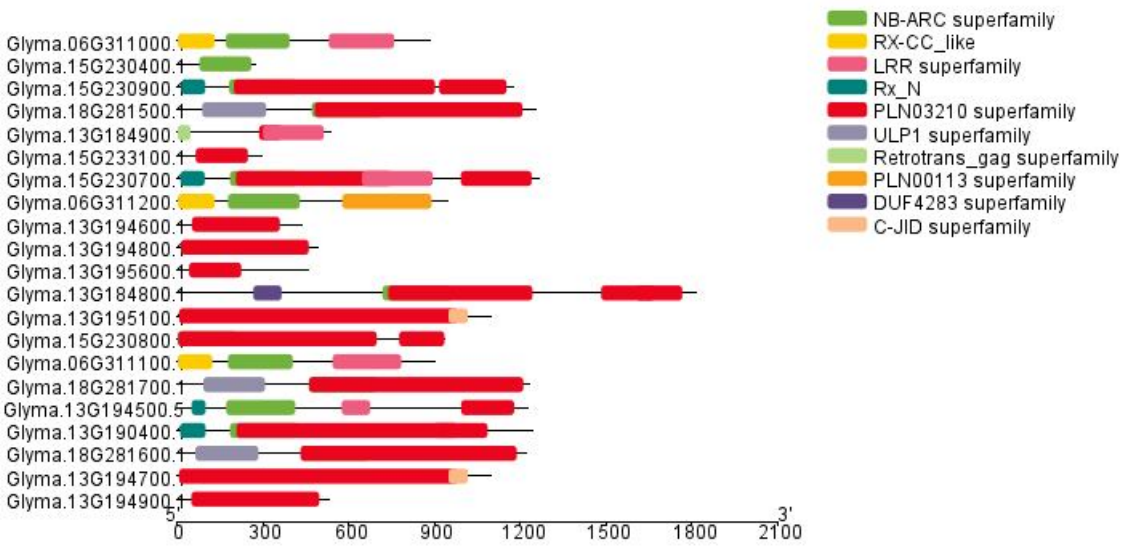


**Figure 3: Domains compositions of NB-ARC- LRR containing genes in Soybean.**Domain analysis was conducted using the batch-CD tool in NCBI (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) for 21 different LRR containing genes including Rpp1 locus underlined genes. Further, TB-tool was employed for the visualization of domain structures. Each domain is represented in the different colored box as per legends. A distinct domain ULP1 was observed within the Rpp1 locus genes.

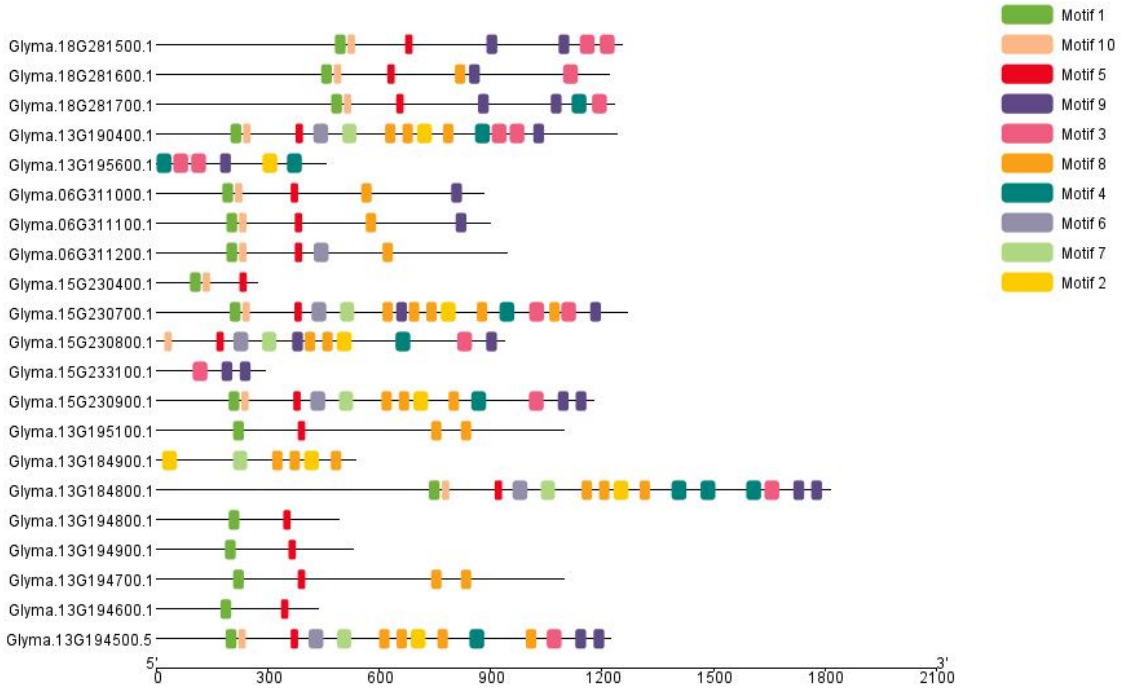


**Figure 4: Distribution of different motif on 21 different genes in soybean*:*** Motif analysis was done for 21 different LRR containing genes including Rpp1 locus underlined genes by using MEME tool (http://meme.nbcr.net/meme/). Different motif was shown in different color as per legends for each genes. All three Rpp1 genes (Glyma18G281700 and Glyma18G281600) exhibits distinct pattern of motif than other tested genes.

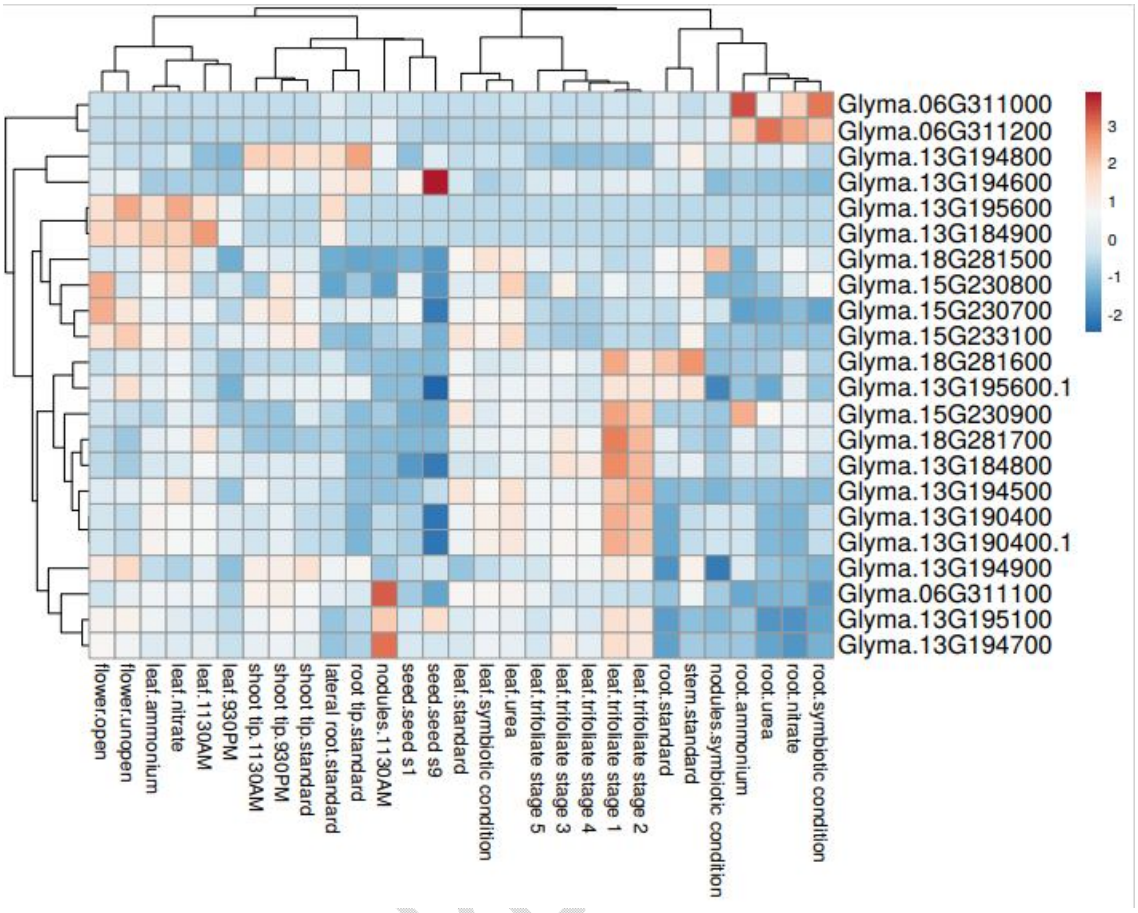


**Figure 5: Expression patterns of NB-ARC- LRR containing genes in Soybean.**Expression data were collected from Legume Mine database via the Legume Information System (LIS) of various stages and tissues for 21 different LLR containing genes including Rpp1 locus. Data were presented in the form of heatmap.Two Rpp1 genes (Glyma18G281700 and Glyma18G281600) exhibited relatively higher expression levels at leaf trifoliate Stage-I and stage II, suggesting their active involvement during stress conditions.

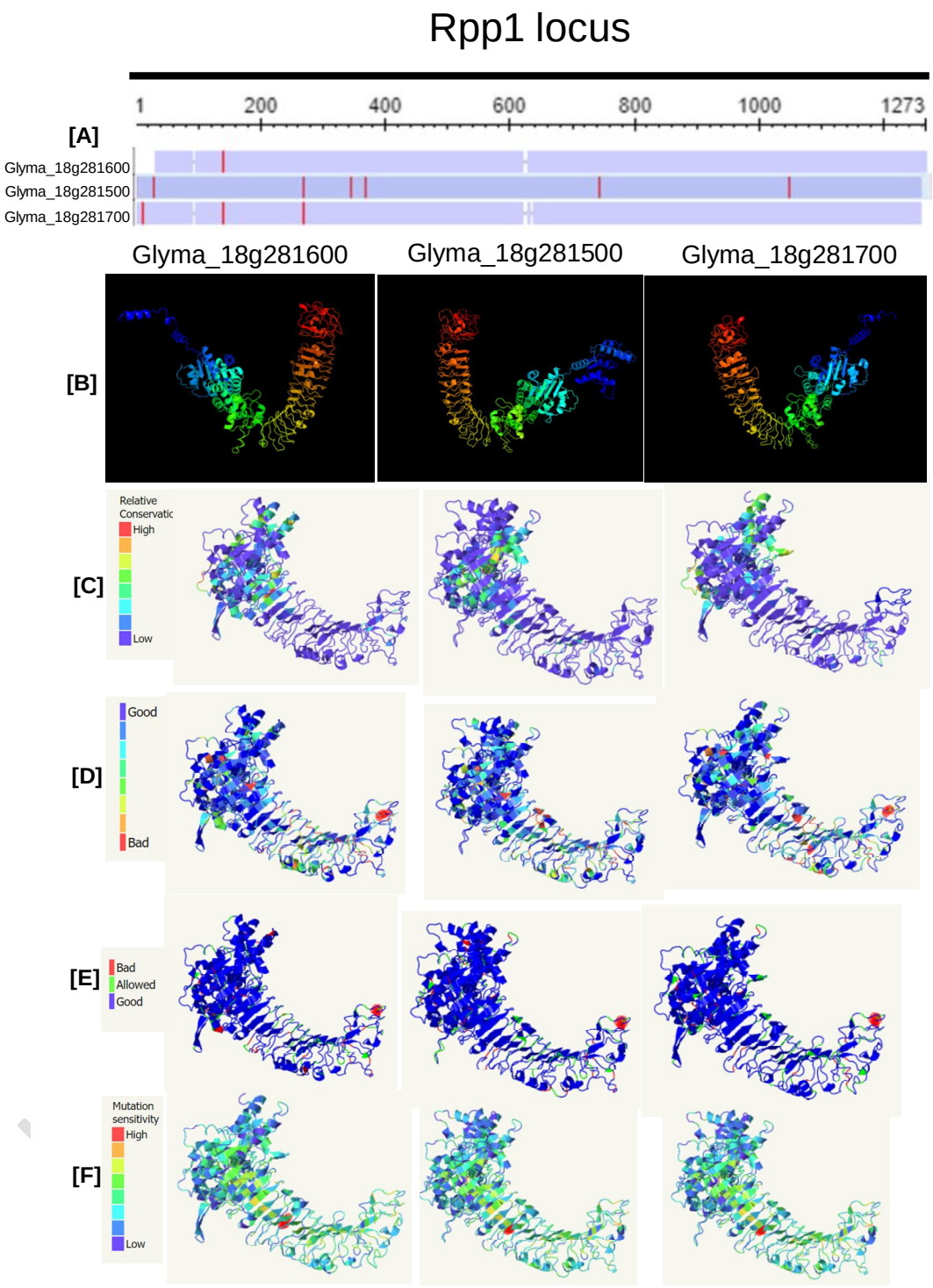


**Figure 6: Evaluation of different Protein structure based parameters for Rpp1 containing genes.**Protein sequence were compared with Constraint-based Multiple Alignment Tool (COBALT) (https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web) (A).The Phyre2 automatic fold recognition server was utilized to predict the structure and/or function of three Rpp1 containing genes (Glyma18G281700, Glyma18G281600, and

Glyma18G281500) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (B). Furthermore, structure-based analyses were conducted that includes relative conservation analysis (C), clashes analysis between atoms in specific residues (D), Ramachandran analysis (E), and mutational sensitivity assessment (F) using the SuSPect tool to predict potential missense mutations in the protein sequences.Glyma18G281500, one of the three genes within the Rpp1 locus, differs significantly from the other two genes and also exhibits a more stable protein structure as it showed less number of clashes.