TÜBİTAK

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2015) 39: 489-502
© TÜBİTAK
doi:10.3906/tar-1409-75

Identification of expressed resistance gene analog sequences in coconut leaf transcriptome and their evolutionary analysis

Muliyar Krishna RAJESH[1,*], Kaitheri Edathil RACHANA[1], Sudalaimuthu Asari NAGANEESWARAN[1], Rahman SHAFEEQ[1], Regi Jacob THOMAS[2], Mohammed SHAREEFA[2], Babu MERIN[2], Karun ANITHA[1]

[1]Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod, Kerala, India

[2]ICAR-Central Plantation Crops Research Institute (Regional Station), Kayamkulam, Kerala, India

Received: 15.09.2014 • **Accepted:** 27.11.2014 • **Published Online:** 12.06.2015 • **Printed:** 30.06.2015

Abstract: Coconut, an important crop of the tropics and subtropics, is susceptible to a variety of diseases and enhancing disease resistance has been the major goal of coconut breeding programs all over the world. Information on the presence and distribution of disease resistance (*R*) genes, which play a primary role in the detection of pathogens and the initiation of specific plant defenses, is scarce in coconut. In this study, RNA-Seq was used to generate the transcriptome of leaf samples of coconut root (wilt) disease-resistant cultivar Chowghat Green Dwarf. Comprehensive bioinformatics analysis identified 243 resistance gene analog (RGA) sequences, comprising 6 classes of RGAs. Domain and conserved motif predictions of clusters were performed to analyze the architectural diversity. Phylogenetic analysis of deduced amino acid sequences revealed that coconut NBS-LRR type RGAs were classified into distinct groups based on the presence of TIR or CC motifs in the N-terminal regions. Furthermore, qRT-PCR analysis validated the expression of randomly selected NBS-LRR type RGAs. The results of this study provide a sequence resource for development of RGA-tagged markers in coconut, which would aid mapping of disease-resistant candidate genes. In addition, we hope that this study will provide a genomic framework for isolation of additional RGAs in coconut via comparative genomics and also contribute to the deciphering of mode of evolution of RGAs in Arecaceae.

Key words: Coconut, disease, phylogeny, resistance gene analogs, transcriptome

1. Introduction

Coconut (*Cocos nucifera* L.) is an economically important and ecofriendly palm widely grown in tropical and subtropical regions that provides food and shelter. Coconut is closely associated with many socioeconomic aspects in lives of millions of people, who directly or indirectly depend on this crop. Each and every part of this palm is highly valuable and can be used for a myriad of functions (Persley, 1990). Due to multifarious uses of this monotypic species, it makes the term 'tree of life' meaningful and apt to coconut (Foale and Ashburner, 2004), the industry providing not only food, income, and raw materials, but also providing employment. Additionally, the palm can be used for ornamental purposes and different parts possess potential medicinal properties (DebMandal and Mandal, 2011).

Coconut palm, unfortunately, is affected by many diseases caused by different pathogens, which directly or indirectly reduce their yield, either by killing or debilitating the palm and hindering production increase. Root (wilt) disease causes the maximum damage of coconut palms in India, with yield loss of up to even 80% in severe cases (Ramjegathesh et al., 2012). Crop improvement programs in coconut, therefore, mainly emphasize the development of disease-resistant cultivars through selective breeding. Long generation time, practical difficulties of vegetative propagation, difficulties of seed production through controlled and systematic pollination, and reduced number of seeds per bunch make breeding efforts in coconut time-consuming and cumbersome.

Plants have developed multiple and sophisticated defensive mechanisms to protect themselves from the attack of various pathogens like fungi, nematodes, bacteria, and viruses; these mechanisms have been formed by the perpetual evolutionary battle waged by plants against the pathogenic invaders. One of the main defense mechanisms of the plant system is characterized by the gene-for-gene interaction concept involving a specific plant resistant (*R*) gene and a cognate pathogen avirulence (*Avr*) gene (Flor, 1971). This kind of specific resistance is often associated with localized hypersensitive response in the plant cells proximal to the site of infection triggered

* Correspondence: mkraju.cpcri@gmail.com

by the recognition of Avr protein by the *R* gene product (Dangl et al., 1996; Heath, 2000). This interaction also triggers various downstream signal transduction cascades for immediate defense mobilization to rapidly suppress pathogen growth (Bent and Mackey, 2007).

Plant genomes contain a large number of *R* genes against various pathogens. Resistance gene analogs (RGAs) or resistance gene candidates (RGCs) are useful putative derivatives of known *R* genes (Kanazin, 1996), identified by their structure and characterized in a wide range of plant species. Sequential and comparative analyses of different disease resistance protein products have revealed the presence of some highly conserved amino acid sequences, which provides signature identity to these types of genes from others in the genome (Hammond-Kosack and Jones, 1997). These conserved sequences of different resistance genes have also been successfully utilized for designing degenerate primers for PCR amplification to find similar regions from other plants (Leister et al., 1996; Gururani et al., 2012).

Most of the cloned and functionally identified RGAs within the plant kingdom to date have been grouped into distinct classes based on the predicted protein structure (Jones and Dangl, 2006). The NBS-LRR type constitutes the largest class of RGAs (Meyers et al., 2003) and is characterized by the presence of 2 conserved domains, a central nucleotide binding site (NBS) and a C-terminal series of leucine-rich repeats (LRRs) (Hulbert et al., 2001). This class is further divided into 2 based on the N-terminus domain structure: the coiled-coiled (CC) domain and the Toll/interleukin-1/receptor (TIR) domain, with homology to the intercellular signaling domain of *Drosophila* Toll and mammalian interleukin-1 receptor (Hammond-Kosack and Jones, 1997). The TIR domain is thought to be absent or rare in monocotyledonous plants, but it has been shown to be present in all dicotyledonous taxa studied to date (Ellis and Jones, 1998; Tarr and Alexander, 2009). The receptor-like kinases encode proteins comprising an extracellular receptor domain connected to a cytoplasmic serine-threonine kinase domain and are further divided into several subclasses based on the presence of various extracellular receptor domains. The receptor domains can contain LRRs and lectin-like or additional domains, which can potentially bind pathogen-derived peptides (Vossen et al., 2013). The protein kinase class of *R* genes encodes only serine-threonine kinases, no transmembrane domains or LRRs (Vossen et al., 2013).

Either for cloning of *R* genes or for utilization of *R* genes in marker-assisted breeding programs for disease resistance, it is imperative to generate molecular markers positioned close to *R* genes in the plant genome. Degenerate primer-based PCR strategy, based on conserved domains, has been successfully utilized to isolate RGAs from many plant species (Gururani et al., 2012). However, since plant genomes are known to contain hundreds of RGAs, PCR-based approaches often result in complex banding patterns (Vossen et al., 2013). Marker techniques like DArT (Soriano, 2009) and AFLP (Li et al., 1998) have been employed for de novo mapping of *R* genes. NBS profiling techniques, which utilize a single degenerate primer in combination with an adapter ligated to a restriction enzyme site, have also been developed for genetic mapping of *R* genes (van der Linden et al., 2004). The advent of next generation sequencing technologies has dramatically accelerated research in genome analysis. Data-mining approaches have been developed and utilized for identification of RGAs in ESTs of sugarcane (Rossi et al., 2003), wheat (Dilbirligi and Gill, 2003), and *Medicago* (Zhu et al., 2002). With the whole-genome sequences of many plants becoming available, genome-wide *R* gene detection has been undertaken in rice (Monsi et al., 2004), *Arabidopsis* (Meyers et al., 2003), *Medicago* (Ameline-Torregrosa et al., 2008), and tomato (Andolfo et al., 2014).

Precise understanding of disease resistance mechanisms and development of molecular techniques for identifying these mechanisms may be beneficial in breeding programs for disease resistance in coconut. To our best knowledge, no studies regarding systematic analysis of *R* genes have been reported so far. The coconut genome is yet to be sequenced. Transcriptomic resources can be an useful substitute for gene discovery in species that lack completely sequenced genomes. *R* gene-like sequences have been mined from transcriptomic sequences and expressed sequence tags (ESTs) of *Coffea* spp. (Alvarenga et al., 2010), *Curcuma longa* (Joshi et al., 2011), and *Phaseolus vulgaris* (Liu et al., 2012).

In India, root (wilt) disease is the major threat to coconut production. Screening trials of coconut cultivars against root (wilt) disease revealed that no cultivars possessed complete resistance to the disease (Jacob et al., 1998). Intensive surveys carried out in farmers' plots in the disease hotspot areas of Kerala State in southern India revealed that 75% of Chowghat Green Dwarf (CGD) palms were disease-free, whereas the predominant West Coast Tall palms standing in the same plot had disease incidence to the extent of 80% or more. These studies indicated the presence of a higher level of resistance in CGD palms to coconut root (wilt) disease (Nair et al., 2004).

In this study, we have carried out RNA-Seq on an Illumina HiSeq 2000 platform and de novo assembly of leaf transcriptome of coconut root (wilt) disease-resistant cultivar CGD. A total of 243 *R* gene sequences were mined from generated coconut transcriptomic data and grouped into different classes based on homology search and by putative domain prediction. Phylogenetic and evolutionary relationships among the coconut RGAs were also evaluated. Expression patterns of selected NBS-LRR type RGAs revealed differential expression patterns in

disease-resistant CGD genotypes compared to susceptible ones.

2. Materials and methods

2.1. Plant materials

Root (wilt) disease-resistant CGD palms maintained at the Institute Farm of the Central Plantation Crops Research Institute (CPCRI) regional station in Kayamkulam, Kerala, India, were utilized for this study. These palms were planted in June 1998 and started flowering from September–December 2001. Leaf samples were collected in 2012 from 3 disease-resistant palms. These 3 palms were serologically tested to ascertain resistance against root (wilt) disease (Sasikala et al., 2010).

2.2. RNA extraction

Spindle leaf samples collected from the 3 disease-resistant palms were snap-frozen immediately in liquid nitrogen. Total RNA was extracted using the Tri-Reagent (Sigma) and treated with DNase I (Fermentas) according to the manufacturer's instructions. The quality and the purity of the extracted RNA were assessed by OD 260/280 ratio and RNA integrity number was analyzed using an Agilent Technologies 2100 Bioanalyzer with the Agilent RNA chip.

2.3. Transcriptome sequencing

Equal amounts of independently isolated RNA from the 3 disease-resistant palms were pooled together and approximately 5–10 µg of total RNA was used to prepare the RNA-Seq library using TruSeq RNA Sample Prep Kits (Illumina) according to the manufacturer's instructions. In short, poly $(A)^+$-containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were then used to synthesize first-strand cDNA using reverse transcriptase and random primers followed by second-strand cDNA synthesis using DNA polymerase I and RNase H. These cDNA fragments were then subjected to an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products were purified and enriched with PCR to create the final cDNA library. Bioanalyzer plots were used at every step to assess mRNA quality, enrichment success, fragmentation sizes, and final library sizes. The size distribution of the sequencing library was determined by gel electrophoresis. The quantity of the library was assessed before sequencing by qPCR. After the library was constructed, a paired end run was performed on the Illumina HiSeq 2000 platform to obtain 2 × 100 bp reads.

2.4. De novo transcriptome assembly and evaluation

The fastq files were trimmed before performing assembly using a Perl script by removing the first 15 bases from all reads to avoid specific sequence bias. Furthermore, reads with average base quality score of <Q20 Phred score were discarded from the analysis. The trimmed reads were then assembled using SOAPdenovo v1.05 (http://soap.genomics.org.cn/soapdenovo.html) with default options.

2.5. Functional annotation and classification

Assembled contigs were further converted into scaffolds. To reduce the redundancy between the scaffolds, we further assembled and obtained unigenes using the CAP3 de novo assembly program (http://seq.cs.iastate.edu/cap3.html). BLAST-based similarity searches of the assembled coconut unigenes were carried out against both date palm and oil palm genomes and unigenes were then assigned function based on similarity. For validation, coconut unigenes were also subjected to tBLASTn alignment using publically available known *R* genes of different classes using HMMER (http://hmmer.janelia.org/). The tBLASTn alignment was used to predict the open reading frame of each class. Coconut unigenes showing matches to known *R* genes, with a cut off value e^{-20}, were scrutinized and used for homology screening of *R* genes in GenBank (NCBI) using BLASTx. After this search, the threshold expectation value (E-value) was set to e^{-10} for excluding spurious BLAST hits.

2.6. Family, domain, and signature motif prediction

The signature motifs of deduced amino acid sequences of coconut RGAs were obtained from the domain profiles retrieved from InterProScan 4.8 stand-alone version (integrated database of PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIR Super Family, SUPERFAMILY Gene3D, PANTHER, and HAMAP; http://www.ebi.ac.uk/interpro/) and from the COILS program (http://embnet.vital-it.ch/software/COILS_form.html). A stringent threshold of 0.9 was used to specifically detect CC domains. Similarly, R proteins were also identified based on the kinase domains in Pfam. To analyze the fine structure and diversity of each class of proteins, the amino acid sequences were again subjected to Multiple Expectation Maximization for Motif Elicitation (MEME) software and the Motif Alignment and Search Tool (MAST) (http://meme.sdsc.edu/meme/website/intro.html). Consensus sequence was depicted using the web-based application WebLogo version 2.8.2 (http://weblogo.berkeley.edu/) using default settings.

2.7. Phylogenetic and evolutionary analysis

To understand the phylogenetic relationship among the coconut RGAs, multiple sequence alignment of the deduced amino acid sequences of the identified coconut *R* genes was carried out using CLUSTALW (http://www.genome.jp/tools/clustalw/) and manually edited using Genedoc (http://www.nrbsc.org/gfx/genedoc/). The aligned sequences were subjected to the MAFFT

program (http://mafft.cbrc.jp/alignment/software/) for construction of an N-J phylogenetic tree. The sequence divergences among deduced amino acid sequences of coconut RGAs were estimated by computing the rate of nonsynonymous versus synonymous substitutions (Ka/Ks) using KaKs_calculator 2.0 (Wang et al., 2010).

2.8. RNA isolation, cDNA synthesis and RT-PCR analysis

Total RNA from leaf samples of resistant (5 palms) and susceptible (4 palms) CGD genotypes from the Institute Farm at the CPCRI regional station in Kayamkulam was isolated using a QIAGEN plant RNA mini kit and DNase l (QIAGEN) treatment was performed. Integrity, concentration, and quality checks of the extracted RNA were carried out. Synthesis of cDNA was performed using a High Capacity RNA-to-cDNA kit (Applied Biosystems). Gene-specific primer pairs were designed using Primer 3.0 software (http://primer3.ut.ee/) to amplify 6 randomly selected coconut NBS-LRR type RGAs. The expression pattern of alpha-tubulin, the endogenous control, was studied in the 9 palms sampled. PCR amplification was carried out with a reaction mixture containing 2 µL of 10X reaction buffer, 1 µL of synthesized cDNA, 1 U of Taq polymerase, 2 µL of 25 mM $MgCl_2$, 0.2 µL of 10 mM dNTP mix, and 1 µM of each forward and reverse primer. RT-PCR reactions were carried out using a BIO-RAD 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, and 1 min at 72 °C; and a 10-min extension at 72 °C. Each PCR pattern was verified by triple replicate experiments and a mixture without template was used as the negative control.

2.9. Quantitative RT-PCR and data analysis

Quantitative RT-PCR (qRT-PCR) reactions were carried out in a total volume of 25 µL containing 12.5 µL of 2X SYBRGreen PCR Master Mix (Applied Biosystems), 1 µL of each primer, 1 µL of template (10× diluted cDNA from samples), and 9.5 µL of sterile distilled water. The thermal conditions were as follows: initial holding stage at 52 °C for 2 min and 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. All reactions were performed in triplicate in 48-well reaction plates using a Step One Real Time PCR machine (Applied Biosystems). Alpha-tubulin (KP126009) was used as an endogenous control for the expression analysis of coconut NBS-LRR type RGAs (Rajesh et al., 2014). Five biological replicates of resistant CGD palms and 4 biological replicates of susceptible palms were considered for the expression analysis. Three technical replicates were included for each sample. No-template controls were also included for each gene to detect any spurious signals arising from amplification of any DNA contamination or primer dimer formed during the reaction. The results (Cq values) were generated using sequence detection software SDS version 1.1 (Applied Biosystems). The Cq values were imported to qBASE software (Biogazelle) and relative fold changes in transcript levels in the resistant genotypes compared to susceptible genotypes were generated using the comparative Cq method. Normalization was done using the alpha-tubulin gene. Finally, calibrated normalized relative quantity (CNRQ) values were exported from the qbase software and statistically investigated using Student's t-test. Significance was accepted at $P < 0.05$. These CNRQ values with the corresponding standard errors were then plotted.

3. Results

3.1. Identification of RGAs from transcriptomic sequences

A total of 102,579,213 paired-end reads (~17.64 Gb) were generated by Illumina sequencing of leaves of root (wilt) disease-resistant CGD cultivar (SRX 436961). After clustering and assembly, these sequences were assembled into 254,302 contigs, 159,932 scaffolds, and 130,942 unigenes with sequence size of more than 100 bp. These unigenes were used for identification of RGAs.

Full-length protein sequences of known *R* genes were used to perform tBLASTn searches against coconut unigenes. A total of 243 nonredundant unigenes were identified that showed homology to proteins encoded by known *R* genes, with scores above or equal to 100 and E-values below or equal to $1e^{-10}$. The putative functions of these 243 RGAs were identified by a BLASTx search against the GenBank databases. Among these, 31 belonged to the NBS-LRR type, 3 belonged to the CC-NBS-LRR type, 1 belonged to the TIR-NBS-LRR type, 181 belonged to the serine/threonine type protein kinase class, 12 belonged to the TM-LRR class, and 15 belonged to the kinase-TM-LRR type.

3.2. Motif characterization of coconut RGAs

To analyze the structural diversity of conserved motifs in coconut RGAs, the conserved motifs and their relative positions were predicted using MEME software and SMART analysis.

The 5 major conserved motifs (P loop, RNBS A-non-TIR, Kinase 2, Kinase 3A, and GLPL), which determine the structural characteristics of the NBS-domain of NBS-LLR type RGAs, were found in the deduced amino acid sequence of the NBS-LRR type RGAs (Figure 1). A search for the protein signature motif through SMART analysis verified the presence of the NB-ARC domain (PF00931) and LRR domain (PF00560). The CC type NBS-LRR RGAs also possessed the 5 major conserved motifs (P loop, RNBS A-non-TIR, Kinase 2, Kinase 3A, and GLPL) (Figure 2). A search for protein signature motif through SMART analysis confirmed the presence of the NB-ARC domain (PF00931), LRR domain (PF00560), and CC domain (PF04942).

P loop
(E value: 1.5e-101)

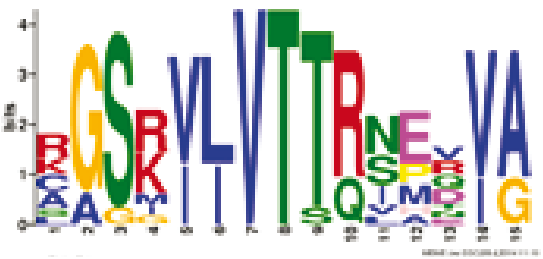
GG[LVI]GKTTL[AV][QK][LK][VI]Y[ND][DH]

RNBS-A-non TIR
(E value: 1.0e-066)

HF[ES]LRAW[VI]CVS[QE][DN]FD

Kinase 2
value: 6.7e-050)

ALR[GDN]K[RK][FY][LF]LVLDD[VL]W



RNBS-B (Kinase 3A)
(E value: 1.0e-040)

RGS[KR][VI][LI]VTT[RQ][NS]Ex[VI][AG]

HD-GLPL
(E value: 4.4e-033)

[IM]G[KMR][EK]I[VS][KE]K[CL][GDH]GLPLA

Figure 1. Sequence logo of prominent conserved motifs in the NBS region of coconut NBS-LRR type RGAs along with their corresponding E-values.

Ploop
(E value: 1.9e-010)

GG[ILV]GKTTLAQL[VI]YND

RNBS-A-non TIR
(E value: 2.1e-009)

[HC]F[DEP][ILQ]R[AMN]W[VC][CY]V[SP][DE]NFD

Kinase 2
(E value: 1.3e-005)

V[LF]DD[VI]W[DN]E[TD][PG][AES][KH]W

RNBS-B (Kinase 3A)
(E value: 1.3e-003)

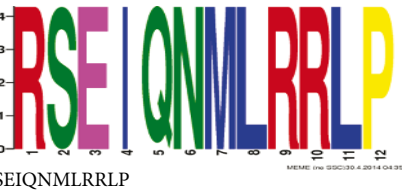
VTT[QR][SN][AEP][NRV]V[AG][DER]
[IMV]M[GR][TG][ML]

HD-GLPL
(E value: 9.7e-001)

I[CSV][EKR]KL[GHK]GLPLA

CC Domain
(E value: 2.7e-002)

I[QY][AT][RLN][IL][DEN][DY][AM]EE



CC Domain
(E value: 3.9e-000)

RSEIQNMLRRLP

Figure 2. Sequence logo of prominent conserved motifs in coconut CC-NBS-LRR type RGAs along with their corresponding E-values.

The N-terminal regions of one of the coconut RGAs showed sequence similarity to TIR type NBS-LRR proteins, the presence of which is very rare in monocots. The presence of a conserved aspartate (D) residue at the end of the Kinase 2 domain established its classification with TIR subclass of NBS-LRR type *R* genes. A search for the protein signature motif through SMART analysis revealed the presence of the TIR domain (PF01582). The putative motifs characterizing TIR domain (TIR-1, TIR-2, TIR-3) were clearly predicted using MEME (Figure 3). Both TIR and non-TIR type NBS-LRR resistance proteins shared other conserved domains like P-loop, RNBS-B, and GLPL.

The protein kinase type coconut RGAs possessed 5 conserved motifs (Figure 4). A search for protein signature motifs of these RGAs through SMART analysis proved the presence of the serine/threonine domain (PF07714).

3.3. Phylogenetic relationships of coconut RGAs

The phylogenetic tree of NBS-LRR type coconut RGAs revealed clear clustering of all the NBS-LLR type RGAs together, while the CC and TIR-type NBS-LRR RGAs were distinctly clustered (Figure 5). The phylogenetic tree of serine/threonine kinase type RGAs is given in Figure 6. Protein kinase type RGAs in coconut were found to be distributed in 2 major clades.

3.4. Comparing evolutionary rates among coconut RGAs

To examine the evolutionary forces acting on coconut and palm RGAs, the ratio of nonsynonymous (Ka) to synonymous nucleotide substitutions (Ks) was calculated. A Ka/Ks ratio of >1 indicates a positive or diversifying selection, implying the accumulation of advantageous mutations during the course of evolution. In contrast, a Ka/Ks ratio of <1 indicates a purifying or negative selection, implying the elimination of most of the nonsynonymous substitutions. In the current study, for coconut RGAs, all Ka/Ks values were <1, suggesting the operation of purifying selection (Table 1).

3.5. Validation of expression of selected RGAs

In order to confirm the expression patterns of 6 randomly selected NBS-LRR type RGAs, cDNA was prepared from leaf samples of 5 root (wilt) disease-resistant and 4 susceptible CGD genotypes. RT-PCR was used to amplify the 6 selected NBS-LRR type RGAs and alpha-tubulin using the designed primers (Table 2). All 6 primer pairs produced a single product. The amplicons were cloned and sequenced and sequence alignment revealed no differences between the original sequences and those that were amplified from cDNA (data not shown). RT-PCR of 9 samples (5 root (wilt) disease-resistant and 4 susceptible CGD genotypes) revealed uniform expression pattern of alpha-tubulin, the endogenous control (Figure 7). We then performed qRT-PCR using the designed primers to examine the differential expression of RGAs in resistant

G[VI]GKTTLAK[EA][FL][FY]N

Motif 1
(E-Value: 3.5e-050)

LDDVDHV[DE]Q[LV]EAL

Motif 2
(E-value: 2.4e-023)

F[VL][EK]A[CA]GGLPL[SAG]L

Motif 3
(E-value: 2.2e-021)

VFPSFSGVDVRKTFL

TIR 1
(E-value: 3.0e-009)

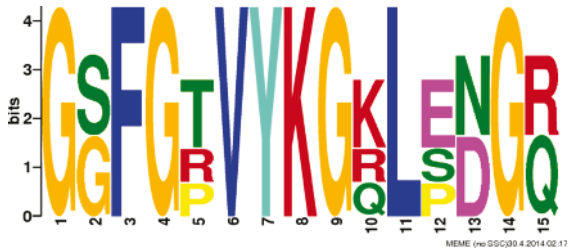
[KQ][NM][LV][VI]P[IV][FY][FY]
D[LV][SD][PA][SAF][DE][CV]

TIR 2
(E-value: 3.0e-009)

[TS]K[KN][SY][FA][KSG][NS]
[PT][YW][CTS][IL][EN]EL[RVK]

TIR 3
(E-value: 3.4e-037)

Figure 3. Sequence logo of prominent conserved motifs in coconut TIR-NBS-LRR type RGAs along with their corresponding E-values.

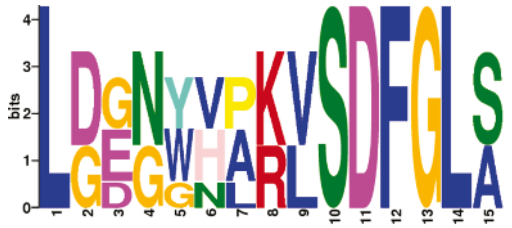


G[GS]FG[TPR]VYKG[KQR]L[EPS][DN]G[QR]

Motif 1
(E value: 4.6e-008)

[HQ]LTVKSDVY[GS][FL]GVV[FM]

Motif 2
(E value: 1.1e-013)

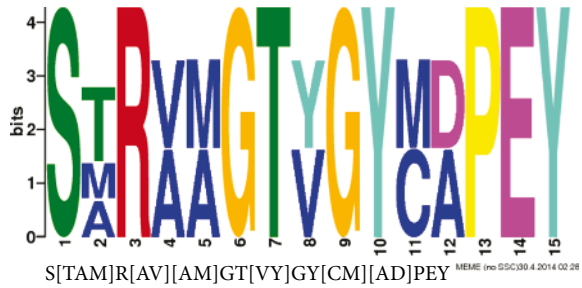


L[DG][EGD][NG][WYG][HVN][APL][KR][VL]SDFGL[SA]

Motif 3
(E value: 3.1e-010)

A[NV]PP[IV]I[HY]RD[FI]KSSNI

Motif 4
(E value: 1.3e-021)



S[TAM]R[AV][AM]GT[VY]GY[CM][AD]PEY

Motif 5
(E value: 2.7e-016)

Figure 4. Sequence logo of prominent conserved motifs in coconut serine/threonine protein kinase type RGAs along with their corresponding E-values.

and susceptible genotypes. Melt curve analysis indicated that the 6 primer pairs amplified a specific PCR product. The analysis indicated significantly higher expression of the RGAs in disease-resistant genotypes compared to susceptible genotypes (Figure 8). However, the rate of expression varied between the different resistant genotypes.

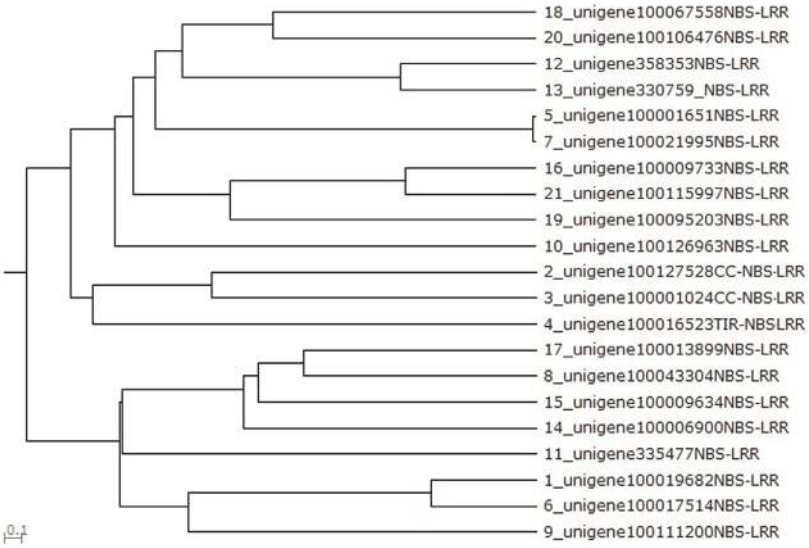


Figure 5. Phylogenetic N-J tree generated from deduced amino acid sequence alignment of NBS-LRR type coconut RGAs. The scale bar corresponds to a distance of 0.1 amino acid substitution per site.

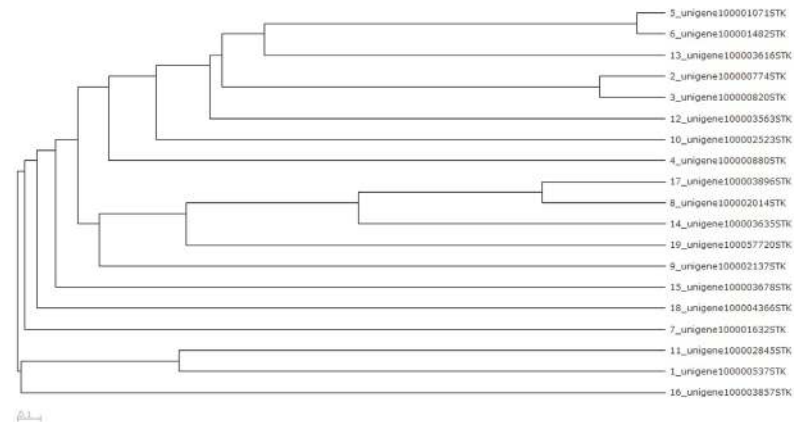


Figure 6. Phylogenetic N-J tree generated from deduced amino acid sequence alignment of serine/threonine kinase type coconut RGAs. The scale bar corresponds to a distance of 0.1 amino acid substitution per site.

4. Discussion

Development and use of resistant cultivars represents the most effective and economic approach to control plant pathogens. Therefore, identification of new sources of resistance has been the top priority in crop breeding programs for disease resistance. Plant pathogen interaction involves activation of signal transduction cascades resulting in the deployment of an array of defense responses against microbial invaders, which enables the plant to circumvent further disease infection (Anil et al., 2014). Induction of these defense signaling pathways in plants involves recognition of specific pathogen effectors, encoded by *Avr* genes, by products of specialized host genes, called the *R* genes (Dangl and Jones, 2001). In the last 2 decades, many *R* genes have been cloned from a large number of plant species, with the majority encoding for NBS-LRR proteins (Holt et al., 2003). Structural similarity among different *R* genes conferring resistance to diverse pathogens indicates the highly conserved nature of the plant resistance mechanism (Dangl et al., 1996). RGAs have been utilized to develop molecular markers for tagging and mapping disease resistance traits in many plant species, for isolation of full-length *R* genes, and to analyze the evolutionary patterns of *R* genes in different plant species (Gururani et al., 2012). Utilizing plant *R* genes for development of disease resistant varieties is an

Table 1. Average evolutionary rates (Ka, Ks) and their ratios (Ka/Ks) in different domains of coconut RGAs, and P-values between different domains.

Domain	Ka	Ks	Ka/Ks
TIR	0.36	2.09	0.17
NBS	0.48	2.15	0.23
LRR	1.17	3.37	0.35
CC	0.62	1.97	0.31
Kinase	1.18	3.40	0.35
Domain	P-value		
NBS vs. TIR	4.45e-020	2.67e-035	4.89e-023
NBS vs. LRR	8.24e-070	4.54e-011	5.49e-090
NBS vs. CC	4.67e-050	3.08e-013	6.52e-008
NBS vs. Kinase	7.42e-006	5.32e-070	4.34e-023
LRR vs. TIR	2.39e-023	1.06e-007	1.32e-025
LRR vs. CC	4.57e-080	6.31e-008	2.54e-003
LRR vs. Kinase	3.21e-005	5.32e-076	1.25e-005
CC vs. TIR	2.53e-011	6.34e-010	2.56e-002
CC vs. Kinase	3.67e-080	5.21e-003	8.36e-001
Kinase vs. TIR	4.37e-040	6.21e-001	1.24e-007

ecofriendly alternative to employment of chemical control measures for disease control (Gururani et al., 2012).

The root (wilt) disease is one of the most devastating diseases of coconut palms in India. All the available coconut cultivars are susceptible to the root (wilt) disease (Jacob et al., 1998). The CGD cultivar has been reported to possess a relatively higher level of resistance compared to other cultivars against the root (wilt) disease (Nair et al., 2004). Identification of RGAs has been considered as a valuable and rapid strategy for generation of genomic resources for development of cultivars with durable resistance against different diseases in crop breeding programs (Sharma et al., 2009). In the first study aimed at elucidating the diversity and evolution of RGAs in coconut, we have obtained a collection of expressed RGAs by mining the leaf transcriptome of the root (wilt) disease-resistant CGD cultivar, assessing their phylogenetic and evolutionary relationships, and characterizing their expression profile in root (wilt) disease-resistant and susceptible CGD genotypes.

NBS-LRR type RGAs constitute the largest family of *R* genes, with the NBS region consisting of conserved domains like the P loop, RNBS-A, Kinase 2, Kinase 3A, and GLPL domains. These motifs have been known to play significant roles in signal transduction and induction of defense responses through activation of kinases in response to pathogen attack (Hammond-Kosack and Jones, 1997). The alignment of deduced amino acid sequences of coconut NBS-LRR type RGAs revealed the presence of the 5 consensus sequences (P loop, RNBS-A, Kinase 2, Kinase 3A, and GLPL domains), which suggest their possible function as disease resistance protein genes (McHale et al., 2006). NBS-LRR sequences are abundant in plant genomes and our results reveal that a large number of NBS-LRR type sequences are also present in coconut.

Table 2. Gene specific primers used to amplify NBS-LRR type RGAs (RGA1–RGA7) and the endogenous control, alpha-tubulin (TUB), in coconut in RT-PCR and qRT-PCR experiments.

Sl. no.	Coconut unigenes	Primer name	Sequence (5'-3')	Product size (bp)
1.	Unigene100000387	RGA1F	CGGGTACATGATCTTATCCATG	139
		RGA1R	GAACATCTTTACCGGTGTTATG	
2.	Unigene100001651	RGA 3F	CAAAGAGGAACTCATTGGGTG	151
		RGA3R	CCAAGCATGACAACTAAACC	
3.	Unigene100017514	RGA4F	GAGGTTGACGACGTGCTGGAT	165
		RGA4R	AGATTTCCCACAAGTTTCCTC	
4.	Unigene100019682	RGA5F	GGTGAAAACATTCTGGAGGA	152
		RGA5R	CTAAGATCATCTTCCAGTAAC	
5.	Unigene100021995	RGA6F	GATTTCGGTAGTTGGCATGG	127
		RGA6R	AATTGTGAACAACTGCGTCAC	
6.	Unigene100111200	RGA7F	CAATTCGACACTTGACTCTGA	135
		RGA7R	CATGCCATCAAGAAAACTAT	
7.	-	TUBF	CTGGTGTCCTACTGGCTTC	123
		TUBR	GACCATGATTACGCCAAG	

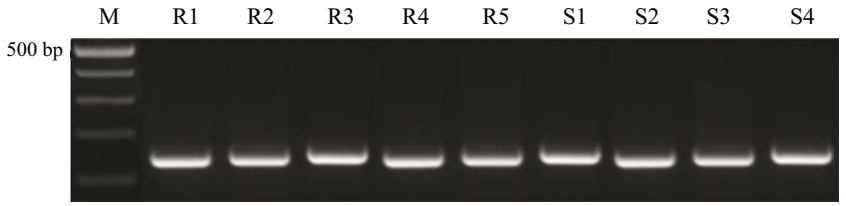


Figure 7. RT-PCR expression pattern of endogenous control alpha-tubulin (123 bp) in coconut root (wilt) disease-resistant (R1–R5) and susceptible (S1–S4) CGD genotypes. M: 100-bp DNA marker.

Plant NBS-LRR type proteins are classified into 2 subclasses, defined by the presence of TIR or CC motifs in the N-terminal domain. Although both are involved in pathogen recognition, these 2 subfamilies are distinct in sequence composition and the signaling pathways in which they are involved (McHale et al., 2006). The presence of aspartate (D) and tryptophan (W) residue at the end of the Kinase-2 motif is used to differentiate TIR and non-TIR classes of NBS-LRR type RGAs (Meyers et al., 2003). Interestingly, one of the RGAs examined in this work had an aspartate (D) residue at the end of the Kinase 2 domain, classifying it as a TIR NBS-LRR class RGA. The presence of the TIR domain in *R* genes could be predicted by the presence of the motif RNBS-A TIR (FDxxxD) near the P-loop (Cannon et al., 2002). TIR NBS-LRR type RGAs have been rarely reported in monocots and it was suggested that although TIR-type NBS-LRR sequences were present in early land plants, their numbers diminished significantly in monocots during evolution (Tarr and Alexander, 2009).

Clear grouping of the different NBS-LRR type RGAs into their respective classes could be visualized through cluster analysis and this suggests the existence of substantial sequence diversity and possible functional variations among these classes in coconut with high evolutionary significance. The evolution of *R* genes is facilitated by cluster formation, thereby causing recombination and sequence exchange, and this subsequently results in haplotypic diversity (Joshi et al., 2011).

A majority of RGAs identified in the present study belonged to the protein kinase class, encoding serine/threonine kinases. The serine/threonine kinases are known to interact with other proteins and affect a wide range of processes, especially signaling during pathogen recognition and subsequent activation of plant defense machinery (Romeis, 2001; Afzal et al., 2008).

Since a high mutation rate exists in many of the plant pathogens, plants should be able to generate and maintain useful levels of diversity at the resistance loci in order to subsist over evolutionary time (Chen et al., 2007). There has been much speculation regarding the origin of genetic diversity within *R* genes, which are arrayed in complex clusters in plants. Genetic mechanisms including unequal crossing-over, recombination, point mutations, and gene conversions have been proposed for generation of genetic diversity in *R* genes (Dixon, 2000). Characterization of nucleotide substitution patterns offers insight into the evolution of *R* gene families. The ratio of nonsynonymous to synonymous changes (Ka/Ks) is an indicator of the evolutionary pressures acting on a class of genes. Estimation of Ka/Ks ratios among coconut RGAs revealed that the ratios were significantly less than unity, indicating the presence of purifying selection rather than divergent selection. These results are consistent with earlier studies (Meyers et al., 1999; Noir et al., 2001; Xu et al., 2005; Chen et al., 2007; Nair and Thomas, 2007). These results obtained in coconut indicate that *R* genes might not be evolving in a rapid manner in order to keep pace with mutations in plant pathogens, but rather fairly slowly for imparting resistance against pathogen populations that are heterogeneous in time and space (Michelmore and Meyers, 1998; Stahl et al., 1999).

Constitutive expressions of NBS-LRR type *R* genes at low levels in plants have been reported under uninduced asymptomatic conditions. These *R* genes are known to be induced only after infection by pathogens (Nair and Thomas, 2007; Peraza-Echeverria, 2008). In our study, we have found significantly higher differential expression patterns of NBS-LRR type RGAs in the root (wilt) disease-resistant leaf samples compared to susceptible, symptomatic leaf samples for the 6 RGAs. Induction of *R* genes in plants following a pathogen attack has been observed for many *R* genes in rice (Yoshimura et al., 1998; Wang et al., 1999), *Arabidopsis* (Xiao et al., 2001), and ginger (Nair and Thomas, 2013). The *R* genes act as molecular receptors and are constitutively expressed and induced as soon as a pathogen is perceived, which subsequently triggers innate basal defense responses (Ellis and Jones, 1998; Richter and Ronald, 2000; Bent and Mackey, 2007).

The present investigation identified *R* genes in leaf transcriptome of the root (wilt) disease-resistant coconut CGD cultivar, making this study the first attempt at the targeted isolation of RGAs in coconut. The identified RGAs were classified into different classes based on the conserved motifs in the *R* genes. Detailed studies are further required

Figure 8. Calibrated and normalized relative gene expression (CNRQ) levels of 6 NBS-LRR type RGAs (RGA1–RGA7) in coconut in root (wilt) disease resistant (R1–R5) and susceptible (S1–S4) CGD genotypes, along with the endogenous control, alpha-tubulin (α-TUB).

to decipher the involvement of the upregulated RGAs, identified using qRT-PCR, in imparting disease resistance to resistant genotypes. The identified RGAs can act as a valuable resource towards development of RGA-based molecular markers for genetic mapping for root (wilt) disease resistance in coconut. Marker-assisted selection using markers tightly linked to coconut root (wilt) disease resistance could be used to screen a large number of germplasm accessions for the presence of these genes, which will be our future focus area.

Acknowledgments
The authors are thankful to the Director of CPCRI, Kasaragod, India, for facilities and support extended and to the Indian Council for Agricultural Research (ICAR) and Kerala Biotechnology Commission (KBC) for funding this research work. The authors also thank the 3 anonymous reviewers for their constructive comments and suggestions to improve the manuscript.

References

Afzal AJ, Wood AJ, Lightfoot DA (2008). Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21: 507–517.

Alvarenga SM, Caixeta ET, Hufnagel B, Thiebaut F, Maciel-Zambolim E, Zambolim L, Sakiyama NS (2010). *In silico* identification of coffee genome expressed sequences potentially associated with resistance to diseases. Genet Mol Biol 33: 795–806.

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu HY, Roe B, Young ND, Cannon SB (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146: 5–21.

Andolfo G, Sanseverino W, Aversano R, Frusciante L, Ercolano MR (2014). Genome-wide identification and analysis of candidate genes for disease resistance in tomato. Mol Breeding 33: 227–233.

Anil K, Das SN, Podile AR (2014). Induced defense in plants: a short overview. P Natl Acad Sci USA 84: 669–679.

Bent AF, Mackey D (2007). Elicitors, effectors, and *R* genes: the new paradigm and a lifetime supply of questions. Annu Rev Phytopathol 45: 399–436.

Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002). Diversity, distribution and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. J Mol Evol 54: 548–562.

Chen G, Pan D, Zhou Y, Lin S, Ke X (2007). Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.). J Biosci 32: 713–721.

Dangl JL, Dietrich RA, Richberg MH (1996). Death don't have no mercy: cell death programs in plant-microbe interactions. Plant Cell 8: 1793–1807.

Dangl JL, Jones JD (2001). Plant pathogens and integrated defense responses to infection. Nature 411: 826–833.

DebMandal M, Mandal S (2011). Coconut (*Cocos nucifera* L.: Arecaceae) in health promotion and disease prevention. Asian Pac J Trop Med 4: 241–247.

Dilbirligi M, Gill KS (2003). Identification and analysis of expressed resistance gene sequences in wheat. Plant Mol Biol 53: 771–787.

Dixon MS, Golstein C, Thomas CM, van der Biezen EA, Jones JDG (2000). Genetic complexity of pathogen perception by plants: the example of *Rcr3*, a tomato gene required specifically by *Cf-2*. P Natl Acad Sci USA 97: 8807–8814.

Ellis J, Jones D (1998). Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant Biol 1: 288–293.

Flor HH (1971). Current status of the gene-for-gene concept. Annu Rev Phytopathol 9: 275–296.

Foale MA, Ashburner R (2004). The coconut palm. In: Chopra VL, Peter KV, editors. Handbook of Industrial Crops. New Delhi, India: Indian Council for Agricultural Research.

Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW(2012). Plant disease resistance genes: current status and future directions. Physiol Mol Plant P 78: 51–65.

Hammond-Kosack KE, Jones JDG (1997). Plant disease resistance genes. Annu Rev Plant Physiol Plant Mol Biol 48: 575–607.

Heath MC (2000). Non-host resistance and non-specific plant defenses. Curr Opin Plant Biol 3: 315–319.

Holt BF, Hubert DA, Dangl JL (2003). Resistance gene signaling in plants—complex similarities to animal innate immunity. Curr Opin Immunol 15: 20–25.

Hulbert SH, Webb CA, Smith SM, Sun Q (2001). Resistance gene complexes: evolution and utilization. Annu Rev Phytopathol 39: 285–312.

Jacob PM, Nair RV, Rawther TSS (1998). Varietal resistance. In: Nampoothiri KUK, Koshy PK, editors. Coconut Root (Wilt) Disease. Kasaragod, India: CPCRI, pp. 97–104.

Jones JDG, Dangl JL (2006). The plant immune system. Nature 444: 323–329.

Joshi RK, Kar B, Nayak S (2011). Survey and characterization of NBS-LRR (R) genes in *Curcuma longa* transcriptome. Bioinformation 6: 360–363.

Kanazin V, Marek LF, Shoemaker RC (1996). Resistance gene analogs are conserved and clustered in soybean. P Natl Acad Sci USA 93: 11746–11750.

Leister D, Ballvora A, Salamini F, Gebhardt C (1996). A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nat Genet 14: 421–429.

Li X, van Eck HJ, Rouppe van der Voort JNAM, Huigen DJ, Stam P, Jacobsen E (1998). Autotetraploids and genetic mapping using common AFLP markers: the R2 allele conferring resistance to *Phytophthora infestans* mapped on potato chromosome 4. Theor Appl Genet 96: 1121–1128.

Liu Z, Crampton M, Todd A, Kalavacharla V (2012). Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (*Phaseolus vulgaris* L.). BMC Plant Biol 12: 42.

McHale L, Tan X, Koehl P, Michelmore RW (2006). Plant NBS-LRR proteins: adaptable guards. Genome Biol 7: 212.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999). Plant disease resistance genes encode members of an ancient and diverse protein family with the nucleotide-binding superfamily. Plant J 20: 317–332.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15: 809–834.

Michelmore RW, Meyers BC (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res 8: 1113–1130.

Monsi B, Wisser RJ, Pennill L, Hulbert SH (2004). Full-genome analysis of resistance gene homologues in rice. Theor Appl Genet 109: 1434–1447.

Nair RA, Thomas G (2007). Isolation, characterization and expression studies of resistance gene candidates (RGCs) from *Zingiber* spp. Theor Appl Genet 116: 123–134.

Nair RA, Thomas G (2013). Molecular characterization of *ZzR1* resistance gene from *Zingiber zerumbet* with potential for imparting *Pythium aphanidermatum* resistance in ginger. Gene 516: 58–65.

Nair RV, Jacob PM, Thomas RJ, Sasikala M (2004). Development of varieties of coconut (*Cocos nucifera* L.) resistant/tolerant to root (wilt) disease. J Plantation Crops 32: 33–38.

Noir S, Combes MC, Anthony F, Lashermes P (2001). Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees. Mol Gen Genomics 265: 654–662.

Peraza-Echeverria S, Dale JL, Harding RM, Smith MK, Collet C (2008). Characterization of disease resistance gene candidates of the nucleotide binding site (NBS) type from banana and correlation of a transcriptional polymorphism with resistance to *Fusarium oxysporum* f.sp. *cubense* race 4. Mol Breed 22: 565–579.

Persley GJ (1990).The Coconut Palm: Prosperity or Poverty. Canberra, Australia: ACIAR.

Rajesh MK, Rachana KE, Fayas TP, Merin Babu, Kiran AG, Anitha Karun (2014). Selection and validation of reference genes for quantitative gene expression studies by real-time PCR in coconut. In: Muralidharan K, Rajesh MK, Muralikrishna KS, Jesmi Vijayan, Jayasekhar S, editors. Book of Abstracts of National Seminar on Sustainability of Coconut, Arecanut and Cocoa Farming-Technological Advances and Way Forward, 22-23 August 2014. Kasaragod, India: CPCRI, p. 34.

Ramjegathesh R, Karthikeyan G, Rajendran L, Johnson I, Raguchander T, Samiyappan R (2012). Root (wilt) disease of coconut palms in South Asia–an overview. Arch Phytopathol Pfl 45: 2485–2493.

Richter TE, Ronald PC (2000). The evolution of disease resistance genes. Plant Mol Biol 42: 195–204.

Romeis T (2001). Protein kinases in the plant defense response. Curr Opin Plant Biol 4: 407–414.

Rossi M, Auaujo PG, Paulet F, Garsmeur O, Dias VM, Chen H, Van Sluys MA, D' Hont A (2003). Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. Mol Genet Genomics 269: 406–419.

Sasikala M, Rajeev G, Prakash VR, Amith S (2010). Modified protocol of ELISA for rapid detection of coconut root wilt disease. J Plantation Crops 38: 16–19.

Sharma TR, Das A, Kumar SP, Lodha ML (2009). Resistance gene analogues as a tool for rapid identification and cloning of disease resistance genes in plants: a review. J Plant Biochem Biotechnol 18: 1–11.

Soriano JM, Joshi SG, van Kaauwen M, Noordijk Y, Groenwold R, Henken B, van de Weg WE, Schouten HJ (2009). Identification and mapping of the novel apple scab resistance gene *Vd3*. Tree Genet Genomes 5: 475–482.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400: 667–671.

Tarr DEK, Alexander HM (2009). TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2: 197.

van der Linden G, Wouters D, Mihalka V, Kochieva E, Smulders M, Vosman B(2004). Efficient targeting of plant disease resistance loci using NBS profiling. Theor Appl Genet 109: 384–393.

Vossen JH, Dezhsetan S, Esselink D, Arens M, Sanz MJ, Verweij W, Verzaux E, van der Linden CG (2013). Novel applications of motif-directed profiling to identify disease resistance genes in plants. Plant Methods 9: 37.

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. Genomics Proteomics Bioinformatics 8: 77–80.

Wang ZX, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999). The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19: 55–64.

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG (2001). Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291: 118–120.

Xu Q, Wen XP, Deng XX (2005). Isolation of TIR NBS-LRR resistance gene analogues and identification of molecular marker linked to a powdery mildew resistance locus in chestnut rose. Theor Appl Genet 111: 819–830.

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang ZX, Kono I, Kurata N, Yano M, Iwata N, Sasaki T(1998). Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. P Natl Acad Sci USA 95: 1663–1668.

Zhu H, Cannon SB, Young ND, Cook DR (2002). Phylogeny and genomic organization of the TIR and non-TIR NBS-LRR resistance gene family in *Medicago truncatula*. Mol Plant Microbe Interact 15: 529–539.