Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Map Based Cloning of a Dominant Rust Resistance Gene and Mapping of its Duplicated Paralogues in Cultivated Groundnut (Arachis hypogaea L.)

Suvendu Mondal (suvenduhere@yahoo.co.in)
Bhabha Atomic Research Centre https://orcid.org/0000-0001-7638-140X
K. Mohamed Shafi
NCBS: National Centre for Biological Sciences
Avi Raizada
Homi Bhabha National Institute
Devyanee S. Kesvad
University of Mumbai
Pallavi Bhat
Saint Aloysius College
Song Hui
Qingdao Agricultural University
Ramanathan Sowdhamini
National Centre for Biological Sciences
Anand M. Badigannavar
Bhabha Atomic Research Centre

Original Article

Keywords: Arachis hypogaea L., Indel marker, Ka/Ks ratio, Map-based cloning, Non-synonymous mutation, R gene, SNP marker

Posted Date: February 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-158595/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Understanding the mechanism and nature of resistance genes in crop plants is essential for its use in new breeding techniques. Previously, a dominant rust resistance gene was fine-mapped within a 1.2 cM interval in chromosome A03 of groundnut. Here, the rust resistance gene, *VG9514-Rgene* was isolated through map based cloning. Sequencing of the gene from resistant and susceptible plants revealed non-synonymous mutations in the TIR, NBS and LRR region of R-protein. Genetic mapping of these SNPs-based markers confirmed the position of *VG9514-Rgene* in between FRS 72 and SSR_GO340445 markers in chromosome A03. Homology searching identified four homologous R-genes in groundnut genome. Of them, *Arahy.R8KUIR*, *Arahy.T6DCA5* and *Arahy.ZZ0VZ9* are paralogues. These paralogous genes had several small InDels. Mapping of these InDels-based markers revealed tandem duplication of these paralogous R-genes at distal portion of chromosome A03. K_a/K_s calculation revealed that this unique *VG9514-Rgene* had undergone positive selection. Homology based structure modelling of this R-protein revealed a typical consensus three dimensional folding of TIR-NBS-LRR protein. Non-synonymous mutations in susceptible version of R-protein were mapped in this protein model and found E268Q mutation in hhGRExE motif, Y309F in RNBS-A motif and I579T in MHD motif of NB-ARC domain are probable candidates for loss of function.

Introduction

Plant genome contains many resistance genes (*R*-genes) and their analogues (RGAs), which provide non-host resistance to plants (Schulze-Lefert and Panstruga 2011). In nature, specific *R*-genes have been evolved against specific plant pathogens (Flor 1942; Guido et al. 1992). Such specific interaction between pathogen effector (Avr gene product) and *R*-gene impart vertical resistance to crop plants, where such resistance is mostly controlled by oligogenes. *R* genes are usually dominant in nature which provides full or partial resistance to one or more pathogens. Way back in 1992, first *R* gene *Hm1* was cloned from maize (Johal and Briggs 1992). Since then, many *R* genes were cloned from different plants including crop species. This account a total of 314 cloned functional *R* genes which operate through nine distinct mechanisms inside plant cell towards imparting disease resistance (Kourelis and van der Hoorn 2018). Jiang et al. (2018) isolated a broad spectrum and durable NB-LRR gene *R8* that provide resistance against *Phytophthora infestans*. Similarly, a unique tandem kinase-pseudokinases R gene that provides broad spectrum resistance to *Puccinia striiformis f. sp. tritici* (Pst), was isolated based on map based cloning approach (Klymiuk et al. 2018). Most resistance genes encode intracellular nucleotide binding/leucine-rich repeat (NLR) immune receptor proteins that have three distinct domains, Toll/Interleukin 1 Receptor (TIR) or Coiled Coil (CC), Nucleotide Binding-ARC (NB-ARC) Site (NBS) and Leucine Rich Repeat (LRR). NB-ARC domain named so due to the presence in APAF-1 (apoptotic protease-activating factor-1), R proteins and CED-4 (*Caenorhabditis elegans* death-4 protein) along with nucleotide binding sites (van der Biezen and Jones 1998). Cloning of R-genes can be achieved with genetic map based positional cloning or mutational genomics approach (Arora et al. 2019).

Prior to genome sequence availability in groundnut, R-genes were predicted from expressed sequence tags (EST) using known R-gene protein sequence from model plant. Liu et al. (2013) identified six different classes of R genes from 1053 ESTs which were later assembled into 156 contigs and 229 singletons as groundnut-expressed RGAs. Genome sequence initiative in wild diploid progenitor has yielded 278 NBS from *A. duranensis* and 303 NBS from *A. ipaensis* (Song et al. 2017). Subsequently, Zhuang et al. (2019) identified 661 NBS domain containing R-genes in tetraploid groundnut. All these R-genes were divided into three groups: coiled coil (CC)-NBS-leucine-rich repeat (LRR) (CNL), Toll/interluekin-1 receptor (TIR)-NBS-LRR (TNL) and resistance to powdery mildew8 (RPW8)-NBS-LRR (RNL). Efforts on mapping and development of molecular markers against various disease resistance genes in *Arachis* species has progressed considerably, however no reports are available on isolation and cloning of resistance gene/R-gene for a particular disease.

Breeding and cytogenetic efforts were undertaken to introgress rust resistance gene containing-chromosome fragment from *A. cardenasii* into cultivated groundnut species through wide hybridization techniques in India and abroad (Varman 1999; Simpson 2001; Stalker 2017). Such a wide hybridization derived breeding line VG 9514 was used in crossing with a cultivar TAG 24 to develop a recombinant inbred line population towards tagging a rust resistance gene. By using this population in genetic mapping and QTL mapping, SSR (Mondal et al. 2012 a, b) and MITE markers (Mondal et al. 2014a) were developed for the dominant rust resistance gene. Interestingly, same genomic loci controlling rust resistance was also detected by Khedikar et al. (2010) and Sujay et al. (2012) by using another mapping population derived from cross between TAG 24 and GPBD 4. These reports deal with the rust resistance gene which was introgressed from an A-genome species *A. cardenasii* and mapped in chromosome A03. Leal-Bertioli et al. (2015) mapped another rust resistance gene which was sourced from *A. magna* (a B-genome member) in B08 chromosome. Recently, fine mapping effort on VG 9514 x TAG 24 derived RIL population has delimited the rust resistance gene within a 1.2 cM fragment flanked by two SSR markers SSR_GO340445 and FRS 72 in chromosome A03 (Mondal and Badigannavar 2018). In Parallel, the position of the rust resistance gene in the same genomic fragment of chromosome A03 was confirmed by utilizing ddRAD-seq and whole-genome resequencing approach (Shirasawa et al. 2018). Genome sequence initiative has further helped to look into the nucleotide sequence level and revealed a 331 kb segment corresponding to the 1.2 cM fine mapped region. This 331 kb chromosome A03 fragment within the fine mapped region has several disease resistance related protein coding genes like a single TIR-NBS-LRR gene and three glucan endo-1,3 β glucosidase genes (Mondal and Badigannavar 2018). Present paper reports the research work on cloning of the TIR-NBS-LRR gene and genetic validation towards the candidature of this gene for rust resistance in cultivated groundnut. Mapping of other paralogous *R* genes in cultivated groundnut genome and position of non-synnonymous mutations in homology model of this R-protein were also undertaken in this study.

Materials And Methods

Plant materials

A rust resistant breeding line VG 9514 was used as a source of resistance, while TAG 24 was used as susceptible line. VG 9514 was bred at Tamilnadu Agricultural University, Vriddachalam from an interspecific cross between Co 1 (*A. hypogaea* L.) and *A. cardenasii* (Varman 1999). TAG 24 is a high yielding cultivar bred at Bhabha Atomic Research Centre, Mumbai and is cultivated widely across India (Patil et al. 1995). A recombinant inbred line population (VG 9514 X TAG 24) with 164 lines was used for mapping of SNP based PCR markers and InDel markers designed specifically in this study.

Cloning of *R*-gene within fine mapped region

Based on the finding from Mondal and Badigannavar (2018), a TIR-NBS-LRR gene (*Aradu*.Z87JB) was predicted as a candidate gene for rust resistance. The predicted gene was downloaded from Peanutbase (https://peanutbase.org/home). The total size of the genomic fragment of the predicted gene was about 3.9 kb. Towards cloning, four pairs of primer (Table 1; Fig 1a) were designed in such a way that some overlapping region between the PCR products exist. This overlapped region will help to assemble the derived sequence from each PCR products and extract the whole TIR-NBS-LRR sequence. Gradient PCR (60 to 70 ºC) was followed to standardize the selective amplification from each primer pairs. After that, selective amplifications of respective fragment from both VG 9514 (resistant) and TAG 24 (susceptible) were carried out in a 20 µl PCR reaction volume contained 1x colorless buffer, 2 mM $MgCl_2$, 0.2 mM each dNTPs, 0.2 µM each primer and 1.0 U of Q5® high fidelity DNA polymerase (New England Biolabs, Madison, USA). The amplification profile consisted of initial denaturation for 5 min at 95°C; 35 cycles of 30s denaturation at 95°C, 30s annealing at 62-70ºC (depending of primer pairs; Table 1), 1 min extension at 72°C and a final extension at 72°C for 10 min. The desired size of the PCR product was eluted from agarose gel (1% agarose gel in 1X TAE) and purified with a gel purification kit (Qiagen, Humburg, Germany). The purified PCR product of desired size was then ligated into pcDNA3.1-myc-HisB plasmid cut with *Eco*RV. The ligation mixture was then used for transformation into *E.* coli DH5α competent cells and the viable cells were selected within ampicilin (100 µg/ml) containing Luria broth agar plates. The positive recombinant cells were confirmed through colony PCR using T7 promoter forward primer and BGH reverse primer (Table 1). The positive colony thus obtained was inoculated in LB broth containing 100 µg/ml ampicilin. After overnight growth at 37ºC, the recombinant plasmids were isolated from positive cells and the insert size was confirmed through double digestion with *Xho*I and *Hind*III enzymes. For each fragment, 10 colonies were isolated and further confirmed through double digestion of recombinant plasmid with above enzymes. The recombinant plasmids from five such positive clones were then sequenced bi-directionally using vector-specific primers.

Sequencing of the target *R*-gene

The positive recombinant plasmid DNA from each *R*-gene fragment was sequenced bi-directionally in an ABI PRISM 3700 DNA Analyzer (Applied Biosystem, CA, USA) using T7 promoter forward and BGH reverse primer at Dr. KPC Life Sciences Pvt. Ltd., Falta SEZ, West Bengal, India. Removal of the extra dye was performed according to the manufacturer's protocol. Vector sequence was trimmed from each sequence

and all the fragments were aligned and the whole R-gene sequence was extracted. The extracted *R*-gene sequence from VG 9514 and TAG 24 were then submitted at NCBI.

Identification of SNPs and design of SNP-based PCR marker

Both the *R*-gene sequence from VG 9514 and TAG 24 were aligned in ClustalW tool (https://www.genome.jp/tools-bin/clustalw) and similarity and/or mismatch were pointed out. The DNA sequences were then translated *in silico* and protein sequences were generated. Based on protein sequence alignment non-synonymous mutations were found and PCR-based SNP primer pairs (Table 1) based on 3'-mismatches were designed in WebSNAPPER tool (Drenkard et al. 2000). Each of the PCR primer pairs (0.2 µM) were then tested for polymorphism in between VG 9514 and TAG 24. PCR was performed in a 10 µl reaction mixture with specific annealing temperature (Table 2) and as per following thermal condition: initial denaturation for 5 min at 95 °C; 25 to 32 cycles of 30 s denaturation at 95 °C, 30 s annealing at 58-67 ºC (depending of primer pairs; Table 2),30s extension at 72°C and a final extension at 72°C for 5 min.PCR amplification of SNP markers were carried out by using 1 X buffer, 0.2 mM dNTP and 1.5 U *Taq* DNA polymerase supplied by Board of Radiation and Isotope Technology (BRIT), Mumbai, India. The amplified product was size separated in 1.5% agarose gel along with 100 bp DNA ladder and stained with 0.1% ethidium bromide. Later, the stained agarose gel was photographed in a gel documentation system (Kodak, Rochester, USA)

Identification of Indel in other homologous sequence

The identified *R*-gene sequence of VG 9514 was used in BLASTn against all three genome in Peanutbase (https://peanutbase.org/home). Based on the searching parameter (coverage >80%, E value = 0, and identity >90%), four homologous sequences in *A. hypogaea* and one each in *A. duranensis* and *A. ipaensis* were found. Within these sequences, small InDel (3 to 18 bp) were identified through multiple sequence alignment in ClustalW2. Primer pairs were designed from the flanking sequences of those identified InDel and later used for detection of polymorphism and genetic mapping studies (Table 1).

Genotyping of mapping population with SNP and InDel markers

Genomic DNA was isolated freshly from young leaves of all 164 RILs (VG 9514 X TAG 24) along with parents. The polymorphic SNP-based PCR markers and InDel markers were used for genotyping in all the 164 RILs. Standard PCR conditions along with optimum annealing temperatures (Table 1) were followed for PCR amplification of SNP-based PCR markers and InDel markers. For the PCR amplification of InDel markers 1U of Go *Taq* polymerase was used with 0.2 µM of each primer, 200 µM dNTPs and 1X colorless buffer (Promega, Madison, USA). The amplified products from InDel markers were resolved in a capillary gel electrophoresis (Qiagen, Germany). The size of the PCR product (allele size) and scoring were done using the QiaExcel software (Qiagen, Germany).

Linkage analysis

The linkage analysis was performed using QTL IciMappingver. 4.1 (Wang et al. 2016). Minimum LOD score of 3.5 was set as thresholds for linkage group determination. For ordering and rippling of grouped markers 'nnTwoOpt' and 'SAD' (Sum of adjacent distances) command were used, respectively. The map distance was expressed in centiMorgan (cM) using the Kosambi (1944) map function. The newly developed SNP based marker and InDel markers were included in the existed genetic linkage map of Mondal et al. (2014b) by following the above mapping protocol in QTL IciMapping ver. 4.1.

Validation of SNP-based PCR marker in other resistant genotypes

A set of 11 rust resistant (including VG 9514) and 11 rust susceptible (including TAG 24) groundnut genotypes were used to validate the strong association/co-segregation of *R*-gene sequence derived SNP based PCR marker with rust resistance in field (Table 3). Genomic DNA was isolated freshly from all these above genotypes. 10 µg DNA from each genotype was used for PCR reaction of SNP based PCR marker by following the temperature profile and reaction mixture condition as mentioned in above.

HRM Realtime-PCR amplification and data analysis

High Resolution Melting (HRM) assay was performed in Rotor-Gene Q realtime PCR Thermocycler (Qiagen, Germany) using pre-screened amplified primer pairs as per the protocol of Type-it HRM PCR Kit (Qiagen, Hamburg, Germany). Specific amplification for each primer pair was checked on 2.5% agarose gel. Three primer-pairs (HRM-1, HRM-4 and HRM-6) harbouring multiple SNPs with product size of ~100 bp were designed (Table 1). These three primers were screened in three genomic DNA samples (in duplicate): VG 9514 (Resistant), TAG 24 (Susceptible) and heterozygous sample made by pooling equal volume of VG 9514 and TAG 24 gDNAs. For each HRM PCR reaction, four independent experiments were carried out. HRM PCR amplification were carried out in 10 µl reaction volume containing 5 µl eva green master mix (2x), 4.8 µl genomic DNA (2 ng/ul) and 0.2 µl primer mix (forward and reverse) of 10 pmol/ul. Thermal cycling was carried out with initial denaturation of 5 min at 95ºC followed by 3-steps cycle of initial denaturation at 94ºC for 30 sec, annealing at 60ºC for 30 sec and extension at 72ºC for 20 sec. HRM of amplified PCR products was done at temperature range from 65ºC to 90ºC with 0.1ºC increase in temperature at every step. Raw HRM data was analysed with the help of inbuilt software in Rotor-Gene Q. For data quality control, PCR amplification was analysed through the assessment of the C_T value and amplification efficiency (Wu et al. 2008). HRM runs with CT value $\leq$ 30 with amplification efficiency >1.4 was considered for analysis. Raw HRM curves were analysed using the HRM analysis module as per Rotor-Gene Q Software 2.3.3.5 TECHNICIAN (Qiagen, Germany). Melting temperature (T_m) difference was calculated manually with the help of negative derivative plots. Homozygotes melt in a single transition whereas, heterozygotes showed multiple melt phases. Genotypes were differentiated based on normalised HRM curve, derivative melt curves and difference plots.

Gene Expression analysis

Total RNA was isolated from the healthy young leaves of VG 9514 and TAG 24 plants using the RNeasy Plant Mini Kit followed by DNase I (RNase free) treatment. The quality and quantity of isolated RNA was

checked by electrophoresis on 1% agarose gel and A_{260}/A_{280} absorbance measurement in a Nanodrop spectrophotometer (Thermo-Fisher Scientific Inc., USA). cDNA was prepared by reverse transcription of 1.2 µg of total RNA using high-fidelity reverse transcriptase and oligo-dT primer supplied in a Prime Script^{TM} Hi-fidelity RT-PCR kit. Next, 4 µl of the 1/5-diluted cDNA was added to the 20 µl reaction mix containing 1X SYBR mix and 0.2 µM of each primer pairs (Table 1). Quantitative real time PCR (qRT-PCR) analysis of the set of genes including *ACT2* as housekeeping (Morgante et al. 2011) was performed using premix *Ex-Taq* DNA polymerase (Takara Bio Inc., Japan) in Rotor-Gene Q (Qiagen, Germany). Reactions were carried out in triplicates from four independent samples. The two-step PCR program consisted of 15 s of initial denaturation at 94°C followed by 40 cycles of 5 s at 94°C and 31 s at 60°C, with a final amplification for 1 min at 60°C. The amplification fluorescence signals were acquired at 60°C in each set of 40 cycles. Specificity of target genes amplification was checked by melting curve analysis at 60 ˚C to 90˚C followed by electrophoresis separation of the same PCR products on 2.5% agarose gel. Transcription levels were expressed as a fold change relative to the control samples, as calculated by the comparative ($2^{-\Delta\Delta Ct}$) method (Livak and Schmittgen 2001).

Analysis of evolutionary significance of the cloned *R*-gene based on K_a/K_s ratio

All the homologous (including homeologous) genes were identified based on blast search in Peanutbase from three available genomes of *Arachis*. Pairwise alignment was done using clustal-omega (Sievers et al. 2011) and converted to AXT format. AXT file was used as input to K_aK_s Calculator 2.0 (Wang et al. 2010) for pairwise calculation. P-Value (Fisher) value for K_a/K_s was computed by Fisher exact test. The cloned rust resistance gene (*VG9514-Rgene*) was used as query against each target homologues. Phylogenetic relationship among the homologues was carried out by maximum likelihood method with 1000 bootstrap. Phylogeny tree was constructed by IQ-tree (Nguyen et al. 2015) and alignment was performed using MAFFT (Katoh and Standley 2013). Model Finder (Kalyaanamoorthy et al. 2017) was used to select the best-fit model using Bayesian Information Criterion (BIC). The best-fit evolutionary model according to BIC was TVM+I.

Homology modeling of VG9514-R gene from groundnut

Modeller, a modeling program (Sali and Blundell 1993) was used to perform homology modeling of VG9514-R gene. Amino acid sequences for both the R-gene sequence of VG 9514 (MK791522) and TAG 24 (MK791523) were aligned and residue differences were noted by aligning them using Clustal omega (Sievers et al. 2014). A BLASTp (Altschul et al. 1990) search against PDB database (Berman et al. 2000) was performed to identify suitable templates for VG9514-R protein. Two top hit structures in the PDB were shortlisted as potential templates and aligned with VG9514-R protein sequence for modeling using the multiple templates modeling approach. Secondary structure was predicted using PSIPRED (McGuffin et al. 2000) and nonaligned regions were removed from the alignment. This alignment was used as query to Modeller v9.12 software and generated 20 models and 5 loop refinement runs per model. A top model was selected based on the DOPE score and energy minimized using OPLS3 force field. Selected model

was validated using MolProbity (Christopher et al. 2018) and ProSA webservers (Wiederstein and Sippl, 2007). Motifs in NB-ARC domain were indicated as per the suggestion from van Ooijen et al. (2008).

Results

Cloning and Sequencing of the targeted *R*-gene

Using specific primer pairs (Table 1) for four different fragments (R1, R2, R3, R4) of the target *R*-gene, the genomic regions were amplified from both VG 9514 (resistant) and TAG 24 (susceptible) genotypes. The R1 fragment was amplified at 62ºC annealing temperature and has produced 866 bp PCR product. While, the PCR product from R2, R3 and R4 fragments were 1115 bp, 915 bp and 1281 bp, respectively (Fig 1a,b). There was no length variation of any PCR fragments between the resistant and susceptible genotypes. The whole sequence (3998 bp) was later extracted from these four fragments. A sequence for *VG9514-Rgene* was submitted in Genbank as an accession MK791522. The other sequence from susceptible parent (TAG 24) was mentioned in Genbank as MK791523. Since the *VG9514-Rgene* sequence was amplified from genomic DNA, it had 31 bp 5' UTR, 73 bp 3' UTR and 3066 bp ORF. The gene comprised of 8 exons and 7 introns. There were no large introns within the gene. The length of introns varied between 78 bp (intron 3) to 189 bp (intron 7). Among the eight exons, exon 3 was found with the longest length 938 bp followed by exon 2 (624 bp). The length of exons varied between 71 bp (exon 1) to 938 bp (exon 3). The translated protein had 1021 amino acids with three conserved domains, TIR domain (pfam 01582), NBS/NB-ARC domain (pfam 00931) and LRR domain (pfam 13855). Based on the alignment of both resistant and susceptible version of this R-protein, non-synnonymous mutations were found in all the three domains (Fig 2).

Identification of SNPs and their mapping

Multiple alignments between each respective fragment from VG 9514 and TAG 24 revealed no SNPs in R1, 39 SNPs in R2, 51 SNPs in R3 and 86 SNPs in R4 fragment. There were many non-synonymous mutations in R2 fragment which mostly correspond to both TIR and NBS region of R-protein. Based on amino acid sequence comparison, it was found that TIR domain had four changes in amino acid sequence, 18 changes in NBS domain and 11 changes in LRR domain (Fig 2). Some of these non-synonymous regions were chosen to develop six PCR based SNP markers. Of the six PCR-based SNP markers, three showed reproducible polymorphism between VG 9514 and TAG 24. In all the three cases, VG 9514 produced positive amplification, while TAG 24 had no bands (Fig S1). All these three SNP-based PCR markers were used for genotyping all 164 RILs and had no segregation distortion (Table 2). They all co-segregated with the rust resistance in RILs. Further linkage mapping placed all three SNP-based PCR markers (FR4, FR1 and FR6) in the same place of A03 linkage group in between the SSR markers, FRS 72 and SSR_GO340445 within 0.6 cM interval (Fig 3). A representative figure for co-segregation of FR6 markers were shown in Figure S2. Thus, co-segregation of marker data of these gene-based PCR markers and positioning within the identified map interval confirmed the candidature of *VG9514-Rgene* for rust resistance in cultivated groundnut.

Validation of SNP-based PCR markers

To validate the strong association of these PCR-based SNP markers, 11 rust resistant cultivated groundnut genotypes and 11 susceptible genotypes were screened with these allele specific diagnostic markers. In all cases, these three markers had proper one to one relation with the resistant/susceptible genotypes. In all the 11 resistant genotypes, the diagnostic allele-specific markers amplified the PCR product, while no amplification was noticed in all the susceptible genotypes (Table 3).

HRM analysis of the SNP region in the identified *R*-gene

Agarose gel electrophoresis of HRM qPCR products of all three SNP loci showed single band with no primer-dimers (Fig S3). The normalized HRM curve, derivative melt curves and difference plots distinguished resistant and susceptible genotypes clearly for all the three tested HRM-primer pairs. The primer pair HRM-FR1 and HRM-FR6 clearly differentiated resistant and susceptible parent but not its heterozygotes. The HRM marker HRM-FR4 distinguished the three genotypes namely, homozygous resistant, heterozygote and homozygous susceptible due to difference in shape and shift in melt curves (Fig 4a&b; Table 4). The melting temperature difference in normalized HRM curve and derivative melting curve of these HRM PCR products confirmed the presence of SNPs and allelic difference between resistant and susceptible parents.

Identification of homologous sequence in *Arachis* genome

When the cloned *R* gene (*VG9514-Rgene*) sequence was used in BLASTn against *A. duranensis* genome, a single hit in A03 chromosome was found. This is the same gene *Aradu.Z87JB* which was proposed in our earlier report (Mondal and Badigannavar 2018). It had 18 bp deletion in intron-2 compared to the isolated *VG9514-Rgene*. In the genome of *A. ipaensis*, a single gene *Araip.0R3VU* was found to be homologous with the isolated *R* gene. *Araip.0R3VU* had 3 bp deletion in intron-3, 3 bp insertion in exon-5, 11 bp insertion in intron-7 and 3 bp deletion in exon-8 and thus found to be non-functional. In the cultivated groundnut genome, four homologous genes were detected. The gene *Arahy.GFGJ54* was found in chromosome 13 (B03). *Arahy.GFGJ54* was found to be the same gene as found in *A. ipaensis* (*Araip.0R3VU)* and had multiple indels like in *Araip.0R3VU*. While three multiple genes *Arahy.6DCA5*, *Arahy.R8KUIR* and *Arahy.ZZ0VZ9* were detected in chromosome 3 (A03) of *A. hypogaea*. *Arahy.6DCA5* had two small indels: a 6 bp insertion in exon-7 and a 15 bp deletion in exon-8 and thus it evolved as non-functional. *Arahy.R8KUIR* had 4 bp deletion in intron-2, 3 bp deletion in intron-3 and 11 bp deletion in intron-7. The gene *Arahy.ZZ0VZ9* had two deletions as compared to *VG9514-Rgene*: 18 bp deletion in intron-2 and 3 bp deletion in intron-3 (Fig 5). The position of 18 bp deletion in *Arahy.ZZ0VZ9* is same as in *Aradu.Z87JB*. With the hope to get polymorphism based on these identified indels in these homologous genes, we designed primer pairs (Table 1) around these indels and screened polymorphism between parents (VG 9514 and TAG 24).

Identification of polymorphic *R* gene specific Indel markers and their mapping

Screening of these new InDel markers revealed that all are polymorphic between VG 9514 (resistant) and TAG 24 (susceptible). Of the five Indels, three had amplified bands as per the InDel size and found polymorphic (Table S1; Fig S4). These three are InDel 3, InDel 11 and InDel 15. Further these InDels are corresponding to the three homologous *R*-genes, *Arahy.ZZ0VZ9*, *Arahy.R8KUIR* and *Arahy.T6DCA5*, respectively. In every case, multiple bands were amplified due to the presence of homologous genes in the groundnut genome (Fig S4&S5). Based on the amplification profile in the entire RIL population, a linkage map was constructed. These four polymorphic InDel markers were mapped at the distal region of A03 chromosome and they were tightly linked to each other. Further, they were mapped almost 19.3 cM away from the cloned *VG9514 R-gene* in the same chromosome (Fig 3).

Analysis of non-synonymous mutations within the PCR-based SNP marker of *R* gene

Based on the sequence comparison between resistant and susceptible amplified genes, we had identified several SNPs and later developed polymorphic SNP markers. Three such markers showed reproducible polymorphism and thus validated the position of mutations within the R-gene. The SNP marker FR4 was designed for 'C to T', 'G to T' and 'T to C' mutations within exon-3. This exon-3 code for some parts of TIR and NBS domain, where non-synonymous mutations changed 'lysine to asparagine', 'serine to leucine' and 'valine to serine' in the distal portion of TIR domain which is in close proximity of hhGRExE and walker A/P-loop motif of NBS domain. Another SNP marker FR1 was designed for 'T to A' and 'C to G' in the same exon-3. These mutations corresponded to the amino acid changes of 'isoleucine to tyrosine' and 'glutamine to glutamate' between RNBS-A and Walker B motif of NBS domain. The third polymorphic SNP marker FR6 harbored 'G to T' and 'A to G' in exon-7. Such point mutations lead to change of aspartate to cysteine in LRR domain of the R-gene of susceptible version (Table 5).

Gene expression of *R*-gene and other pathogenesis-related genes

To check the transcript level of this *R*-gene and other associated pathogenesis related genes (mentioned in Mondal and Badigannavar 2018), we have designed gene specific primer pairs for qRT-PCR analysis (Table 1). No amplification was observed in negative control (-RT). Comparison of transcript in healthy young leaves of VG 9514 and TAG 24 revealed a 4 fold enhanced expression of the cloned R-gene in the resistant parent. The other pathogenesis related (PR) genes (mainly glucan endo-1,3 β glucosidases) were also had higher expression in the resistant parent (Fig 6). Of the three PR genes, transcript level of *Aradu.1WV86* was quite low (Ct value >34) and thus was not considered for analysis. Both the other PR genes, *Aradu.T44NR* and *Aradu.NG5IQ* had higher level of expression in the resistant plant.

Evolutionary significance of the cloned *VG9514-R-gene* compared to other homologues

According to the homology standard based on BLAST, it was found that three genes were duplicated in tandem on distal portion of chromosome A03 of *A. hypogaea*. These three genes were *Arahy.ZZ0VZ9*, *Arahy.T6DCA5* and *Arahy.R8KUIR* and they are paralogs of the identified *VG9514-R-gene*. However, no orthologous relationship was noticed between these three genes and *Aradu.Z87JB* (Fig 7). The phylogenetic tree showed that there is a stronger homologous relationship between *Araip.0R3VU* and

Arahy.GFGJ54 (bootstrap 100), but lower relationship between *Arahy.ZZ0VZ9* and *Aradu.Z87JB* (bootstrap less than 70). Taken together, there was orthologous relationship between *VG9514-Rgene* and *Araip.0R3VU*. In addition, there was homoeologous relationship between *Arahy.GFGJ54* and three orthologous genes (*Arahy.ZZ0VZ9*, *Arahy.T6DCA5* and *Arahy.R8KUIR*) of *VG9514-Rgene*. Based on these leads, K_a/K_s values were determined from all possible combinations of *VG9514-Rgene* with one homeologous (*Arahy.GFGJ54*) and three orthologous (*Arahy.ZZ0VZ9*, *Arahy.T6DCA5* and *Arahy.R8KUIR*) genes in *A. hypogaea* genome. The identified *R*-gene displayed >1 K_a/K_s value with respect to *Arahy.GFGJ54* and *Arahy.R8KUIR*. With respect to two other paralogous genes, K_a/K_s values were found close to 1 (Table 6). These results indicated that these paralogous genes underwent positive selection or relaxed purifying selection.

Structure model of VG9514 R-protein and localization of mutations

Based on PDB search of VG9514-R protein sequence, a crystal structure of TIR domain from *Vitis rotundifolia* was found homologous to TIR domain of the cloned R-protein. This top hit (having 16% coverage of VG9514-Rprotein) was PDB:600W, is a crystal structure of the TIR domain from the grapevine disease resistance protein RUN1 in complex with NADP+ and Bis-Tris (Horsefield et al. 2019). Another potential hit was found with PDB:6J5T (with a percentage identity 46%), a Cryo-EM structure of plant NLR resistosome from *Arabidopsis thaliana* (Wang et al. 2019). (Fig 8). Validated model of VG9514-Rprotein was supported by having more than 99% of the amino acid residues in the allowed region of Ramachandran plot (Fig S6a). Further, ProSA program checked the quality of the model, with a good Z-score (-8.28) and not much deviation of energy was observed in the proposed model (Fig S6b). Amino acid differences from R-protein of TAG 24 sequence were mapped to the modelled structure of VG9514-Rprotein. There were only three residue differences observed at the end of the TIR domain, whereas most of the SNPs were in the NB-ARC and LRR domain of the modelled R-protein.

Mapping of non-synonymous mutations (found in R-gene sequence of TAG 24) identified that the mutations were scattered throughout the whole VG9514 R-protein model (Fig 8). At the TIR domain three specific non-synonymous mutations K252N, S253L, V254S which are positioned in close proximity of hhGRExE motif of NBS domain. Another important non-synonymous mutation E268Q was revealed within hhGRExE motif. In the RNBS-A motif of NB-ARC domain, an important non-synonymous mutation Y309F was deciphered (Fig 8). Other three amino acid changes Q349H, I350Y, Q351E were observed in between RNBS-A and Walker B motif of NB-ARC domain. An interesting non-synonymous mutation K380N was observed in close proximity to Walker B motif. Similarly K420E was placed in between RNBS-B/sensor I and RNBS-C motif. Another K457E mutation was found in GLPL motif of the NBS domain of this R-protein in TAG 24. Two other mutations Y527W and R528I were at the end of RNBS-D motif of NBS domain. At the MHD motif, an I579T mutation was also deciphered in the susceptible R-protein in TAG 24 (Fig 8). Many other mutations were also noticed in LRR domain which is responsible for recognition of avirulence (AVR) protein in pathogen.

Discussion

Genetic resistance provided by the crop cultivars is an environmentally viable approach to combat the yield losses in plants due to deadly diseases. Often such resistances are not available in primary gene pool. Disease resistance was then introgressed into cultivated species from wild crop relatives. Such a dominant rust resistance gene was introgressed into cultivar Co1 (*A. hypogaea* L.) through wide hybridization in hexaploid pathway from *A. cardenasii* (Varman 1999; Mondal et al. 2007). The resultant rust resistant breeding line VG 9514 was thus evolved and later used for making a mapping population to tag this dominant rust resistance gene (Mondal et al. 2007; 2012 a, b). Usage of several chromosome specific SSR markers later fine mapped this rust resistance loci within a 1.2 cM map interval in chromosome A03 of cultivated groundnut (Mondal and Badigannavar 2018). Gene ontology and specific protein function analysis of the ORFs present in this fine mapped region revealed a TIR-NBS-LRR domain containing *R*-gene that may have functional genetic polymorphism.

Cloning of the candidate rust resistance gene in groundnut and its validation

The donor *A. cardenasii* is a member of A-genome of section *Arachis*. Based on the availability of genome information of another A-genome member *A. duranensis* (https://peanutbase.org/gbrowse_aradu1.0), specific primer pairs were designed to PCR amplify the identified TIR-NBS-LRR gene from both resistant and susceptible parents. Genomic DNA amplification using those specific primer pairs revealed 3998 bp length of this *R*-gene in both resistant and susceptible genotypes and thus no InDel was found between them. Both the sequences have considerably high sequence identity of 95.35%. However, they shared lot of single nucleotide polymorphisms which generated non-synonymous mutations in the TIR, NB-ARC and LRR domains of R-protein sequence. High frequency of presence of SNPs was also reported through whole genome re-sequencing analysis of diverse groundnut genotypes (Pandey et al. 2017; Clevenger et al. 2017). We used those non-synonymous SNP regions in the identified *VG9514-Rgene* and generated six SNP-based PCR primer pairs (Table 1) for studying polymorphism and genetic mapping. Three of them showed reproducible polymorphism (presence vs. absence) between resistant and susceptible parents.

Genetic mapping of these polymorphic SNP markers (FR4, FR1 and FR6) positioned them in between the identified fine mapped region of FRS 72 and SSR_GO340445 markers in chromosome A03 that harbored the consensus major rust QTL (Mondal and Badigannavar 2018). Thus, the genetic mapping of these SNP markers revealed complete co-segregation of them with rust resistance and position of these markers was located within the fine mapped region without any recombination between them (Fig 3). Such non-synonymous SNP regions were later used for high resolution melting (HRM) analysis through quantitative PCR technique and examined for their allelic difference. All these three could differentiate both the genotypes clearly through HRM reaction. The particular primer pairs of HRM-4 distinguished even the heterozygous plant from both the parent and thus displayed its usage in marker assisted selection (MAS) (Fig 4a & b). Identification of such gene based marker in the present study will greatly helpful in the introgression of rust resistance in high yielding cultivars without linkage drag of undesirable agronomic characters from wild crop relatives. It was interesting to know the gene expression pattern in presence of rust in this two contrasting genotypes VG 9514 and TAG 24. We could not succeed to get a

rust infection in artificial growth chamber condition and thus we proceeded with comparative gene expression analysis between them in normal plant growth condition. A significantly higher basal expression of the identified *R*-gene was noticed in VG 9514 probably due to presence of a strong promoter element. Further two other pathogenesis related genes (*Aradu.NG51Q* and *Aradu.T44NR*) also had higher expression in resistant genotype compared to susceptible one (Fig 6).

Mapping of other homologous R-gene in cultivated groundnut genome and evolutionary significance of the isolated *VG9514-R gene*.

Sequence analysis revealed several homologous *R*-gene sequences in cultivated groundnut genome in comparison to the isolated gene *VG9514-Rgene*. In the A genome, we could find *Arahy.ZZ0VZ9* and *Aradu.Z87JB*. Similarly *Araip.0R3VU* and *Arahy.GFGJ54* were found in B genome. According to the homology standard based on BLAST method, it was found that *Arahy.ZZ0VZ9* had two paralogs, *Arahy.T6DCA5* and *Arahy.R8KUIR* on chromosome A03. When the isolated *VG9514-Rgene* sequence was used as query sequence for BLAST in *A. hypogaea* genome in Peanutbase, we could detect several InDels within these paralogous genes. The size of InDel ranged between 3 to 15 bp (Fig 5). Vishwakarma et al. (2017) found 515,223 InDels over the groundnut genome when compared two diploid *A. hypogaea* progenitor's genome sequences. Such small sized InDels in paralogous genes probably occurred in the process of chromosome doubling and tetraploidization during the domestication of cultivated groundnut. Based on their polymorphism between VG 9514 and TAG 24, these InDels were used in genotyping of RILs and genetic mapping. Mapping revealed all these three InDel markers were mapped in close proximity to each other at the distal place of chromosome A03 and positioned themselves around 19.3 cM apart from the identified *VG9514-Rgene* (Fig 3). Thus, the sequence analysis and mapping of these InDel markers revealed tandem duplications of R-genes in the genome of cultivated groundnut. These genes are non-functional against the rust pathogens and thus introgression event of a new R-gene from *A. cardenasii* in other locus of the same chromosome imparted resistance in VG 9514. Shirasawa et al. (2018) had detected a common 2.7 Mb genome fragment (131.9–134.6 Mb in A03) introgression from *A. cardenasii* in four rust resistant genotypes, RIL97 (TAG 24 x GPBD 4), GPBD 4, ICGV 86855 and VG 9514. The present study revealed the sequence of this novel resistance gene (*VG9514-Rgene*) in groundnut and also mapped three other paralogues of the cloned *R*-gene at the distal portion of chromosome A03.

Based on K_a/K_s analysis, it was found that *VG9514-Rgene* had positive selection during the process of evolution compared to *Araip.0R3VU*, *Arahy.GFGJ54* and *Arahy.R8KUIR*. Whereas with respect to the paralogous genes *Arahy.ZZ0VZ9* and *Arahy.T6DCA5*, the K_a/K_s value was almost close to 1 and thus they have been conserved along with the introgressed gene *VG9514-Rgene* in chromosome A03. Selection pressure on protein coding sequence is normally assessed by the ratio of the non-synonymous substitution rate (K_a) to the synonymous substitution rate (K_s) (Hughes and Nei 1988). If $K_a/K_s > 1$, positive selection is assumed to have occurred during the evolution of the sequence. Typically natural selection shapes the evolution of species through one of the two mechanisms: purifying selection, which promotes the conservation of existing phenotypes, and positive selection, which favors the emergence of new phenotypes (Vallender and Lahn 2004). Taken together, it is obvious that the *VG9514-Rgene* was

introduced through an introgression event from *A. cardenasii* and this gene has undergone positive selection that emerged as a novel dominant rust resistance in groundnut. It is postulated here that resistance genes at the distal portion of chromosome A03 had undergone tandem duplications, possibly through transposon activities. Further, both intergenic and intragenic gene conversions and unequal crossing-over occurred during repeated backcrossing in hexaploid pathway of gene introgression event from wild species, might have evolved the functional rust resistance in VG 9514.

Implications of the non-synonymous mutations in respect to homology modelling of VG9514-R protein

There are two major subfamilies of plant NBS-LRR proteins, defined by the presence of Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) motifs in the amino-terminal domain. Although TIR-NBS-LRR proteins and CC-NBS-LRR proteins are both involved in pathogen recognition, the two subfamilies are distinct both in sequence and in signalling pathways (McHale *et al*., 2006). The isolated *VG9514-Rgene* belongs to TIR-NBS-LRR category. At the FR4 SNP marker region three non-synonymous amino acid changes K252N, S253L, V254S were noticed in the close upstream of hhGRExE motif of NBS domain (Fig 2). All these mutations were present in TIR domain and may involve in transducing the conformational changes due to ATP binding in NBS domain (in response to effector molecule binding) to TIR domain for the signalling event of program cell death (Swiderski et al. 2009). Zhang et al. (2017) studied in detail the importance of self-association of interfaces in plant TIR domains of disease resistance proteins L6, RPS4 and SNC1. They found that the impairment of interaction in AE interface (interface formed by αA and αE helices of TIR domain) and DE interface (interface formed by αD and αE helices of TIR domain) are required for self association and cell death signalling in TIR domain of plants R protein. It was shown that R164 and K200 residues played important roles in stabilizing the DE interface dimer of L6-TIR and mutation of either residue suppresses L6-TIR self-associations and autoactivity (Bernoux et al. 2011). In the SNP region of FR1 marker, Q349H, I350Y, Q351E mutations occurred between RNBS-A and Walker B motif of NBS domain in the susceptible version of the *VG9514-R*gene product. For FR6 SNP marker, a significant change from aspartate to cysteine (D768C) was taken place in LRR domain of this isolated *R* gene. This particular aspartate residue is a part of '**L**DLS**D**' sequence where both the L and the terminal D are conserved in L6 R-protein. The LRR domain is a common motif found in many proteins, and it is involved in protein-protein interactions and effector binding (Jones and Jones 1997). In many NBS-LRR proteins, the putative solvent exposed residues of LRR domain show significantly elevated ratios of non-synonymous to synonymous substitutions, indicating that diversifying selection has maintained variation at these positions. The LRR domain is involved in determining the recognition specificity of several R proteins (Hwang and Williamson 2003). NBS domain is responsible for specific binding and hydrolysis of ATP that is thought to result in conformational changes in response to effector molecule binding at LRR domain. This conformational change in NBS domain has to transduce to TIR domain to regulate downstream signalling for programmed cell death, a hypersensitive response against the obligate parasite in resistant genotype (Mestre and Baulcombe 2006; van Ooijen et al. 2008; Sueldo et al. 2015). Based on the protein sequence homology, van Ooijen et al. (2008) proposed nine conserved motifs in NB-ARC domain. These motif sequences are conserved across the R-protein and alternations of these sequences may interfere with structure-functions relationship of the R-gene. Based on mapping of

mutations in susceptible R-protein of TAG 24, it was found that three mutated amino acid residues are lying in these conserved domains: E267Q in hhGRExE motif, Y309F in RNBS-A motif and I579T in MHD motif (Fig 8). We proposed here the critical role of these mutations in providing rust resistance function in groundnut.

In conclusion, the present investigation has isolated a TIR-NBS-LRR domain containing dominant rust resistance gene in groundnut for the first time in literature and developed gene-based SNP markers for marker assisted selection. The gene based markers and their segregation pattern further proved its co-segregation with the rust resistance and its candidature. This novel *VG9514-Rgene* was distinct from the other homologous R-gene in the tetraploid genome and found to undergo positive selection. Homology based protein modelling study later revealed that the non-synonymous mutations in susceptible version of this gene in TAG 24 were positioned in all the domains of the R-protein. Of them, three were in conserved motifs (hhGRExE, RNBS-A and MHD motif) of NB-ARC domain which is responsible for ATP binding, hydrolysis and transduce the ATP-binding-induce conformational change to TIR domain for program cell death.

Declarations

Acknowledgement

We are thankful to Head, Nuclear Agriculture and Biotechnology Division, BARC for constant encouragement and constructive comments during the work. We also extend our special thanks to Associate Director, Bioscience group, BARC for timely support. Authors duly acknowledge Dr. Swathi Kota, MBD, BARC for extending her laboratory facility for transformation and plasmid isolation during this work. Efforts made by T. Chalapathi and Sujit Tota at field experiments are duly acknowledged.

Data Availability Statements

All data supporting the findings of this study are available within the paper and within its supplementary data published online. Sequence information of MK791522 (isolated R-gene sequence from VG 9514 genotypes) and MK791523 (isolated R-gene sequence from cv. TAG 24) is available at NCBI (https://www.ncbi.nlm.nih.gov).

Conflict of Interest Statement

Authors do not have any conflict of interest

Conflict of Interest

Authors declare no conflict of interest.

Funding: The research was funded from 12th five year project, Bhabha Atomic Research centre, Government of India.

***Conflicts of Interest/Competing Interests**:* Authors declare no conflicts of interest

***Availability of data and materials**:* All data are included in the manuscript. The Genbank accession (MK791522 and MK791523) was deposited in NCBI.

***Code Availability**:* The study involved usage of all freely available software program.

***Authors' Contribution**:* Included

***Ethics Approval**:* The study was conducted by following all ethical standards for laboratory experiments.

***Consent to Participate**:* Appropriate consent was taken from institute

***Consent for Publication**:* Appropriate consent was taken from institute

Author contribution statement

SM conceptualized the work, developed mapping population, executed and supervised the experiments, wrote the draft and edited the contributed draft version from different authors. KMS did homology modelling of R-protein, KaKs analysis and drafted the manuscript. AR performed real time PCR and high resolution melting analysis. DSK and PB performed genotyping work with SNP and other markers. SH helped in analyzing sequence data for evolution genetics and identified orthologs of cloned R-gene. RS conceptualized the protein modelling work and supervised it. AMB conceptualized the work, developed mapping population and edited manuscript.

Note on ethical standards

All the experiments (both field and laboratory) were conducted as per the standard ethical guidelines. The study did not involve any experiments on animals and human volunteer.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403-410
2. Arora S, Steuernagel B, Kumar G, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Cob E, Periyannan S., Singh N, MdHatta MA, Athiyannan NK, Cheema J, Yu G, Kangara N, Ghosh S, Szabo L, Poland J, Bariana H, Wulff B (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nature Biotechnol 37:139-143
3. Berman HM, Westbrook J, Feng Z, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28(1): 235-242.
4. Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN (2011) Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signalling and autoregulation. Cell host Microbe 9(3):200-211

5. Clevenger J, Chu Y, Chavarro C, Agarwal G, Bertioli DJ, Leal-Bertioli SCM, Pandey MK, Vaughn J, Abernathy B, Barkley N, Hovav R, Burow M, Nayak SN, Chitikineni A, Isleib T, Holbrook C, Jackson SA, Varshney RK, Ozias-Akins P (2017) Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. Mol Plant 10:309-322
6. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164-1165
7. Drenkard E, Richter BG, Rozen S, Stutius LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol 124:1483-1492
8. Flor HH (1942) Inheritance of pathogenicity in *Melarnpsora lini*. Phytopathol 32:653-669
9. Gayathri M, Shirasawa K, Varshney RK,Pandey MK, Bhat RS (2018) Development of *Ah*MITE1 markers through genome-wide analysis in peanut (*Arachis hypogaea*). BMC Res Notes 11:8-13
10. Guido M., VandenA, Jan AL, Van K, Pierre JGM, De W (1992) Molecular analysis of the avirulence gene avr9 of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant J 2:356-369
11. Horsefield S, Burdett H, Zhang X, Manik MK, Shi Y, Chen J, et al. (2019) NAD+ cleavage activity by animal and plant TIR domains in cell death pathways. Science 365:793-799
12. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170
13. Hwang CF, Williamson VM (2003) Leucine-rich repeat-mediated intramolecular interactions in nematode recognition andcell death signaling by the tomato resistance protein Mi. Plant J 34:585-593
14. Jiang R, Li J, Tian Z, Du J, Armstrong M, Baker K, Lim JTY, Vossen JH, et al. (2018) Potato late blight field resistance from QTL *dPI09c* is conferred by the NB-LRR gene J Exp Bot 69:1545-1555
15. Johal GS, Briggs SP (1992) Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258:985-987
16. Jones DA, Jones JDG (1997) The role of leucine-rich repeat proteins in plant defenses. Adv Bot Res 24:90-167
17. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587-589
18. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780
19. Khedikar YP, Gowda MVC, Sarvamangala C, Patgar KV, Upadhyaya HD, Varshney RK (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea*). Theor Appl Genet 121:971-984
20. Klymiuk V, Yaniv E, Huang L, Raats D, Fatiukha A, Chen S, Feng L, Frenkel Z, Krugman T, et al. (2018) Cloning of the wheat Yr15 resistance gene sheds light on the plant tandem kinase-pseudokinase

family. Nature Comm 9:3735 doi: 10.1038/s41467-018-06138-9.

21. Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172-175.
22. Kourelis J, van der Hoorn RAL (2018) Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. Plant Cell 30:285-299
23. Leal-Bertioli SCM, Cavalcante U, Gouvea EG, Ballén-Taborda C, Shirasawa K, Guimarães PM, Jackson SA, Bertioli DJ, Moretzsohn MC (2015) Identification of QTLs for rust resistance in the peanut wild species *Arachis magna*and the development of KASP markers for marker-assisted selection. G3 (Bethesda) 5:1403-1413
24. Liu Z, Feng S, Pandey MK, Chen X, Culbreath AK, Varshney RK, Guo B (2013) Identification of expressed resistance gene analogs from peanut (*Arachis hypogaea*) expressed sequence tags. J Integ Plant Biol 55:453-461
25. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ Methods 25:402-408
26. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16(4):404-405
27. Mestre P, Baulcombe DC (2006) Elicitor-mediated oligomerization of the tobacco N disease resistance protein. Plant Cell 18:491-501
28. Mondal S, Badigannavar AM, D'Souza SF (2012a) Molecular tagging of a rust resistancegene in cultivated groundnut (*Arachis hypogaea*) introgressed from *Arachis cardenasii*. Mol Breed 29:467-476
29. Mondal S, Badigannavar AM, D'Souza SF (2012b) Development of genic molecular markers linked to a rust resistance gene in cultivated groundnut (*Arachis hypogaea*). Euphytica 188:163-173
30. Mondal S, Badigannavar AM, Murty GSS (2007) RAPD markers linked to a rust resistance gene in groundnut (*Arachis hypogaea*). Euphytica 159:233-239
31. Mondal S, Badigannavar AM (2018) Mapping of a dominant rust resistance gene revealed two R genesaround the major Rust_QTL in cultivated peanut (*Arachis hypogaea*). Theor Appl Genet 131:1671-1681
32. Mondal S, Hande P, Badigannavar AM (2014a) Identification of transposable element markers for a rust (*Puccinia arachidis*) resistance gene in cultivated peanut. Phytopathol 162:548-552
33. Mondal S, Hadapad AB, Hande P, Badigannavar AM (2014b) Identification of quantitative trait loci for bruchid (*Caryedon serratus* Olivier) resistance components in cultivated groundnut (*Arachis hypogaea*). Mol Breed 33:961-973
34. Morgante CV, Guimarães PM, Martins ACQ, Araújo ACG, Leal-Bertioli SCM, Bertioli DJ, Brasileiro ACM (2011) Reference genes for quantitative reverse transcription-polymerase chain reaction expression studies in wild and cultivated peanut. BMC Res Notes4:339

35. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268-274
36. Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD, Vishwakarma MK, Bertoli SL, Liang X, Bertioli DJ, Guo B, Jackson SA, Ozias-Akins P, Varshney RK (2017) Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci Rep 7:40577
37. Patil SH, Kale DM, Deshmukh SN, Fulzele GR, Weginwar BG (1995) Semi-dwarf, early maturing and high yielding new groundnut variety, TAG 24. J Oilseed Res 12:254-257
38. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779-815
39. Schulze-Lefert P, Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance, pathogens host range, and pathogen speciation. Trends in Plant Sci 16:117-125
40. Shirasawa K, Bhat RS, Khedikar YP, Sujay V, Kolekar RM, Yeri SB, Sukruth M, Cholin S, Asha B, Pandey MK, Varshney RK, Gowda MVC (2018) Sequencing analysis of genetic loci for resistance for late leaf spot and rust in peanut (*Arachis hypogaea*). Frontiers in Plant Sci 9:1727
41. Sievers F, Higgins DG (2014) Clustal omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105-116
42. Sievers F, Wilm A, Dineen D, Gibson TJ,Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgin DG (2011) Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Mol Systems Biol 7:539
43. Simpson CE (2001) Use of wild *Arachis* species/introgression of genes into *hypogaea* L. Peanut Sci 28:114-116
44. Song H, Wang P, Li C, Han S, Zhao C, Xia H, Bi Y, Guo B, Zhang X, Wang X (2017) Comparative analysis of NBS-LRR genes and their response to *Aspergillus flavus* in *Arachis.* PLoS One 12:e0171181
45. Stalker HT (2017) Utilizing wild species for peanut improvement. Crop Sci 57:1102-1120.
46. Sueldo DJ, Shimels M, Spiridon LN, Caldararu O, Petrescu AJ, Joosten MHAJ, Tameling WIL (2015) Random mutagenesis of the nucleotide-binding domain of NRC1 (NB-LRR required for hypersensitive response-associated cell death-1), a downstream signalling nucleotide-binding, leucine-rich repeat (NB-LRR) protein, identifies gain-of-function mutations in the nucleotide-binding pocket. New Phytol 208:210-223
47. Sujay V, Gowda MVC, Pandey MK, Bhat RS, Khedikar YP, Nadaf HL, Gautami B, Sarvamangala C, Lingaraju S, Radhakrishnan T, Knapp SJ, Varshney RK (2012) Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 32:773-788
48. Swiderski MR, Birker D, Jones JDG (2009) The TIR domain of TIR-NB-LRR resistance proteins is a signalling domain involved in cell death induction. MPMI 22:157-165

49. Vallender FJ, Lahn BT (2004) Positive selection in the human genome. Human Mol Genet 13:245-254.
50. Van Ooijen G, Mayr G, Kasiem MMA, Albrecht M, Cornelissen BJC, Takken FLW (2008) Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. J Exp Biol 59:1383-1397
51. van der Biezen EA, Jones JDG (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current Biol 8:R226-R228.
52. Varman PV (1999) A foliar disease resistant line developed through interspecific hybridization in groundnut (*Arachis hypogaea*). Indian J Agric Sci 69:67-68
53. Vishwakarma MK, Kale SM, Sriswathi M, Naresh T, Shasidhar Y, Garg V, Pandey MK, Varshney RK (2017) Genome-wide discovery and deployment of insertions and deletions markers provided greater insights on species, genomes, and sections relationships in the genus *Arachis*. Frontier in Plant Sci 8:2064
54. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinfomatics 8:77-80
55. Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, Qi Y, Wang HW, Zhou JM, Chai J (2019) Reconstitution and structure of a plant NLR resistosome conferring immunity. Science 364:eaav5870
56. Wang J, Li H, Zhang L, Meng L (2016) Users' manual of QTL Ici-Mapping. The quantitative genetics group, Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, China, and Genetic Resources Program, InternationalMaize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico, D.F., Mexico.
57. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35(Web Server):
58. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci 27:293-315
59. Wu S, Wirthensohn MG, Hunt P, Gibson JP, Sedgley M(2008) High resolution melting analysis of almond SNPs derived from ESTs. Theor Appl Genet 118:1-14
60. Zhang X, Bernoux M, Bentham AR, Newman TE, Ve T, Casey LW, et al. (2017) Multiple functional self-association interfaces in plant TIR domains. Proc Natl Acad Sci 114(10):E2046-E2052
61. Zhuang W, Chen H, Yang M. Wang J, Pandey MK, Zhang C, Chang WC, Zhang L, Zhang X, Tang R, et al.(2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51:865-876.

Tables

Table 1: List of different primer pairs used for gene cloning, SNP marker development, InDel marker development, high resolution melting analysis and gene expression studies.

Primer name	Primer sequence (5’ – 3’)	Annealing temperature (°C)	Size
R gene cloning Primer pairs			
R1 segment-FP	GGTATACAAGGAATTCACTCCTCAC	62.0 °C	866 bp
R1 segment-RP	GTATATCCCAACCAGAGAGATTGGCGAG	(with Q5 DNA Pol)	
R2 segment-FP	GCAACAATGGAGGGAATCTCTGCTG	64.5 °C	1115 bp
R2 segment-RP	ATCCATTTGGTGGTTAAAGAAACAAGCGA	(with Q5 DNA Pol)	
R3 segment-FP	CTTCAACTTAGTTTTGATGGATTGGAAC	70.0 °C	915 bp
R3 segment-RP	GATGGATGTATGTGCTTAAGCTTTC	(with Q5 DNA Pol)	
R4 segment-FP	CCTTATAAGATGTAGAAAGCTTAAGCACA	67.5 °C	1281 bp
R4 segment-RP	GACATTGAGATGATGAAAGAATACCTG	(with Q5 DNA Pol)	
PCR based SNP primer pairs			
FR1_SNAP-FP	GAGATCCAAGATATATACAATCAAATCC	58.0 °C	142 bp
FR1_SNAP-RP	ATAGACAATCACGAGTCACGCCA	(BRIT *Taq*Pol)	
FR2_SNAP-FP	TATATACAATCAAATCCAGGCTGAG	Non-polymorphic	131 bp
FR2_SNAP-RP	ATAGACAATCACGAGTCACGCCA		
FR3_SNAP-FP	TTTAACCACCAAATGGATTATCG	Non-polymorphic	253 bp
FR3_SNAP-RP	TTACCGTATTTCCAAACATAATAC		
FR4_SNAP-FP	AGAAGCAAATTAGGTTGCAAATCGCT	67.0 °C	132 bp
FR4_SNAP-RP	CATGCCACAAACGCCTACAAC	(BRIT *Taq*Pol)	
FR5_SNAP-FP	GGGATGCAATCTCGTGTGGTAG	Non-polymorphic	87 bp
FR5_SNAP-RP	CATGCCACAAACGCCTACAAC		
FR6_SNAP-FP	GCGTGTATGATCTTGACCTAAGTGA	59.6 °C	131 bp
RR6_SNAP-RP	CTTGTTTATGAAATCAGGTACTCTAAC	(BRIT *Taq*Pol)	
R gene specific Indel primer pairs			
InDel3-FP	CATCCAAAGATTGGAATAAGTG	50.0 °C	206 bp
InDel3-RP	AAACTGCATACCGTATTTCCA	(Go *Taq* DNA Pol)	
InDel-18-FP	ATATAAGACAACCTAAAACAAAGAG	55.4 °C	189 bp
InDel-18-RP	GCAACAATGGAGGGAATCTC	(Go *Taq* DNA Pol)	
InDel-15-FP	TGGGATGGAAAGCAGCATGA	60.0 °C	194 bp

		(Go *Taq* DNA Pol)	
InDel-15-FP	GACATTGAGATGGATGAAAGAATACC		
InDel-11-FP	GTTGTTGCCATTATTGCTGCTCTTG	60.0 °C	227 bp
InDel-11-RP	CAAAGGGCAATGCCAATCCA	(Go *Taq* DNA Pol)	
InDel-4-FP	GGATATACAAAATAAGTAAGTAATCTTTC	60.0 °C	287 bp
InDel-4-RP	CACACGAGATTGCATCCCAAC	(Go *Taq* DNA Pol)	
SNP containing HRM primer pairs			
HRM4-FP	CAGGCAGGAGAATGGAGAGATTG	62.0 °C	101 bp
HRM4-RP	GAGATTGCATCCCAACCAAC		
HRM1-FP	GTCAAATTACAAATGGAGAGATCCAAG	60.0 °C	108 bp
HRM1-RP	CATTATCTAGAACAATCAGAGCCTTTTGA		
HRM6-FP	GTGGAGGCTTTCTGTGGCTCGC	62.0 °C	91 bp
HRM6-RP	GGATGATATGGCCTCAGGGATAG		
Gene expression analysis primer pairs			
Rust Rgene-FP	AGGTGAGGGGTGAATTGGTT	54.0 °C	160 bp
Rust Rgene-RP	AGTACCACACGTATTCCCCA		
Aradu.T44NR-FP	ATGGCGACCACCATGATGAA	54.0 °C	153 bp
Aradu.T44NR-RP	GCAAAGGCCATTGGCTTGAA		
Aradu.1WV86-FP	TGACAGCTACTATCAGCGCA	54.0 °C	111 bp
Aradu.1WV86-RP	ATTGGTGGGTATCTGCAGCC		
Aradu.NG5IQ-FP	GCAAACTGTGCCGCTATTCA	54.0 °C	159 bp
Aradu.NG5IQ-RP	AGGATCCCTTGTAGTAATTGTAGCA		
ACT2-FP	GAGCTGAAAGATTCCGATGC	54.0 °C	178bp
ACT2-RP	GCAATGCCTGGGAACATAGT		

Table 2: Segregation of allelic data of new SNP and Indel marker in mapping population

	Marker name	VG 9514 allele	TAG 24 allele	χ^2 value	P value	Missing allele
1.	FR1	85	79	0.22	0.64	0
2.	FR4	85	79	0.22	0.64	0
3.	FR6	85	79	0.22	0.64	0
4.	Indel 15	92	66	4.28	0.04	6
5.	Indel 11	83	75	0.41	0.52	6
6.	Indel3	74	86	0.9	0.34	4
7.	*Ah*TE1220 #	81	83	0.02	0.87	0

This particular *Ah*TE markers were synthesized based on the reports in Gayathri et al. (2018)

Table 3: Details of groundnut genotypes used for SNP based PCR marker validation study

	Genotypes	Pedigree	Rust score		Marker reaction		
			2016	2018	FR4	FR1	FR6
1.	VG 9514	Co 1 x *A. cardenasii*	1	2	+	+	+
2.	DTG 58	TG 26 x Mutant 28-2	2	3	+	+	+
3.	TG 60	TG 26 x Mutant 28-2	2	2	+	+	+
4.	TG 66	TAG 24 x VG 9514	2	3	+	+	+
5.	TG 69	Mutant of TG 66	3	3	+	+	+
6.	TG 75	TG 26 x R 9227	3	3	+	+	+
7.	GPBD 4	KRG 1 x ICGV 86855	2	2	+	+	+
8.	GFDS 272	Not Known	2	1	+	+	+
9.	GPBD 5	TG 49 x GPBD 4	2	2	+	+	+
10.	G 2-52	Mutant of GPBD 4	2	3	+	+	+
11.	JL 776	(ICGV 92069 x ICGV 93184) x ICGV 98300	1	2	+	+	+
12.	TAG 24	TGS 2 x TGE 1	8	9	-	-	-
13.	ICGV 91114	ICGV 86055 x ICGV 86533	8	8	-	-	-
14.	TMV 10	Natural mutant from Argentine	8	8	-	-	-
15.	JL 24	Selection from EC 94943	9	9	-	-	-
16.	SG 99	ICGV 86529 x ICGV 87160	7	8	-	-	-
17.	Chico	Introduction from USA	8	9	-	-	-
18.	Girnar 1	'X'-14-4-B-19B x NCAc 17090	8	8	-	-	-
19.	SB XI	Ah 4213 x Ah 4354	9	9	-	-	-
20.	Co 3	VG 55 x JL 24	7	8	-	-	-
21.	GG 7	S 206 x FESR – 8	8	9	-	-	-
22.	AK 159	JL 24 x CGC 4018	8	9	-	-	-

Table 4: Difference in melting temperature of HRM amplicons in parental genotypes and their constituent heterozygote.

Allele category	Melting temperature of PCR amplicon		
	HRM 4	HRM 1	HRM 6
Homozygous resistant	77.83 (±0.10)	75.23 (±0.27)	78.26 (±0.02)
Heterozygous (constituted)	78.0 (±0.12)	75.38 (±0.24)	78.24 (±0.02)
Homozygous susceptible	78.18 (±0.04)	74.83 (±0.06)	78.56 (0.08)

Table 5: Position and nature of non-synonymous mutations in different conserved domain of the isolated R-protein

SNP marker	Position in genes	Nature of SNP	Changes in amino acids	Domain of R protein	Position of mutations in respect to conserved motif
FR4 (polymorphic)	Exon 3	A to T C to T G to T T to C	Lys to Asn Ser to Leu Val to Ser	TIR domain	Upstream of hhGRExE motif
FR5 (non-polymorphic)	Exon 3	G to C	Glu to Gln	NBS domain	Within hhGRExE motif
FR1 (polymorphic)	Exon 3	T to A C to G	Ile to Tyr Gln to Glu	NBS domain	Between RNBS-A and Walker B motif
FR2 (non-polymorphic)	Exon 3	G to A G to T	Glu to Asn	NBS domain	Between RNBS-A and Walker B motif
FR3 (non-polymorphic)	Exon 3	A to G T to G C to A G to T	Tyr to Trp Arg to Ile	NBS domain	Immediate vicinity of RNBS-D motif
FR6 (polymorphic)	Exon 7	G to T A to G	Asp to Cys	LRR domain	Within LRR domain

Table 6: Sequence analysis (Ka/Ks) of different R-gene homologue in cultivated groundnut genome

Query sequence	Target sequence	Ka	Ks	Ka/Ks	P value (Fisher)
VG9514-Rgene	*Arahy_GFGJ54*	0.058850	0.047245	1.24563	0.3151
	Arahy_R8KUIR	0.084868	0.082300	1.03119	0.8667
	Arahy_T6DCA5	0.070459	0.072121	0.97965	0.8562
	Arahy_ZZ0VZ9	0.067578	0.068653	0.98435	0.8529

Figures

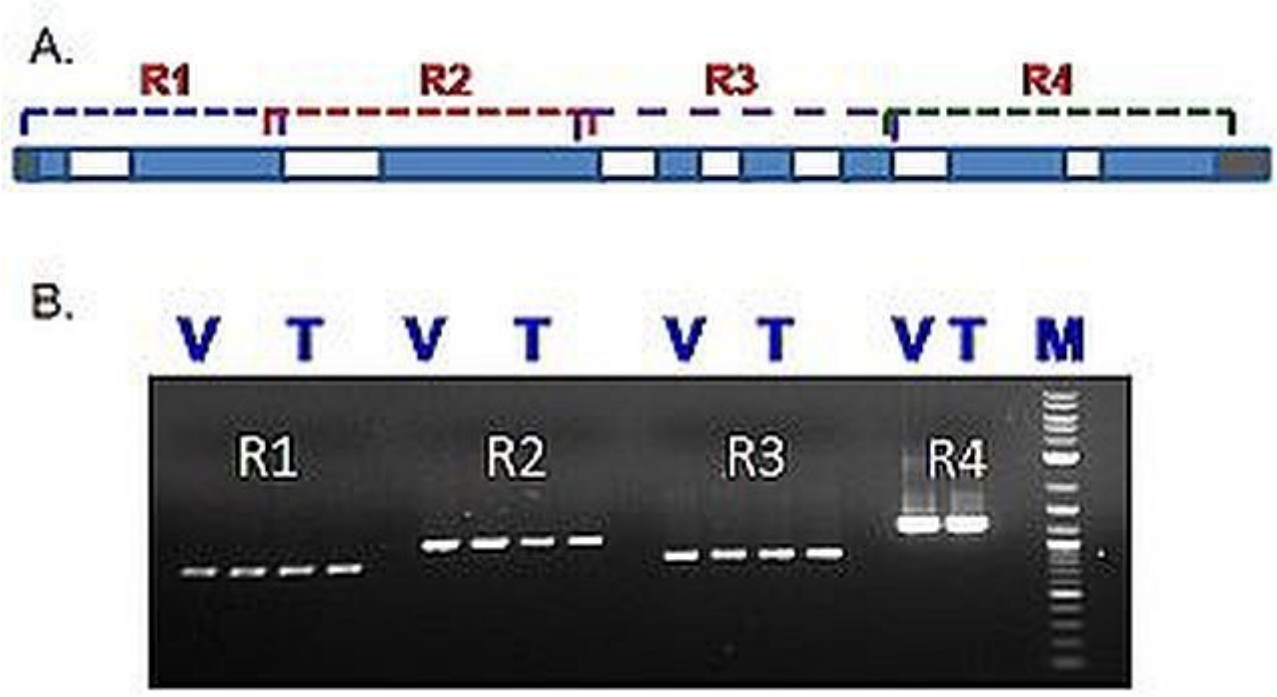

Figure 1

A. Schematic diagram of the isolated R gene and position of four different cloned gDNA fragments; B. Agarose gel electrophoresis of different R-gene fragment PCR amplicons.

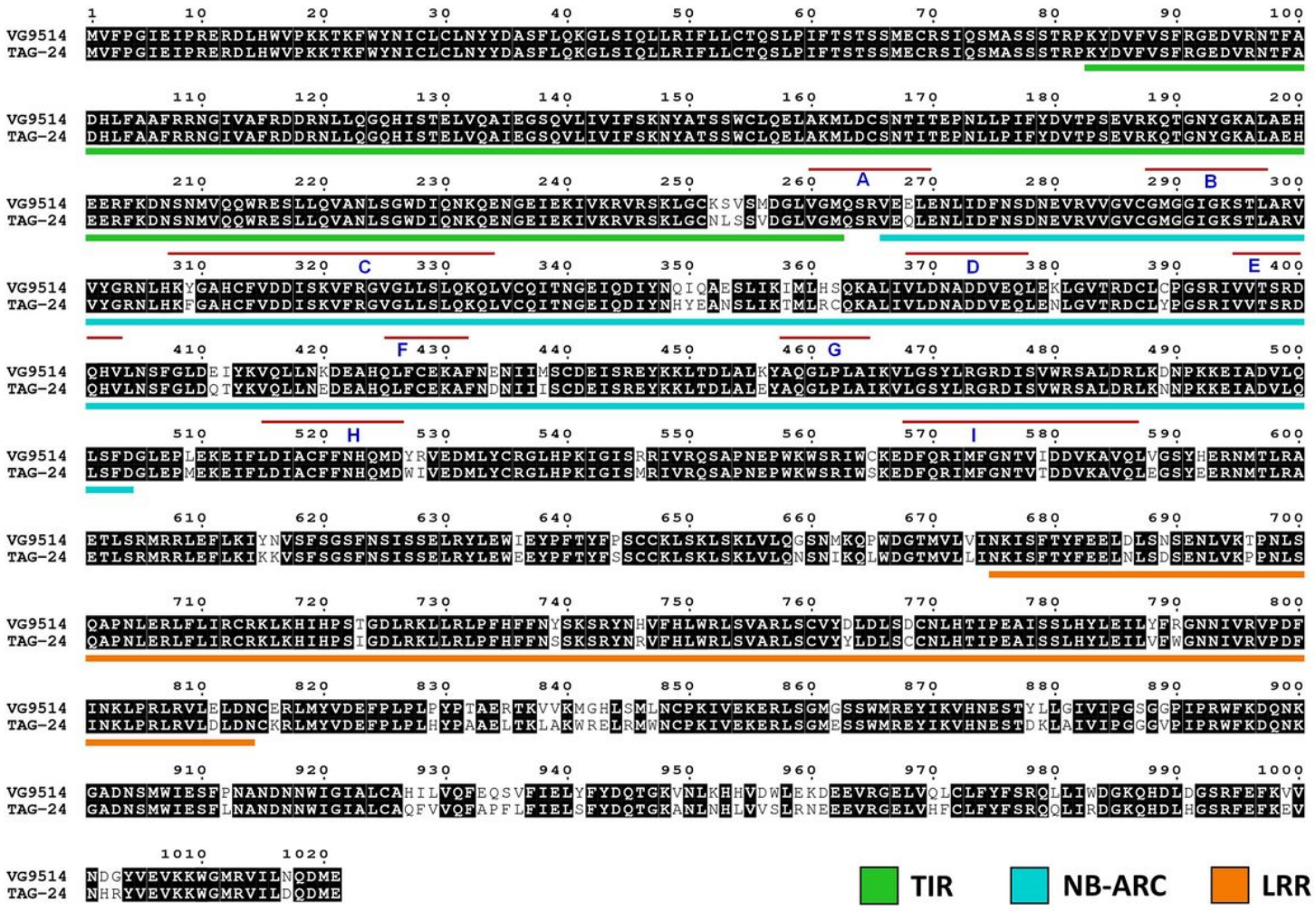

Figure 2

Alignment of R-genes VG9514 and TAG24. Domain boundaries in the alignment have been highlighted. Note: NB: The motif in NBS domain were marked in magenta bar and indicated in alphabet (capital) with blue font color. Motifs were marked above the respective amino acid sequence. A: hhGRExE motif; B: Walker A/P-Loop motif; C: RNBS-A motif; D: Walker B motif; E: RNBS-B motif; F: RNBS-C motif; G: GLPL/GxP motif; H: RNBS-D motif; I: MHD motif

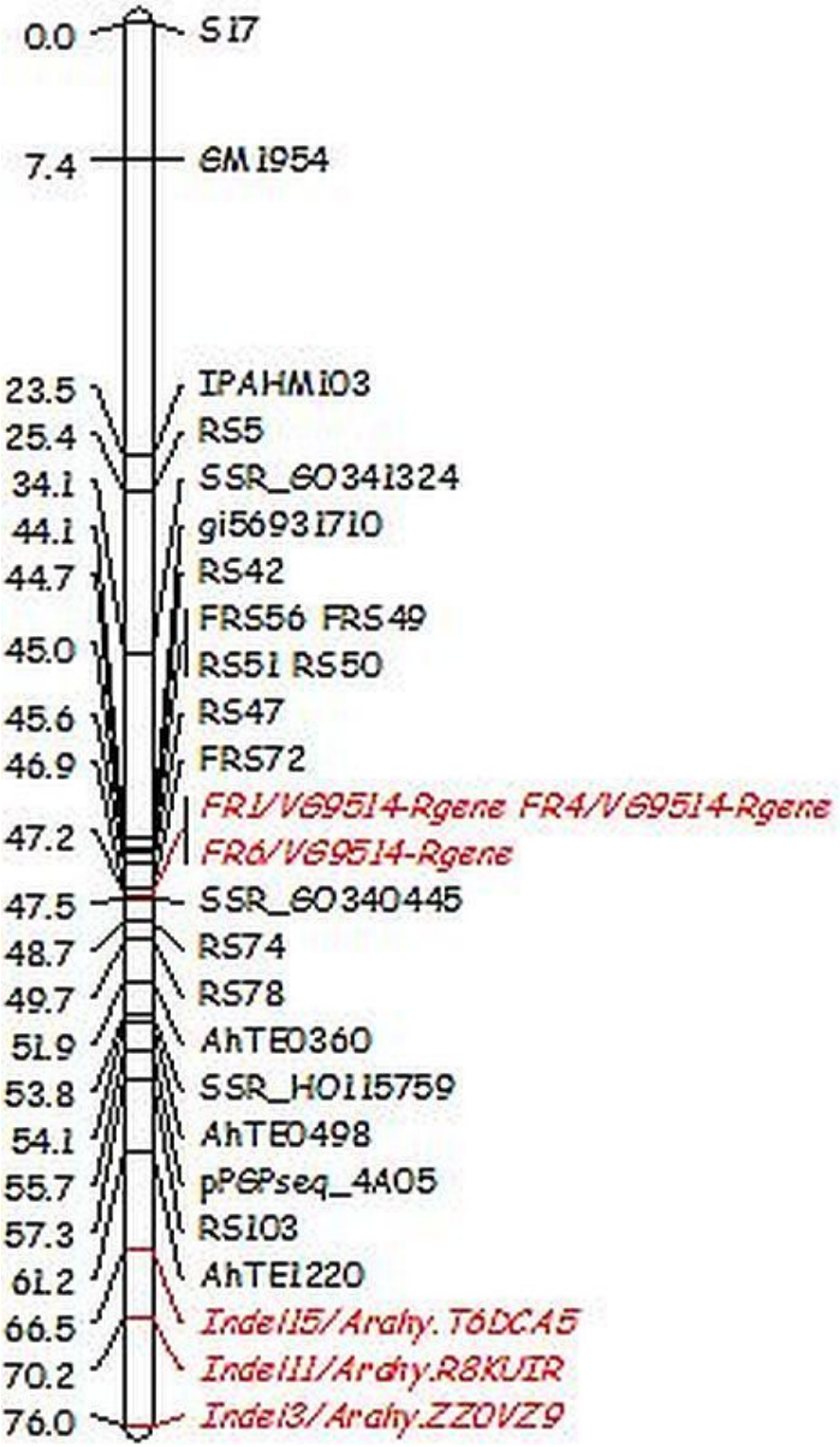

Figure 3

Genetic linkage map of SNP-based PCR markers and InDel markers in A03 chromosome

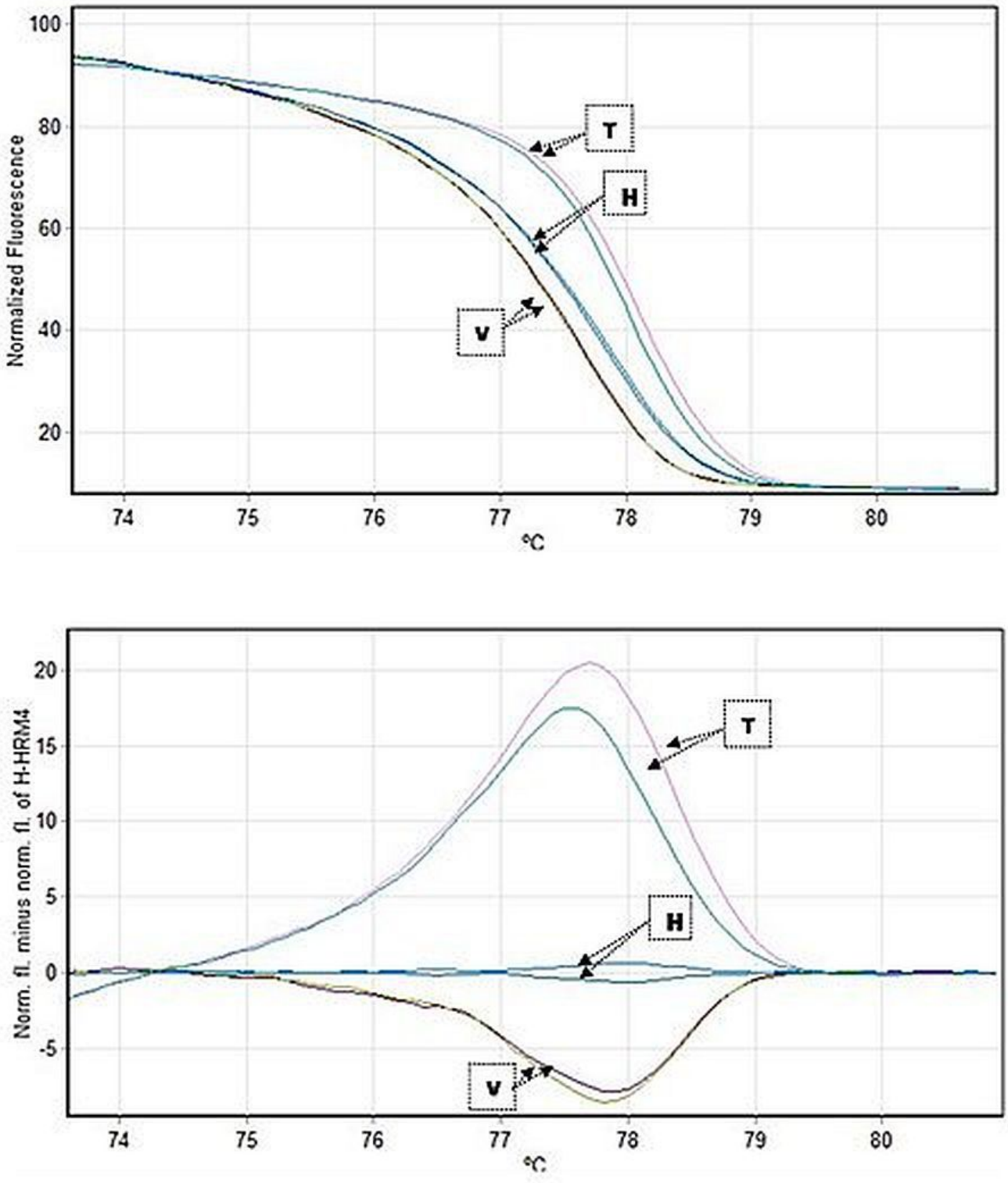

Figure 4

a. Normalized HRM curve of HRM-4 qPCR products of VG 9514, TAG 24 and their constituent heterozygous DNA; b. Difference plots of HRM-4 qPCR products of VG 9514, TAG 24 and their constituent heterozygous DNA. Note: V = VG 9514 (resistant parent), T = TAG 24 (susceptible parent), H = constituent heterozygous DNA.

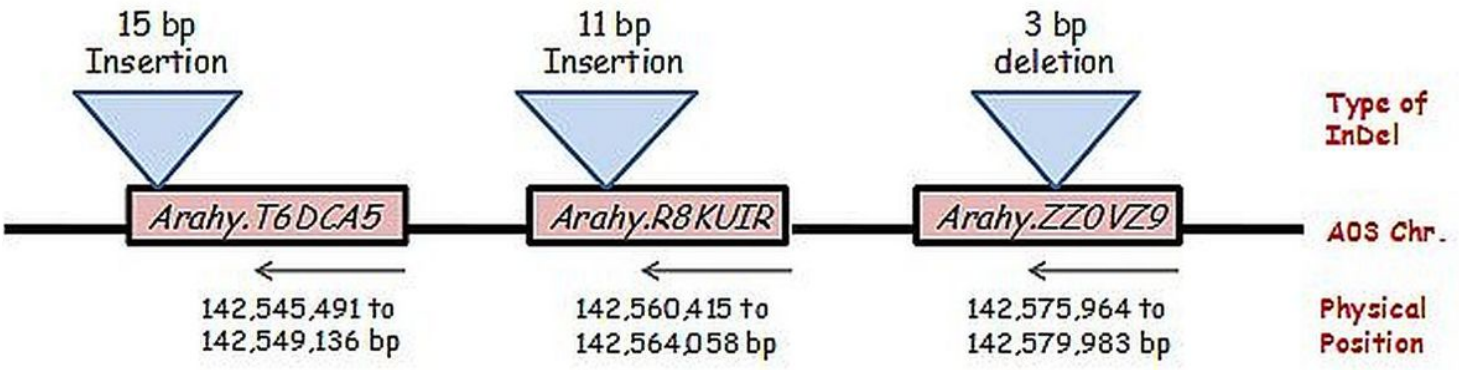

Figure 5

Schematic diagram of R-gene paralogues which were duplicated in tandem on chromosome A03.

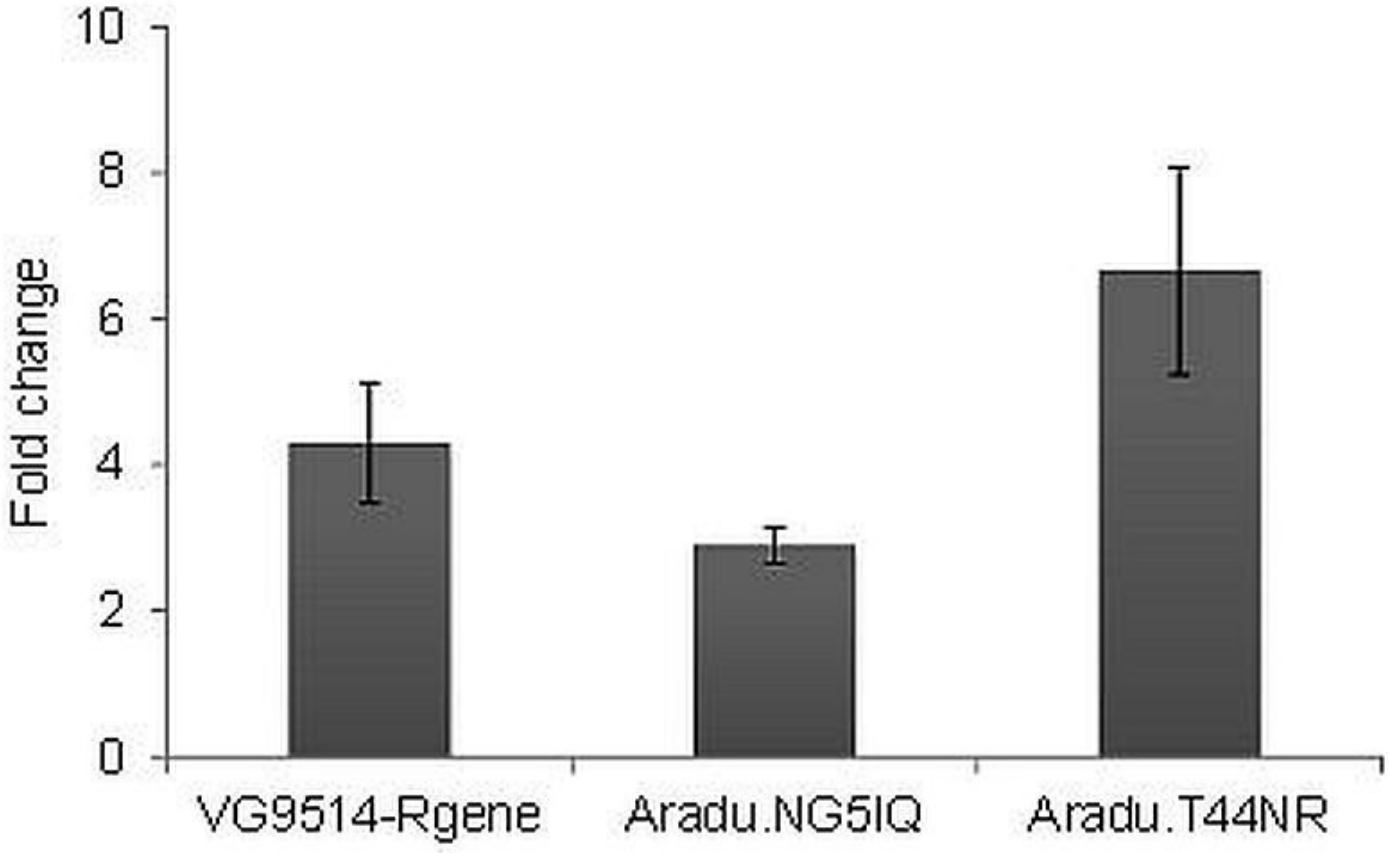

Figure 6

Fold changes in R-gene and two PR genes expression in VG 9514 as compared to TAG 24 in healthy leaves.

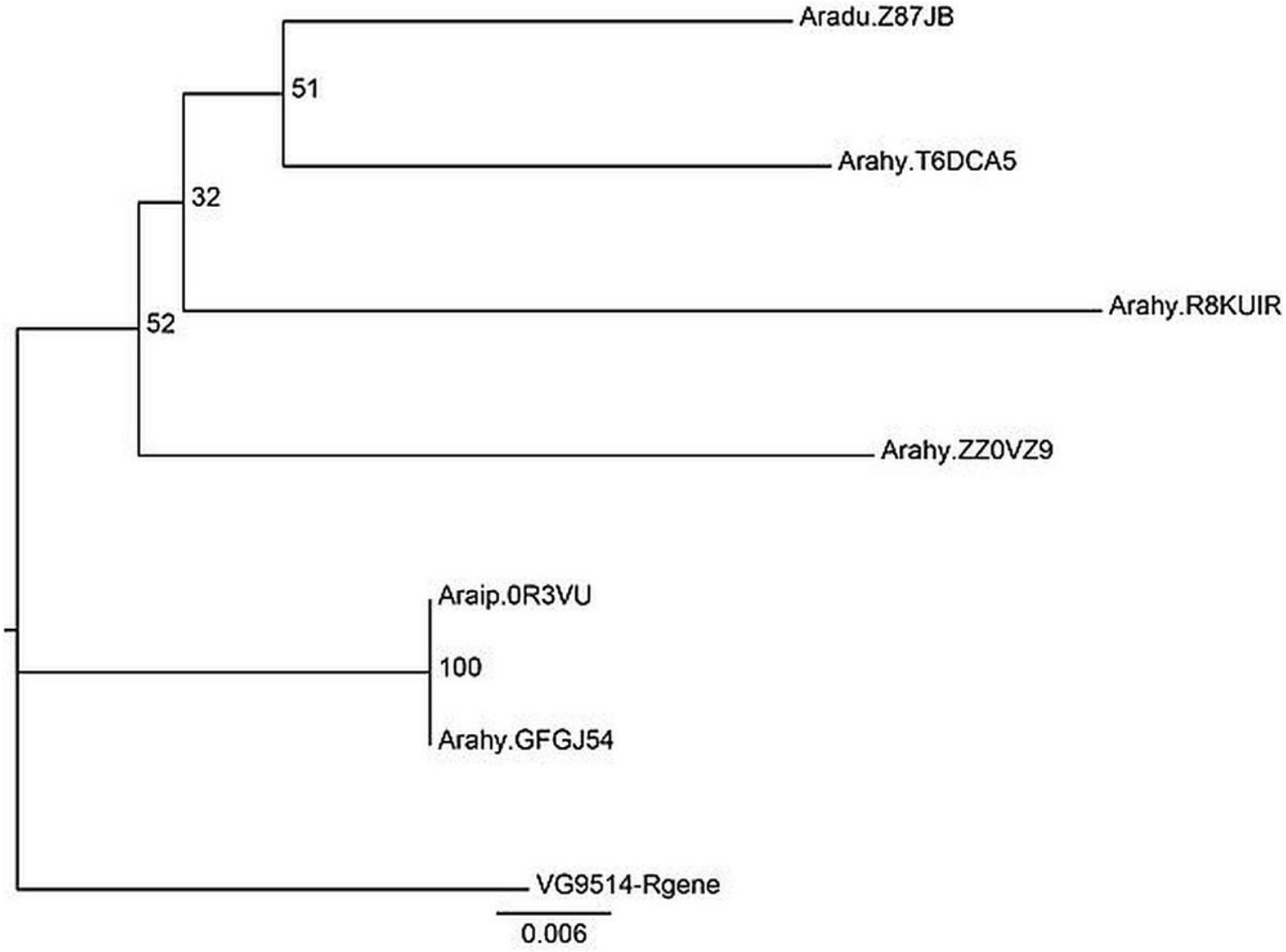

Figure 7

Phylogeny tree of the identified VG9514-Rgene and its homologous genes in groundnut genome

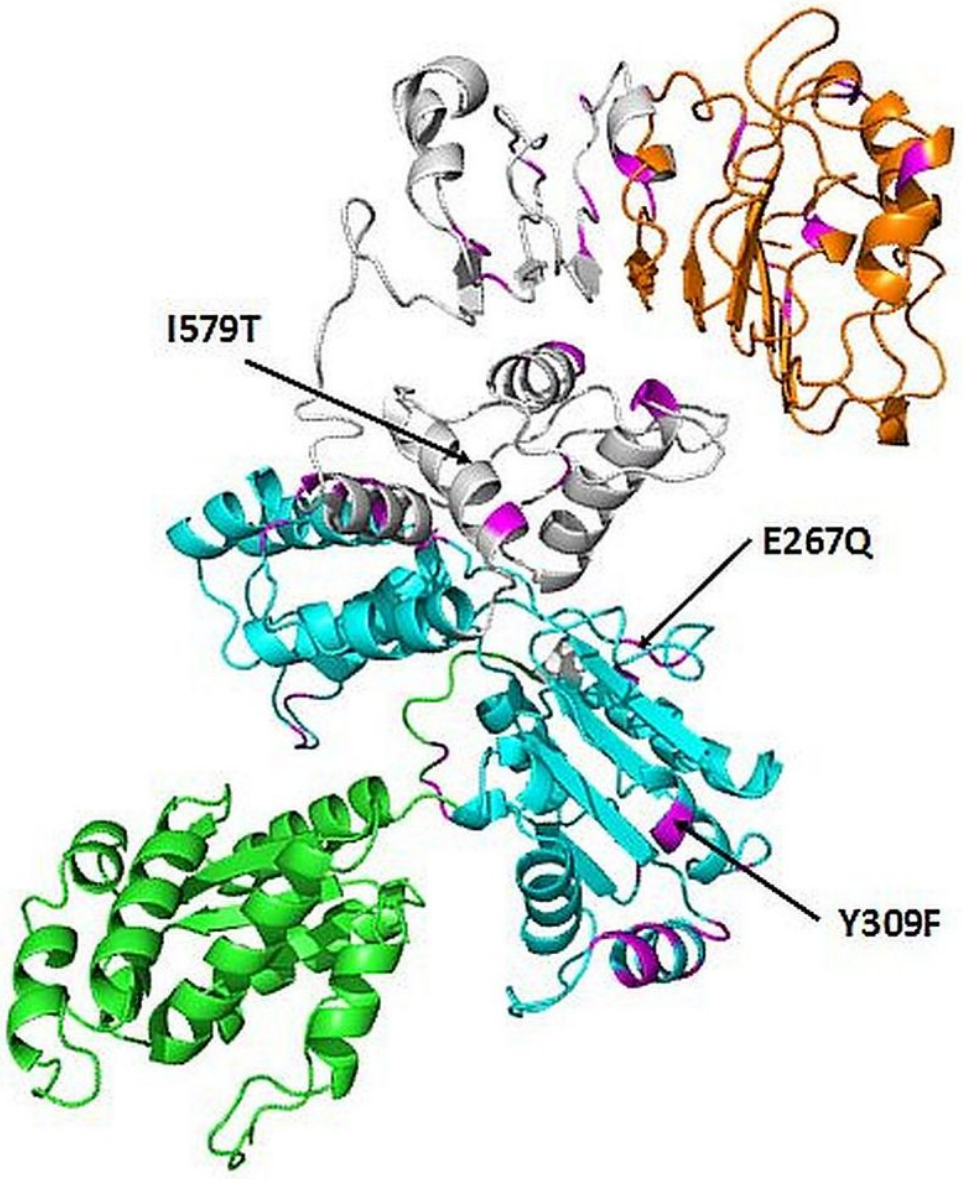

Figure 8

A homology model generated for VG9514 R-protein. Domains have been highlighted in green (TIR), cyan (NB-ARC) and orange (LRR) colors. Residue differences from the TAG 24 gene has been highlighted in magenta color on the model. Three specific mutations that appeared within the conserved motif of NB-ARC domain are highlighted in arrow.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx
- s1.jpg
- s2.jpg
- s3.jpg
- s4.jpg
- s5.jpg
- s6.jpg