DE GRUYTER OPEN

Acta Bot. Croat. 75 (1), 1–10, 2016
DOI: 10.1515/botcro-2016-0003

CODEN: ABCRA 25
ISSN 0365-0588
eISSN 1847-8476

Cloning, characterization and expression analysis of NBS-LRR-type resistance gene analogues (RGAs) in coconut

Kaitheri Edathil Rachana[1], Sudalaimuthu Asari Naganeeswaran[1], Thayale Purayil Fayas[1], Regi Jacob Thomas[2], Muliyar Krishna Rajesh[1]*

[1] Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod 671124, Kerala, India
[2] ICAR-Central Plantation Crops Research Institute (RS), Kayamkulam, Kerala, India

Abstract – Coconut palms are highly susceptible to diseases caused by different pathogens, and replanting with resistant varieties is the best way to manage them. Obtaining a collection of resistance gene analogues (RGAs) is an effective strategy to identify genomic regions linked to disease resistance. We have successfully used a comparative genomics approach to amplify putative RGAs from the coconut root (wilt) disease resistant cultivar Chowghat Green Dwarf (CGD) by using primers designed based on conserved motifs of the NBS-LRR domain of the date palm. The amplified sequences were cloned, sequenced and characterized. The coconut RGAs had high identity to monocot NBS-LRRs. A complete structural analysis and 3-D modeling of the NBS domain of coconut RGA was also undertaken. Real-time quantitative polymerase chain reaction analysis indicated that the isolated coconut NBS-LRR class RGAs was expressed more in root (wilt) disease resistant genotypes than in susceptible ones. This study would provide a base for future efforts to map disease resistant traits in coconut.

Key words: coconut, homology modeling, nucleotide binding site-leucine rich repeat type, real time PCR, resistant gene analogue

Introduction

The coconut palm (*Cocos nucifera* L.) is one of the most important and useful palms in the world and has been a part of human history for over 4000 years. It is an important crop in the agrarian economy of many tropical countries of the world, providing food, drink, shelter and raw materials for industry. The coconut palm is a member of the palm family Arecaceae and is the only accepted species in the genus *Cocos*. Coconut is confronted with an array of diverse pathogens (bacterial, fungal, and phytoplasma), which contributes substantially to yield loss. Thus, breeding in coconut has to aim at the development of high-yielding hybrids with resistance to pathogens.

During the course of evolution, plants have developed complicated defense mechanisms to resist phytopathogens (Staskawicz et al. 1995). The identification of many plant resistance (*R*) genes has helped to understand how plants prevent or slow down infection and disease progression (Hammond-Kosack and Jones 1997). The plant *R* genes encode proteins that can recognize and bind to products encoded by the avirulence genes of the invading pathogens (Scofield et al. 1996). As a result of such a molecular recognition, a signal transduction cascade is initiated, and plant defenses are expressed. These defense responses include hypersensitive response that results in localized cell death, and other general responses such as strengthening of the cell wall, formation of phytoalexins etc. (Dangl et al. 1996).

Plant *R* genes have been cloned and characterized from a number of species, including both mono- and dicotyledonous plants. These *R* genes can be broadly divided into four structurally distinct classes (Ellis et al. 2000). The first class of resistance genes belongs to the serine-threonine kinases (Ritter and Dangl 1996). These protein kinases phosphorylate serine/threonine residues and thus control certain signaling networks during the resistance response. The second class of resistance genes encodes putative trans-membrane receptors with extra-cellular leucine rich repeat (LRR) domains (Jones et al. 1994). The third class encodes for a receptor-like kinase and combines qualities of both the previous classes. Both the LRR domain and the protein kinase regions are encoded in the same protein. The fourth

* Corresponding author, e-mail: mkraju.cpcri@gmail.com

class, which represents the majority of plant disease resistance genes cloned so far, is the nucleotide binding site-leucine rich repeat (NBS-LRR) resistance genes. The leucine-rich repeat (LRR) domains provide pathogen recognition specificity by participating in protein-protein interactions and ligand binding (Ellis et al. 2000).

NBS-LRR proteins are composed of three domains: a variable N-terminal domain (~200 amino acids), a nucleotide binding site domain (~300 amino acids) and a variable LRR domain (~10–40 short Leucine-Rich-Repeat motif) (Cannon et al. 2002). The NBS domain consists of several conserved motifs including the P-loop and the kinase-2 domains, which are ATP- and GTP-binding sites (Meyers et al. 1999); the kinase-3a domain and the putative membrane spanning hydrophobic GLPL domain (Baldi et al. 2004). The conserved motifs offer opportunities for designing degenerate primers for polymerase chain reaction (PCR)-based strategies to recover similar sequences called resistance gene analogues (RGAs) from other plant species (Kanazin et al. 1996).

In the present study, our aim was to clone and characterize RGAs from coconut, whose genome has not been sequenced and for which the genomic resources available in public databases are scarce. We used oligonucleotide primers designed to the conserved motifs of NBS-LRR domain of resistance genes of date palm, the genome of which has been sequenced, to recover RGAs from coconut. The secondary structure of partial R-protein and the 3-D model of NB-ARC domain of isolated RGAs were predicted. Gene expression studies were carried out using qRT-PCR to confirm the up-regulated expression patterns of the coconut RGAs in disease resistant genotypes.

Materials and methods

Retrieving date palm genome sequence and creation of a standalone database

Whole genome sequences of date palm (*Phoenix dactylifera* L. 'Khalas' variety), comprising a total of 1, 42, 304 whole genome shotgun sequence (which includes 28,889 annotated genes), were retrieved (http://qatar-weill.cornell.edu/research/datepalmGenome/index.html) and a standalone date palm sequence database was constructed using formatdb program available in standalone BLAST tool for comparative analysis.

PCR amplification of RGAs and sequencing

Nucleotide sequences of date palm NBS-LRR-type *R* genes were translated using EMBOSS TRANSEQ program and searched against Pfam (http://pfam.sanger.ac.uk) and PRINTS (http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/index.php) databases. Motifs identified with Pfam and PRINTS were confirmed using PROSITE (http:/www.expasy.ch/prosite/).

Based on sequence structural motif prediction by PROSITE/PRINTS database, major disease resistant amino acid motifs characteristic of the NBS-LRR class were identified in NBS-LRR proteins of date palm and these were used to design specific oligonucleotide primers for polymerase chain reaction (PCR) amplification of coconut genomic DNA. Three pairs of forward and reverse primers were designed from different consensus motifs. A fourth primer was designed in the internal region of the first primer. The oligo parameters were optimized by PCR oligonucleotide resources such as OLIGOANALYZER (http://www.idtdna.com/analyzer/applications/oligoanalyzer/) and designed primers for efficient amplification of coconut RGA by FASTPCR primer design software (http://primerdigital.com/fastpcr.html).

DNA was extracted from the spindle leaf samples of the root (wilt) disease resistant cultivar Chowghat Green Dwarf (CGD) palm using a modified SDS-procedure (Rajesh et al. 2013).

PCR was carried out in a 20 μL reaction mixture containing 25 ng template DNA, 200 nM each of forward and reverse primer, 10 mM of each dATP, dCTP, dGTP and dTTP, 10× PCR buffer, 15 mM $MgCl_2$ and 1 U Taq DNA polymerase (M/s Bangalore Genie, India). The cycling profile consisted of an initial denaturation at 94 °C for 2 min followed by 34 cycles each consisting of DNA denaturation at 94 °C for 30 s, primer annealing at the standardized annealing temperature for 1 min and primer extension at 72 °C for 1 min, and a final extension for 10 min in a thermal cycler (BIORAD). PCR products were further analyzed by electrophoresis on 1.2% agarose gel (1× TBE containing 0.1 μg mL^{-1} ethidium bromide), under 80 V for 1 h. After electrophoresis, the gel images were taken using a gel documentation system (BIORAD) for later analysis.

The PCR-amplified products were purified with the use of a PCR purification kit (QIAGEN). Longer amplicons were sequenced by primer walking strategy using the designed internal primers. All amplicons were cloned and sequenced. Sequencing was done using a BigDye® Terminator v3.1 sequencing kit and analyzed on ABI PRISM® 377 Genetic Analyzer.

Sequence analysis

Raw sequences were trimmed to remove the vector contamination using VECSCREEN. The trimmed sequences were analyzed to identify potential open reading frames (ORFs). Amino acid identities of coconut RGAs to well characterized gene sequences were determined by performing translated BLAST (BLASTX) with non-redundant amino acid sequences (http://blast.ncbi.nlm.nih.gov/blast.cgi/). Multiple sequence alignment was carried out using the software CLUSTALX. This alignment was used in the MEGA software to build neighbor- joining tree under the Jones-Thornton-Taylor (JTT) model (Tamura et al. 2011) with 1000 bootstraps. The amino acid sequences of cloned known disease resistance genes *L6* (LUU27081) from flax (*Linum usitatissimum*), *RGC4A* (AF017754) from lettuce (*Lactuca sativa*) and *RPP1* (AF098962), *RPS2* (ATU14158) and *RPM1* (NM111584) from *Arabidopsis* and *Rx* (AJ 011801) from potato (*Solanum tuberosum*) were retrieved from Genbank. RGAs from date palm (gi_672130146) and oil palm (gi_6690745) were also used as references. The NBS-LRR domains were extracted and aligned with coco-

nut RGAs. The comparison of coconut sequence with date palm sequence was also done with the use of mVISTA (MLAGAN algorithm) server (Frazer et al. 2004).

The nucleotide sequences obtained were translated *in silico* into protein sequences using EMBOSS TRANSEQ program and motifs present in the sequences were confirmed using Pfam, PROSITE, PRINTS and SMART database search. The conserved motifs within the NBS domain were identified using MEME server (multiple expectation maximization for motif elicitation) (Bailey et al. 2006).

Additionally, protein sequences obtained by *in silico* translation of isolated coconut RGAs were subjected to BLASTP analysis. Highly significant hits obtained by BLASTP analysis were subjected to the SMART tool (http://smart.embl-heidelberg.de) for predicting the NB-ARC domain. The selected sequences were aligned using MAFFT with jalview (http://mafft.cbrc.jp/alignment/software/), for the multiple sequence alignment and phylogenetic tree with bootstrap method was constructed.

Homology modeling

For further characterization of the translated proteins, we submitted the amino acid sequences to a secondary structure prediction server, PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/). The two feed forward neural networks of PSIPRED perform an analysis on output obtained from PSI-BLAST (position specific iterated BLAST) homology search algorithm (Altschul et al. 1997, Jones 1999). The preliminary 3-D framework of our amino acid sequences based on homology molecular modeling was initiated by the search for a suitable template for 3-D structure prediction of the NBS and LRR region of coconut RGA in Protein data bank (PDB). Templates with significant scores were selected. In fold and function assignment (FFSA03) (Jaroszewski et al. 2005), an alignment profile was created with template and the target protein sequence was then deposited as an input for ProtMod (http://ffas.burnham.org/protmod-cgi/protModHome.pl) structure modelling server for 3-D structure, with SCWRL as default modelling method, and MODELLER (https://salilab.org/modeller/). Desired PDB file of predicted model was generated by ProtMod, subsequently viewed by Pymol software version 1.0. The accuracy of predicted 3-D-protein model was further verified by SAVES (structural analysis and verification server, http://nihserver.mbi.ucla.edu/SAVES), where the PDB file of the protein model was subjected separately for the analysis. Further, the protein model validation was also done based on $\phi - \psi$ ratio of Ramachandran plot obtained from PROCHECK analysis (Wiederstein and Sippl 2007), which provides detailed view on the stereo-chemistry. Finally, PROSA-Web (https://prosa.services.came.sbg.ac.at/prosa.php) search was applied to assess the energy criteria of the constructed protein structure.

qRT-PCR analysis

Total RNA from root (wilt) disease resistant and susceptible CGD genotypes was isolated using plant RNA kit (M/s QIAGEN) and integrity, concentration, quality checks and DNase 1 treatment were performed. Reverse transcription was performed with High Capacity RNA-to-cDNA kit (M/s Life Technologies) using aliquots of total RNA extracted. The cDNA samples were diluted to 20 ng μL^{-1}. CN-NBS-LRR gene specific primers were designed by using PRIMER 3.0 (http://primer3.ut.ee/) software and selected for PCR amplification. Amplification was carried out with a reaction mixture containing 2 µL of 10× reaction buffer, 2 µL of 25 mM $MgCl_2$, 0.2 µL of 10 mM dNTP mix, 1 µM of each forward and reverse primer, 1 µL of synthesized cDNA and 1 U of Taq polymerase. RT-PCR reactions were carried out using BIORAD thermal cycler. The cycling program was as follows: 1 min at 94 °C, 30 cycles of 30 s at 94 °C, 1 min at 55 °C, 1 min at 72 °C and a 10 min extension at 72 °C, with a negative control.

All real-time PCR reactions were performed using the StepOne™ real time PCR system (Applied Biosystems) and the amplifications were done using the SYBR green PCR master mix (Applied Biosystems). The thermal conditions were as follows: initial holding stage 52 °C for 2 min, 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and a final step at 60 °C for 1 min. The experiments were carried out in triplicate for each data point. The relative quantification in gene expression was determined using the 2-ΔΔCt method (Livak and Schmittgen 2001). Using this method, we obtained fold change gene expression in resistant and susceptible genotypes, normalized to an internal control gene, α-tubulin (α-TUB).

Results

Identification of motifs of *R* gene in date palm

Many *R* genes have been cloned from different plant species and these genes have been found to share some structural domains, which include P-loop (kinase-1a), the kinase-2, the GLPLAL and the MHD. Characteristic disease resistant motifs of date palm NBS-LRR RGA sequences (PDK_30s665281g007, PDK_30s742701g004 and PDK_30s816511g003) were identified from the motif prediction tools such as PROSITE/ PRINTS/ Pfam. Based on sequence structural motif prediction by the PROSITE/PRINTS database, four major disease resistant amino acid motifs characteristic of the NBS-LRR class were identified in the three NBS-LRR proteins of date palm and these were used to design specific oligonucleotide primers for polymerase chain reaction (PCR) amplification of coconut genomic DNA (Fig. 1a).

PCR amplification of targeted RGA fragments in coconut

Primers were designed in such a way that around 2200, 1200, 900 and 600 bp PCR products were obtained upon amplification respectively using the designed primers (Fig. 1a, Tab. 1). All four primer pairs, designed according to the consensus motifs of *R* genes in date palm, were used for PCR amplification of template DNA of coconut root (wilt) disease resistant CGD genotype. These primer pairs produced expected size bands according to the date palm species (source) from which the primers were designed (Fig. 1b). Primer CN1F1/R1 amplified the largest band of ap-

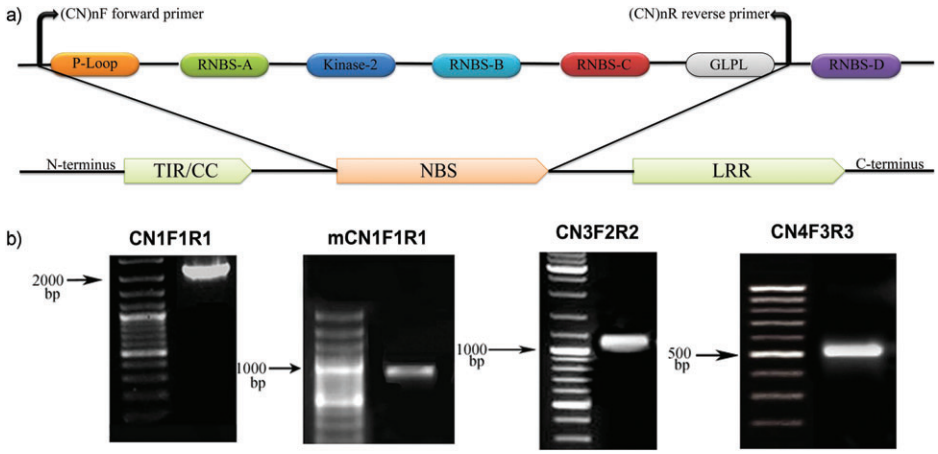


Fig. 1. a) Diagram showing the NBS domain motifs used in primer design. (CN)nF and (CN)nR are general NBS primer sets (CN1F1/R1, CN3F2/R2, CN4F3/R3; see Tab. 1) designed from P-loop and GLPL motifs of three different date palm resistance gene analogue sequences; b) Gel profile of PCR products amplified from four *R* gene of coconut by specific motif-based primers: CN1F1/R1 (around 2200 bp), mCN1F1/R1 (around 900 bp), CN3F2/R2 (around 1200 bp), and CN4F3/R3 (around 600 bp). First line on each gel presents 1 Kb DNA marker.

Tab. 1. Specific primers of disease resistant motifs used for screening for NBS-LRR resistance gene analogues in coconut. S. No. – sample number.

S. No.	Primer name	Forward (5'–3')	Tm (°C)	Conserved motif and R–Gene	Amplified product size (bp)	Annealing temperature (°C)
1	CN1F1	CTCTGGACAATCAGTGATG	56	ALLIVGMAGLGKTTLA	2200	58
	CN1R1	GGGCTTGGGAAAACAACT	58	HNLRSLEKLTITDCPEL		
2	CN3F2	GAGTCCTAGATACCGAAGC	53.1	GLGKTTL	1200	58
	CN3R2	GTTCTCCCAATAGTTGGC	50.1	GNLIQLRYLGLSDAKTK		
3	CN4F3	GATGGGTGGGGTTGGTAAG	55.4	GIGIVGMGGVGKTLLA	600	58
	CN4R3	CTACAGTCTTCGCAGCTAATG	53.2	CGGLPLAAKTVGEIL		
4	mCN1F1	GATGGGTGGGGTTGGTAAG	60	–	900	58
	mCN1R1	CTACAGTCTTCGCAGCTAATG	62	–		

proximately 2200 bp size and CN3 F1/R1 amplified a target of around 1200 bp. In addition, a 600 bp sized fragment was amplified by CN4 F1/R1. Primer mCN1F1/R1, designed from the middle portion of the date palm RGA sequence, gave an additional amplified product of around 900 bp in coconut, which is in between the product amplified by CN1 F1/R1. Target specific large amplicon (2200 bp) were purified, cloned and sequenced by primer walking. These sequences were deposited in the GenBank database and their accession numbers are KC465244 (2211 bp; designated as CN1- NBS-LRR), KF002584 (1165 bp; designated as CN3- NBS-LRR) and KM983337 (616 bp; designated as CN4- NBS-LRR).

Analysis of sequence homology and characterization

Like other RGA sequences, coconut RGA sequences with high identity to each other at the amino acid level may belong to the same gene families or subfamilies. The three putative RGAs from coconut, which showed ≤ 90% identity to each other, were selected for further characterization (On-line Suppl. Tab. 1). Coconut RGAs were highly similar to RGAs cloned from different genera *viz.*, *Oryza, Triticum, Aegilopus, Musa* and *Brachypodium* (On-line Suppl. Tab. 1). The maximum amino acid identity between coconut RGAs and other plants' RGAs ranged from 45–100% (On-line Suppl. Tab. 1).The multiple sequence alignment analysis of coconut RGAs with known resistant genes of Toll and Interleukin receptor-NBS-LRR (TNL) and coiled-coiled-NBS-LRR (CNL) classes showed that coconut RGA sequences were highly clustered with CNL- NBS sequences (Fig. 2a) and contain typical motifs (P-Loop, RNBS-A-non TIR, Kinase- 2, RNBS-B, RNBS-C and GLPL) of the CNL class R proteins within the NBS-domain (Fig. 2b) including RPM1 (gi_440572080) and RPS2 from *Arabidopsis* (gi_549979) and *Rx* from potato (gi_5524754), NBS-LRR sequences from date palm (gi_672130146, gi_672112993) and oil palm (gi_6690745). Coconut RGAs were found to cluster with date palm and oil palm RGAs. TNL sequences formed a separate group which included L6 from flax (gi_862905), RPP1 from *Arabidopsis* (gi_332644385), RGC4A from lettuce (gi_2852690) and contained typical motifs within the NBS domain (P-Loop, RNBS-A-TIR, Kinase-2, RNBS-B, RNBS-C and GLPL). The amino acid identity among the three coconut RGAs was found to be 15–26%.

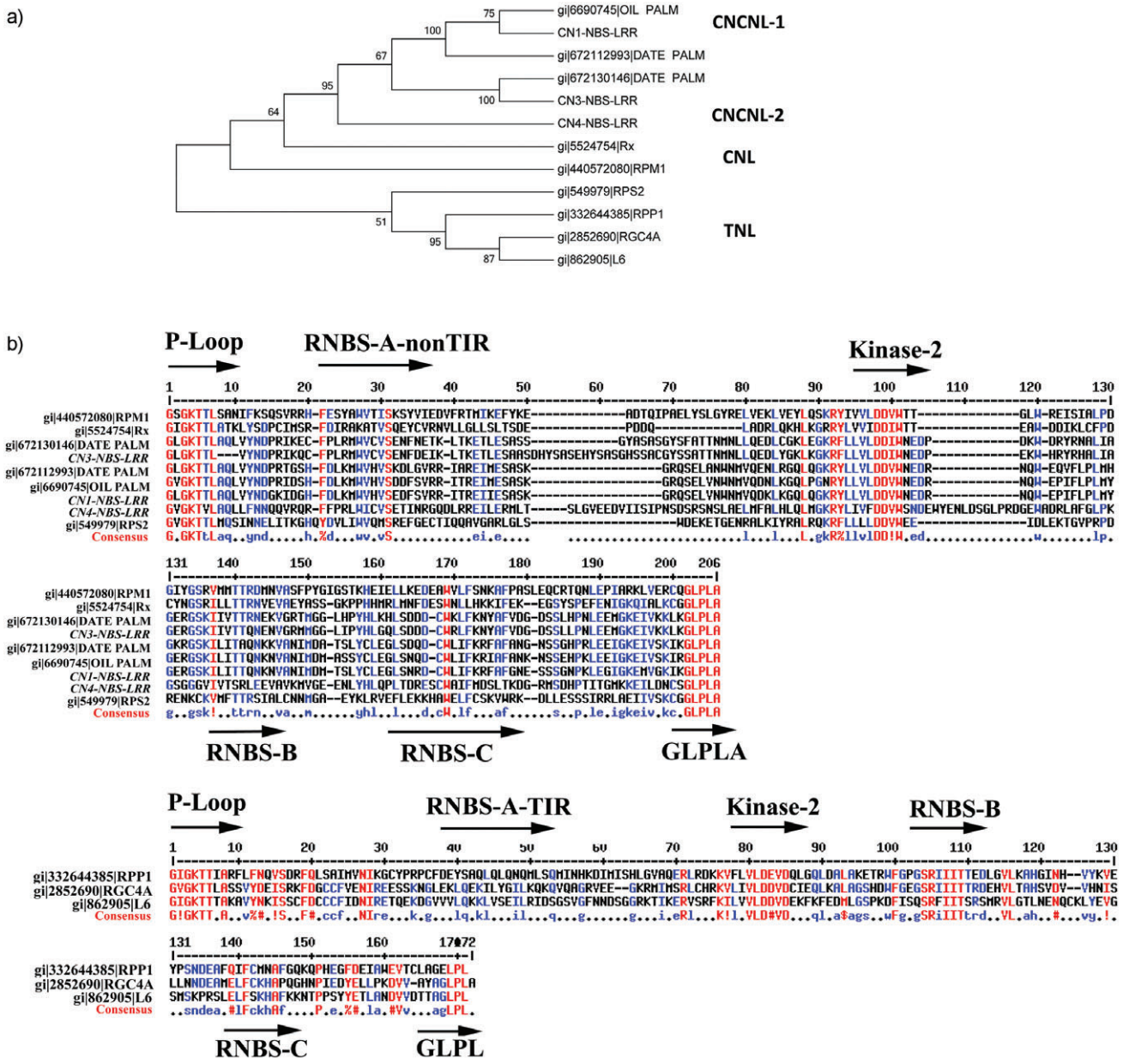


Fig. 2. Relationship among three putative coconut resistance gene analogues (RGAs) and nine known plant NBS-LRR class *R*-genes based on amino acid sequences: a) Phylogenetic tree of three putative coconut resistance gene candidates and RGAs from date palm, oil palm, RPM1, RPP1, RPS2 from *Arabidopsis*, R_x of potato and L6 of flax. Numbers by the branch joints are bootstrap values (out of 1000) that support the clustering. The scale bar represents substitutions per site. b) Amino acid sequence alignment of three coconut RGAs and nine known NBS-LRR class *R*-genes from date palm, RPS2, RPM1 from *Arabidopsis*, Rx of potato of CC-NBS-LRR (CNL) class. Specific motifs of P-Loop, RNBS-A-non TIR, Kinase-2, RNBS-B, RNBS-C and GLPL hydrophobic domain within the NBS-domain are shown. L6 of flax, RPP1 of *Arabidopsis* and RGC4A of lettuce were clustered together and the P-loop, RNBS-A-TIR, Kinase-2, RNBS-B, RNBS-C and GLPL motifs of TIR-NB-LRR class genes (TNL) within the NBS-domain are shown. Only amino acids in the above mentioned motifs are shown in the figure while amino acids between the motifs were replaced with '----' to reduce the space needed.

Alignment and the region of high conservation of *R* genes across two palms, coconut (CN1-NBS-LRR, CN3-NBS-LRR and CN4-NBS-LRR) and date palm (PDK_30s665281g007, PDK_30s742701g004, PDK_30s816511g003) could easily be identified and analyzed using mVISTA server. High percentage of conservation of coconut NBS-LRR sequences with date palm NBS-LRR sequences were observed (data not shown). The BLASTP analysis of the extracted NB-ARC domain of amino acid sequences of coconut RGAs was carried out and the hits obtained were aligned using MAFFT multiple alignment program. Phylogenetic tree was constructed using a neighbour-joining algorithm with 1000 bootstrap trials (On-line Suppl. Fig. 1).

Analysis of conserved motifs in coconut RGA

The three isolated coconut NBS-LRR nucleotide sequences were translated into amino acids using EMBOSS TRANSEQ and subjected to PRINTS/ Pfam databases for

the prediction of conserved resistance motifs. NB-ARC and LRR domains were identified from Pfam database (On-line Suppl. Tab. 2a), while seven signature motifs were identified from PRINTS database (On-line Suppl. Tab. 2b). Newly identified coconut NBS-LRR type RGAs showed conservation in the nucleotide-binding domain. MEME analyses revealed several conserved motifs like P-Loop, RNBS-A-nonTIR, Kinase-2, Kinase-3 and RNBS- C motifs within the NBS-region of all the three coconut NBS-LRR type RGAs. The position and logos of the motifs within the NBS domain are shown in Fig. 3. Exact organization in each protein is obtained by the length and position of each motif.

Secondary structure of CNRGAs

Secondary structure of proteins, comprising constantly repeating local structures like α- helix and β- sheet stabilized by hydrogen bonds, were predicted using the PSI-PRED algorithm. CN1-NBS-LRR was found to possess 20 helices, 50 random coils and 18 strands present at various positions in the protein structure (On-line Suppl. Fig. 2a), while CN3-NBS-LRR possessed 14 helices, 8 strands and 26 random coils (On-line Suppl. Fig. 2b) and CN4-NBS-LRR was made up of 10 helices, 4 strands and 16 random coils (On-line Suppl. Fig. 2c).

3-D structure prediction and validation of NBS-domain

PDB search result for suitable template detection to predict the 3-D structure of CN1-NBS-LRR showed maximum identity to chain A of APAF-1 (PDB ID: 1z6t:A) with an E < 0.0 and with significant identity (61%). FFAS03 pair wise alignment showed that NBS region from 1 to 249 of CN1-NBS-LRR had the expected homology with 147 to 398 of template chain A. ProtMod generated the desired PDB file of the predicted 3-D model of NBS region of CN1-NBS-LRR, which was analyzed in PyMol software, version 1.0. Stereochemical quality, the different parameters, static interactions between the non-bonded atoms, stability etc. of the predicted protein model were verified by SAVES. The 3-D structure of NBS region of CN1-NBS-LRR provides the complete details of different spatial arrangements and function of conserved motifs. The modeled cartoon structure of NBS region is shown in Fig. 4a. NBS domain consists of motifs from N-terminus P-Loop (pink), RNBS-A (light blue), Kinase-2 (wheat), RNBS-B (blue), RNBS-C (yellow), RNBS-D (violet) and GLPL (cyan). The nucleotide binding cleft and relative arrangements of all the motifs of NBS are well conserved in the homology model of CN1-NBS-LRR. The catalytic cleft at the central region was found to be surrounded by other conserved motifs: P-loop, GLPL etc. The P-loop was made up of β-sheets flanked by α-helix, kinase-2 of β-sheets and other motifs of α-helices.

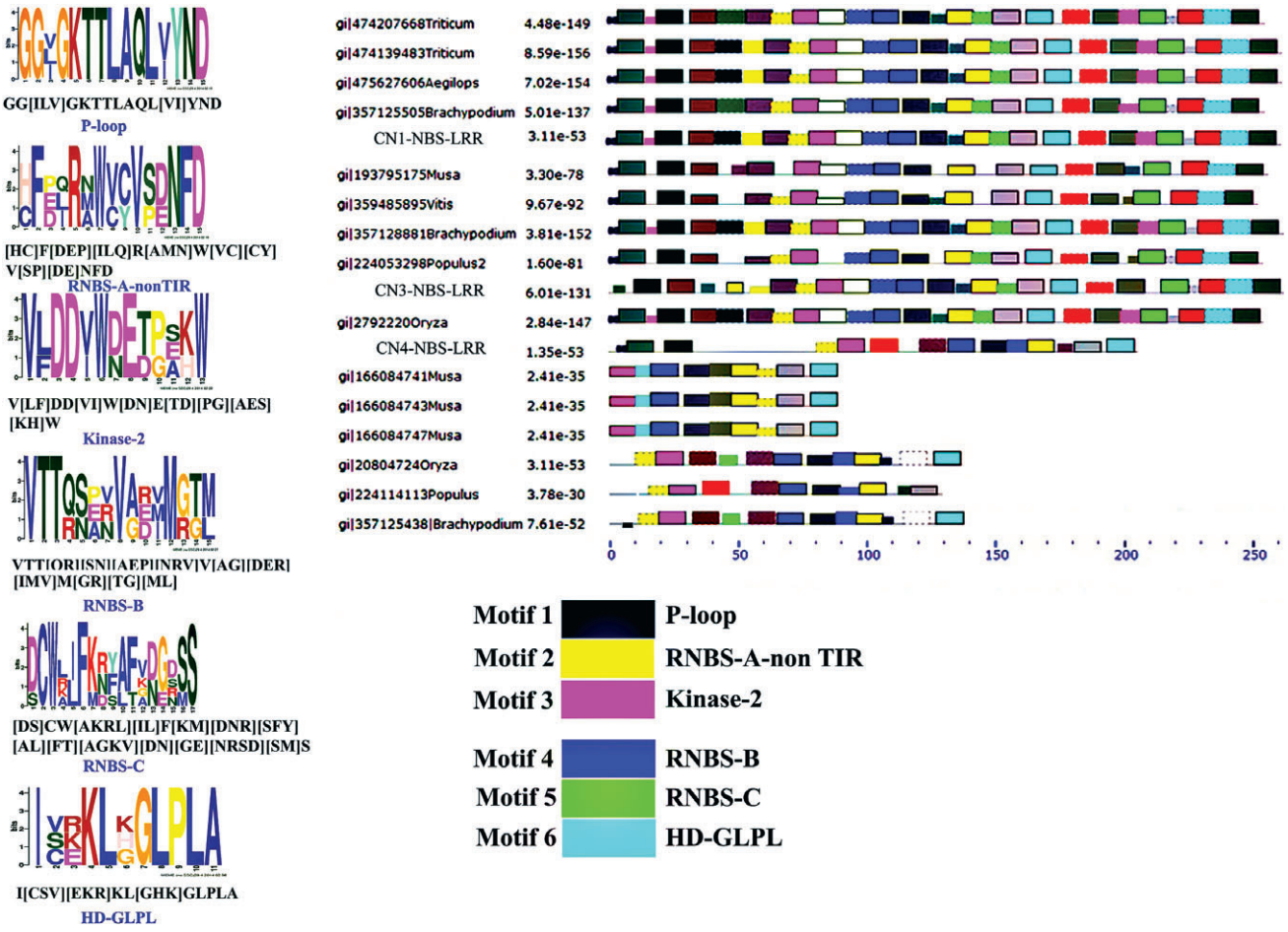


Fig. 3. Different conserved motifs span within nucleotide binding domain identified using MEME. The name of each sequence and combined P-value are shown here. Different motifs are shown by different coloured boxes.

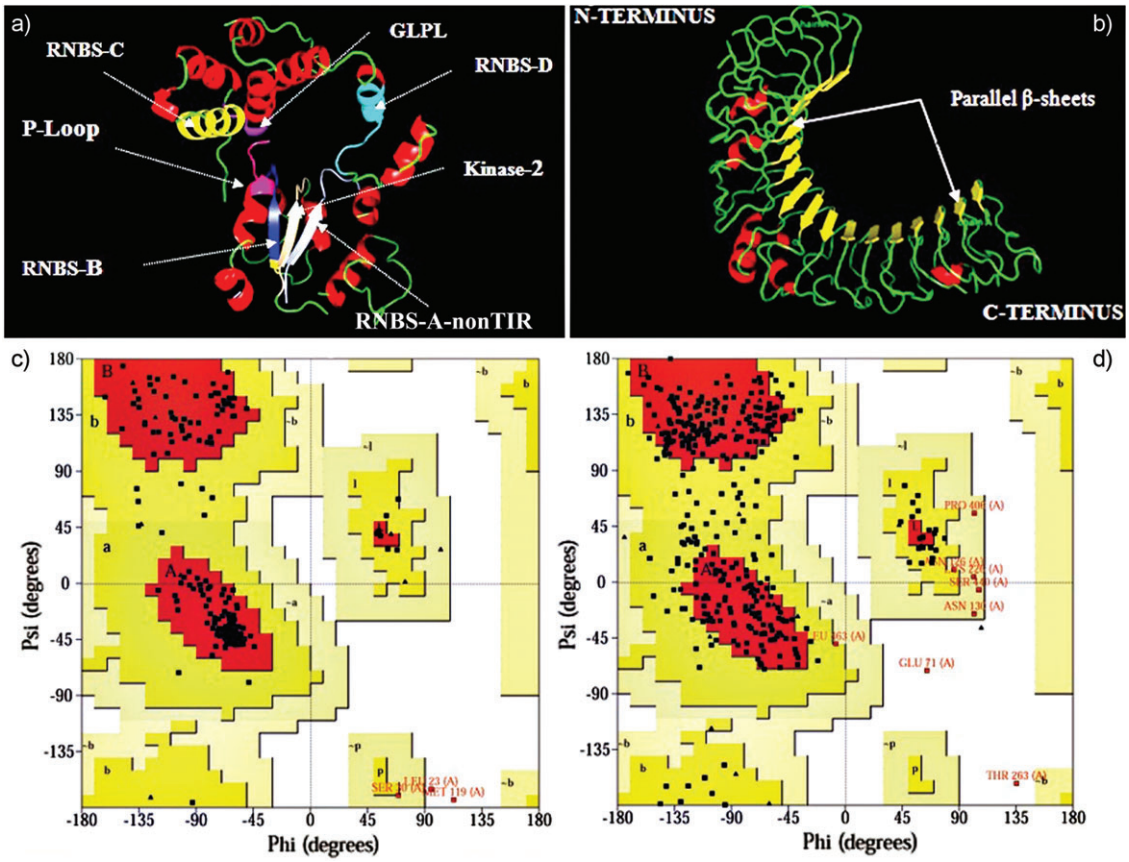


Fig. 4. 3D models of CN1-NBS-LRR domains: (a) NBS domain (PDB ID- 1z6t:A); 3-D cartoon view revealed the relative position of different conserved motifs from N-terminus: P-Loop (pink), RNBS-A (light blue), Kinase-2 (wheat), RNBS-B (blue), RNBS-C (yellow), RNBS-D (violet), GLPL (cyan); (b) LRR domain (PDB ID-2z64:A); Cartoon shows a horseshoe shaped receptor like structure. The parallel β-sheets forming the concave face of the 'horseshoe' are represented as yellow arrows; (c) Ramachandran plot for CN1-NBS model; (d) Ramachandran plot for CN1-LRR model.

The stereo-chemical quality and accuracy of the predicted model was assessed by Ramachandran map calculation by PROCHECK program (Fig. 4c). In the Ramachandran plot, residue classifications were done according to its regions in the quadrangle. The red regions in the graph show most allowed regions and the yellow represent allowed regions. Glycine is represented by triangles and others are represented by squares. The Ramachandran plot derived by PROCHECK showed 91% residues in the most favorable region, 7.5% in additionally allowed region, 0.5% in generously allowed region and only 1.0% in disallowed region as compared to the 1z6t:A template profile 93.1%, 6.1%, 0.0%, 0.0% respectively (Fig. 4c). Most of the amino acids were clustered in a $\phi - \psi$ distribution, consistent with right handed α-helix, indicating a good quality model.

PDB search result for suitable template detection to predict the 3-D structure of CN3-NBS-LRR and CN4-NBS-LRR showed maximum identity to chain A of *Lycopersicon esculentum* NBS domain (PDB ID: 2ft4:A) with an $E < 0.0$ and with significant identity (58%). FFAS03 pair wise alignment showed that NBS region from 1 to 263 of CN3-NBS-LRR had the expected homology with 63 to 320 of template chain A and 1 to 205 of CN4-NBS-LRR with 51 to 339 of template. ProtMod generated the desired PDB file of the predicted 3-D model of NBS region of CN3-NBS-LRR and CN4-NBS-LRR, which were analyzed in PyMol and subjected to SAVES for quality check. The 3-D structure of NBS region of CN3-NBS-LRR provides the complete details of different spatial arrangements and function of conserved motifs and the modeled cartoon structure of NBS region is shown in (On-line Suppl. Fig. 3a). The NBS domain consists of the same motifs identified in NBS domain of CN1-NBS-LRR (from N-terminus P-Loop, RNBS-A, kinase-2, RNBS-B, RNBS-C, RNBS-D and GLPL. The nucleotide binding cleft and relative arrangements of all the motifs of NBS were well conserved in homology model of CN3-NBS-LRR. Also, the catalytic cleft at the central region was found to be surrounded by conserved motifs like P-loop and GLPL. The P-loop consisted of β-sheets flanked by α-helix, kinase-2 of β-sheets and other motifs of α-helices. The stereo-chemical quality and accuracy of the predicted model, assessed by Ramachandran map calculation derived by PROCHECK program (On-line Suppl. Fig. 3b) showed 87.3% residues in the most favorable region, 8.1% in the additionally allowed region, 3.8% in the generously allowed region and only 0.8% in the disallowed region (On-line Suppl. Fig. 3b). Most of the amino acids were clustered in $\phi - \psi$ distribution, consistent with right handed α-helix, indicating a good quality model.

The modeled cartoon structure of NBS region of CN4-NBS-LRR is shown in On-line Suppl. Fig. 3c. Stereo-chemical quality and accuracy of the predicted model was

assessed by Ramachandran map calculation by PROCHECK program (On-line Suppl. Fig. 3d) showed that 81.6% of the residues were in the most favorable region, 12.3% in the additionally allowed region, 4.5% in the generously allowed region and only 1.7% in the disallowed region as compared to the 2ft4:A template profile (On-line Suppl. Fig. 3d). Most of the amino acids were clustered in $\phi - \psi$ distribution, consistent with right handed α-helix, indicating a good quality model.

The complete energy profile and Z-score of the predicted models of the three CN-NBS–LRR RGAs were calculated by PROSA-WEB search. The calculated values of the models ranged from –8.3 to –7.1, indicating the models to be acceptable.

3-D structure prediction and validation of LRR-domain

The LRR region of CN1-NBS-LRR was found to contain a short stretch of amino acid residues with leucine at every second and third position, repeated to form flexible, solvent exposed and parallel β sheets. The amino acid region between residues 54 to 477 showed sequence homology with Toll-like receptor (PDB ID-2z64: A) with significant score and E-value < 0. ProtMod search predicted the 3-D structure of this LRR region by comparing with Toll-like receptor-4. The LRR region of CNRGA contains a series of parallel β-sheets that form a concave face shaped like a horse shoe or banana (Fig. 4b). The number of repeats in this LRR region was approximately 22 and each LRR contained series of about 20–26 amino acids having a consensus sequence.

The Ramachandran plot showed 70.9% residues in the most favourable region, 27.1% in the allowed region and 1.4% in the generously allowed region and 0.6% in the disallowed region as compared to 2z64A template profile as 71.4, 26.8, 1.6 and 0.1%, respectively. 3-D structure of LRR generated using PyMol is represented in Fig. 4d.

Validation of expression of coconut RGAs

To analyze the expression levels of isolated RGAs in the leaves of root (wilt) disease resistant and susceptible coconut genotypes, two representatives (CN1-NBS-LRR and CN3-NBS-LRR) were used for expression analysis using RT-PCR. The primer pairs (Tab. 2) produced a single product (Fig. 5a). Real time profiling of the endogenous control

Tab. 2. Coconut NBS-LRR gene-specific primer pairs for qRT-PCR. S. No. – sample number.

S. No.	Primer Name	Sequence (5'–3')	Amplicon size (bp)
1	CN1F	CCAGCGAGGTAGAGATGAGG	130
	CN1R	GGGAACTCACGACCGAAGTA	
2	CN3F	CAAGCTCCGTATGATCAGCA	125
	CN3R	AATGTCATCGGCATCGTACA	
3	TUBF	ATGTTCAGGCGCAAGGCT	101
	TUBR	CTGCAACCGGGTCATTCA	

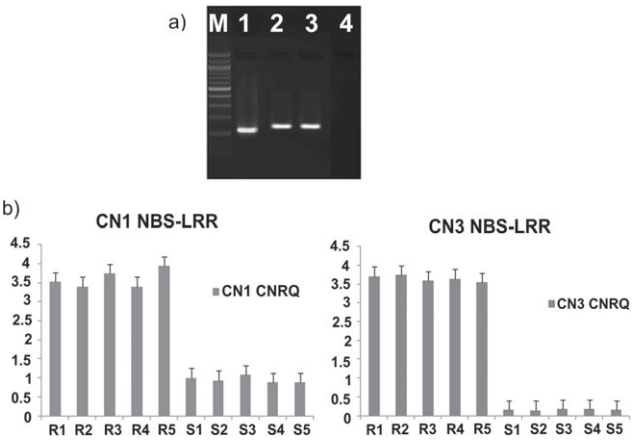


Fig. 5. qRT-PCR analysis of CN1-NBS-LRR and CN3-NBS-LRR in resistant (R) and susceptible (S) genotypes of coconut palm: a) RT-PCR gel image of α-Tubulin (1), CN1-NBS-LRR (2), CN3-NBS-LRR (3) and negative RT control (4); b) transcript level of CN1-NBS-LRR and CN3-NBS-LRR in resistant (R) and susceptible (S) genotypes. Data are average value ± SD, n = 3. R1–R5 and S1–S5 are biological replicates of resistant and susceptible Chowghat Green Dwarf leaves.

α-tubulin clearly indicated their stable expression in all the biological samples used for normalizing data (On-line Suppl. Figs. 4 a, b). Real time data indicated high expression levels of CN-NBS-LRR RGAs in resistant genotypes compared to susceptible genotypes (Fig. 5 b).

Discussion

Root (wilt) disease is one of the most devastating diseases of coconut palms in India. Breeding for disease resistance is one of the most important objectives in any crop breeding program, and is particularly relevant in the case of a long duration crop like coconut, where long generation time, low multiplication rate and requirement of a lot of land for experiments hinder conventional breeding efforts.

To defend against diverse pathogens, all plants have developed immunity by expressing *R* genes as receptors that recognize invading pathogens and mediate downstream signaling defense to either expel the pathogens or to prevent their spread (Staskawicz et al. 1995). New biotechnological approaches have provided much insight into the study of interaction between the plants and pathogens at different stages of infection. Identifying, screening and analyzing the diversity of *R* genes in various plant species are, therefore, vital to understanding the evolution of *R* genes and their functional acquisition. Several *R* genes against distinct pathogen types have been cloned and characterized (Shivaramakrishnan and Seetharama 2001) and this has facilitated our understanding on the mechanism of plant defenses and provided better insight in to effective management of plant diseases. In the present study, we report isolation and characterization of three putative resistant gene analogues from coconut, encoding NBS-LRR type proteins using a comparative genomics approach.

Degenerate primers-based PCR strategies have been used for isolating and identifying putative *R* genes from dif-

ferent plant species in earlier studies (Yu et al. 1996, Mago et al. 1999, Totad et al. 2005). Here we used an alternative, efficient PCR method by aligning and comparing, and designed specific motif based primers of date palm, whose genome has been sequenced. Four sets of primers designed from the conserved motifs (GLPL, P-loop and Kinase-2 etc.) were efficient enough to produce target specific fragments. The designed primers amplified fragments of 600–2000 bp, which, on cloning and sequencing, showed significant homology to known and well-characterized RGAs.

Most NBS-LRR *R* genes implicated in plant disease resistance responses are known to share a common structure and are also presumed to possess a common evolutionary origin. Many studies in different plants have revealed new insights into the molecular evolution of NBS-LRR type *R* genes (McDowell and Simon 2006). Translated polypeptides from the nucleotide sequences were subjected to homology search by BLASTX algorithm for two reasons. Firstly, it is well-known that the protein level homology search and comparison have shown more homology with NBS-LRR region of many RGAs than at the nucleotide level due to genetic code degeneracy issue (Totad et al. 2005). Additionally, amino acid sequences around the structural motifs display a higher degree of conservation. Multiple sequence alignment of the NBS domain of the three coconut RGAs and known *R* genes showed significant homology to the P-loop, kinase-2 and hydrophobic GLPL domain. All three coconut RGAs were classified into the non-TIR-NBS-LRR subfamily of the *R* genes. The NBS-LRR classes of *R* genes are classified into two subfamilies: the TIR-NBS-LRR subfamily, which is characterized by the presence of a highly conserved aspartic acid (D) or aspartate (N) as the last residue of the kinase-2 domain, and the non-TIR-NBS-LRR subfamily with a highly conserved tryptophan (W) as the last residue of the kinase-2 domain (Pan et al. 2000). N-terminal characteristic of only non-TIR class were present in coconut NBS-LRR RGAs isolated in the present study.

3-D structural modeling of R protein facilitates an understanding of its molecular function in recognizing pathogen effector protein. The crystal structure of NB-ARC domain of mammalian APAF1 (Riedl et al. 2005) was used in templates for homology modeling. In our study, the predicted homology modeling based 3-D structure of the NBS domain of CN-NBS-LRR revealed relative 3-D location of various conserved motifs present within this domain. CN-NBS-LRRs possessed conserved motifs that are located around the central catalytic cleft: the P-loop, kinase-2, hydrophobic GLPL etc. all of which serve to orient the ADP molecule (Chattopadhyaya and Pal 2008). The LRR domain is involved in protein–protein interactions, ligand binding (Jones and Jones 1996) and also determines the pathogen recognition specificity of NBS-LRR proteins (Shen et al. 2003, Ellis et al. 2007).

Up-regulation or higher expression level of CN-NBS-LRR genes in coconut root (wilt) resistant compared to susceptible genotypes, as revealed by qRT-PCR, suggests that the sequence similarity-based targeted gene identification approach has a high degree of accuracy. Breeding for resistance to coconut root (wilt) disease has been a long sought goal of coconut breeders. Identification of genes that confer resistance to root (wilt) disease would pave the way for molecular-based strategies to augment the effectiveness of breeding for disease resistant genotypes in coconut.

Acknowledgements

The work was supported by a grant from Indian Council of Agricultural Research (ICAR) and Kerala Biotechnology Commission (YIPB).

References

Altschul, S. F., Madden T. L., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997: Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. Nucleic Acids Research 25, 3389–3402.

Bailey, T. L., Williams, N., Misleh, C., Li, W. W., 2006: MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Research 34, 369–373.

Baldi, P., Patocchi, A., Zini, E., Toller, C., Velasco, R., Komjanc, M., 2004: Cloning and linkage mapping of resistance gene homologues in apple. Theoretical and Applied Genetics 109, 231–239.

Cannon, S. B., Zhu, H., Baumgarten, A. M., Spangler, R., May, G., Cook, D. R., Young, N. D., 2002: Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution 54, 548–562.

Chattopadhyaya, R., Pal, A., 2008: Three-dimensional models of NB-ARC domains of disease resistance proteins in tomato, *Arabidopsis*, and flax. Journal of Bimolecular Structure and Dynamics 25, 357–71.

Dangl, J. L., Dietrich, R. A., Richberg, M. H., 1996: Death don't have no mercy: Cell death programs in plant-microbe interactions. Plant Cell 8, 1793–1807.

Ellis, J. G., Lawrence, G. J., Dodds, P. N., 2007: Further analysis of gene-for-gene disease resistance specificity in flax. Molecular Plant Pathology 8, 103–109.

Ellis, J., Dodds, P., Pyror, T., 2000: Structure, function and evolution of plant disease resistance genes. Current Opinion in Plant Biology 3, 278–284.

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., Dubchak, I., 2004: VISTA: computational tools for comparative genomics. Nucleic Acids Research 32, W273–279.

Hammond-Kosack, K., Jones, J. D. G., 1997: Plant disease resistance genes. Annual Review of Plant Biology 48, 575–607.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., Godzik, A., 2005: FFAS03: A server for profile–profile sequence alignments. Nucleic Acids Research 33, 284–288.

Jones, D. T., 1999: Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology 292, 195–202.

Jones, D. A., Jones, J. D. G., 1996: The role of leucine-rich repeat proteins in plant defences. Advances in Plant Pathology 24, 89–167.

Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., Jones, J. D. G., 1994: Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266, 789–793.

Kanazin, V., Marek, L. F., Shoemaker, R. C., 1996: Resistance gene analogs are conserved and clustered in soybean. Proceedings of the National Academy of Sciences 93, 11746–11750.

Livak, K. J., Schmittgen, T. D., 2001: Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCt method. Methods 25, 402–408.

Mago, R., Nair, S., Mohan, M., 1999: Resistance gene analogues from rice: Cloning, sequencing and mapping. Theoretical and Applied Genetics 99, 50–57.

McDowell, J. M., Simon, S. A., 2006: Recent insights into R gene evolution. Molecular Plant Pathology 7, 437–48.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., Young, N. D., 1999: Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal 20, 317–332.

Pan, Q., Wendel, J., Fluhr, R., 2000: Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. Journal of Molecular Evolution 50, 203–213.

Rajesh, M. K., Jerard, B. A., Preethi, P., Thomas, R. J., Fayas, T. P., Rachana, K. E., Karun, A., 2013: Development of a RAPD-derived SCAR marker associated with tall-type palm trait in coconut. Scientia Horticulturae 150, 312–316.

Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., Shi, Y., 2005: The structure of the apoptotic-protease factor 1 bound to ADP. Nature 434, 926–934.

Ritter, C., Dangl, J. L., 1996: Interface between two specific pathogen recognition events mediated by distinct plant disease resistance genes. Plant Cell 8, 351–257.

Scofield, S. R., Tobias, C. M., Rathjen, J. P., Chang, J. H., Lavelle, D. T., Michelmore, R. W., Staskawicz, B. J., 1996: Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science 274, 2063–2065.

Shen, Q. H., Zhou, F. S., Bieri, S., Haizel, T., Shirasu, K., Schulze-Lefert, P., 2003: Recognition specificity and RAR1/SGT1 dependence in barley MLA disease resistance genes to the powdery mildew fungus. Plant Cell 15, 732–744.

Shivaramakrishnan, S., Seetharama, N., 2001: Identification and isolation of disease resistance genes in crop plants. Journal of Plant Biology 28, 25–38.

Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G., Jones, J. D. G., 1995: Molecular genetics of plant disease resistance. Science 268, 661–667.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731–2739.

Totad, A. S., Fakruddin, B., Kuruvinashetty, M. S., 2005: Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). Euphytica 143, 179–188.

Wiederstein, M., Sippl, M. J., 2007: ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407–W410.

Yu, Y., Buss, G., Maroof, M. A., 1996: Isolation of a super family of candidate disease resistance genes in soybean based on a conserved nucleotide binding site. Proceedings of the National Academy of Sciences 93, 11751–11756.

On-line Suppl. Tab. 1. E-values and amino acid identities of three coconut resistance gene analogues (CN RGAs) to the top hit Genebank accessions revealed by BLASTX searches.

	Top BLASTX hit, organism with Genebank accession number	E-value	Amino acid identity (%)
Coconut NBS-LRR CN1	NBS-LRR resistance protein partial, *Cocos nucifera* (AGI96386.1)	0.0	100
	NBS-LRR like protein, *Oryza sativa Japonica* group (AANO3742.1)	0.0	46
	Putative disease resistance protein RGA4, *Triticum urartu* (EMS58751.1)	0.0	45
Coconut NBS-LRR CN3	NBS-LRR type resistance protein, partial, *Oryza sativa Japonica Group* (AB96985.1)	6e-167	66
	Putative disease resistance protein RGA3-like, *Brachypodium distachyon* (XP_003566098.1)	3e-166	67
	Putative disease resistance RPP13-like protein 1, *Aegilops tauschii* (EMT33711.1)	4e-161	66
Coconut NBS-LRR CN4	NBS-LRR disease resistance protein, *Musa balbisiana x Musa textilis* (CAP66367.1)	6e-91	67
	Disease resistance RPP13-like protein 4-like, *Brachypodium distachyon* (XP_003564401.1)	7e-54	51

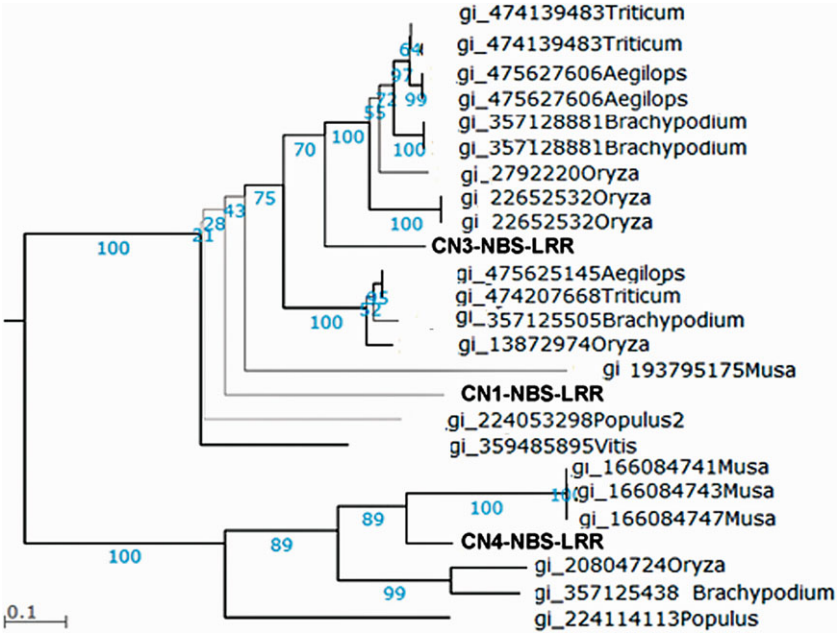
On-line Suppl. Tab. 2. Disease resistant motifs viewed in (a) Pfam and (b) PRINTS.

a) Pfam

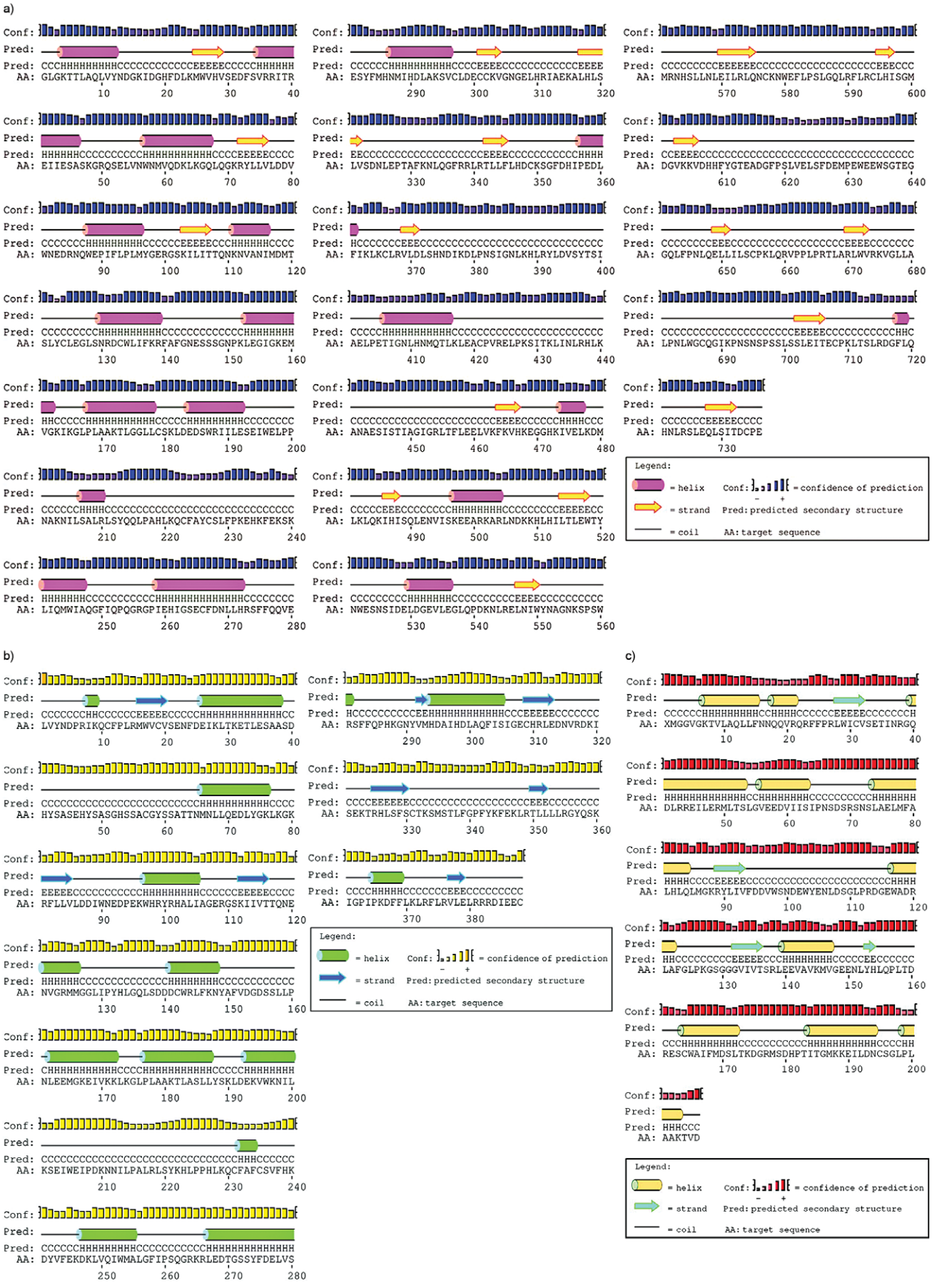
Sequence	Description	Entry type	clan	Envelope		Alignment		HMM		Bit-Score	E-value
				start	end	start	end	from	to		
CN1-NBS-LRR	NB-ARC domain	Domain	CL0023	1	256	1	254	29	285	231.8	6.6e-69
	LRR domain	Family	CL0022	365	405	366	404	2	40	40.6	1.2e-10
CN3-NBS- LRR	NB-ARC domain	Domain	CL0023	1	263	1	261	3	254	201.5	5.4e-29
CN4-NBS-LRR	NB-ARC domain	Domain	CL0023	1	205	1	205	2	170	200.3	4.6e-007

b) PRINTS

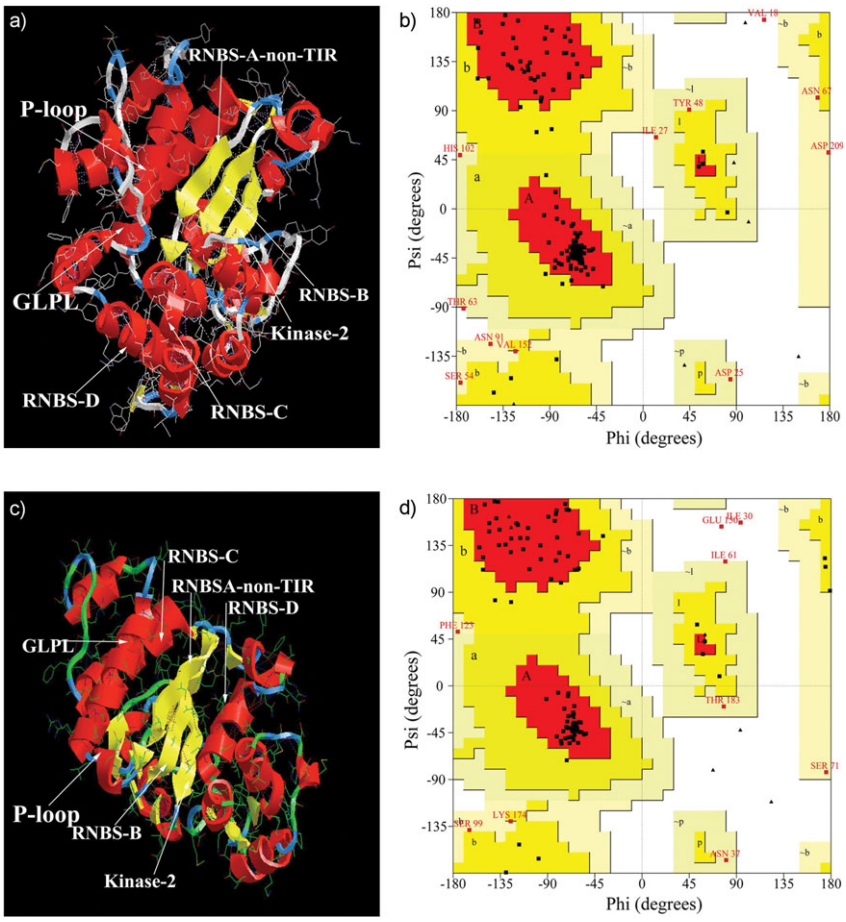
Finger Print	ID score	P VAL	PF Score	Motif No	Sequence	Position	Length	Low	High
Disease	50.83	1.08-e07	386	1 of 4	QLQGKRYLLVLLDW	64	15	0	0
Resistant	46.00	2.39-e06	296	2 of 4	IKGLPLAAKTLGGLL	167	15	0	0
Motifs CN1	36.32	2.9-e04	166	3 of 4	IKIKCLRVLDLSHNDIK	362	17	0	0
Coconut	51.3	2.31e-08	402	2 of 4	KLKGKRFLLVLDDIW	76	15	0	0
RGA CN3	42.8	8.19e-06	206	3 of 4	LKGLPLAAKTLASLL	173	15	0	0
Coconut	44.67	3.67e-07	342	2 of 4	QLMGKRYLIVFDDVW	84	15	0	0
RGA CN4	33.00	2.98e-05	236	3 of 4	CSGLPLAAKTVD	195	15	0	0



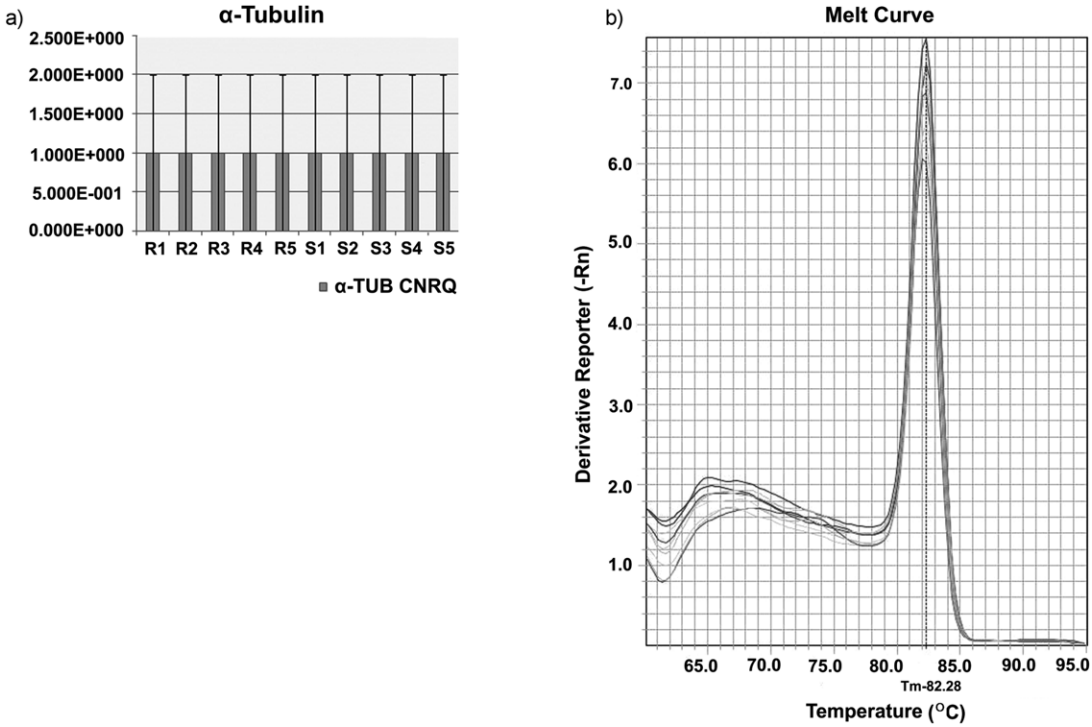
On-line Suppl. Fig. 1. Phylogenetic tree constructed based on the selected resistance gene analogues. Numbers by the branch joints are bootstrap values (out of 1000) that support the clustering. The scale bar represents substitutions per site.



On-line Suppl. Fig. 2. Secondary structure prediction of (a) CN1-NBS-LRR; (b) CN3-NBS-LRR; (c) CN4-NBS-LRR. The prediction line (Pred) denotes the predicted conformation for each residue such as helical (H), random coil (C) and extended (E). The confidence (Conf) indicates the reliability of the prediction for each position.



On-line Suppl. Fig. 3. (a) 3-D model of NBS domain of CN3-NBS-LRR. 3-D cartoon view revealed the relative position of different conserved motifs from N-terminus: P-Loop, RNBS-A,Kinase-2, RNBS-B, RNBS-C, RNBS-D, GLPL; (b) Ramachandran plot for CN3-NBS model; (c) Molecular 3-D- model of NBS domain of CN4-NBS-LRR. 3-D cartoon view revealed the relative position of different conserved motifs from N-terminus: P-Loop, RNBS-A,Kinase-2, RNBS-B, RNBS-C, RNBS-D, GLPL; (d) Ramachandran plot for CN4-LRR model.



On-line Suppl. Fig. 4. (a) Real time profiling of α-tubulin in resistant (R1-R5) and susceptible (S1-S5) samples; (b) Melt curve analysis.