African Journal of Biotechnology Vol. 11(85), pp. 15111-15123, 23 October, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.2373
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Isolation and characterization of resistant gene analogs in cassava, wild *Manihot* species, and castor bean (*Ricinus communis)*

Melaku Gedil[1]*, Manjula Kumar[2] and David Igwe[1]

[1]Bioscience Center, International Institute of Tropical Agriculture, Oyo road, P. M. B. 5320, Ibadan, Nigeria.
[2]Africa Rice Centre, Oyo road P. M. B. 5320, Ibadan, Nigeria.

Accepted 9 September, 2012

Cassava, *Manihot esculenta*, is one of the major food crops in sub-Saharan Africa (SSA) providing the bulk of dietary calories to hundreds of millions of households. Two viral diseases, namely cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) pose a serious threat to cassava production. The emergence of new virus species and strains that overcome the existing resistant/tolerant cultivars entails identification and pyramiding of new sources of resistance using marker-aided selection. The isolation of resistance gene analogues (RGAs) using a homology-based approach can provide useful resources towards this goal. Degenerate primers based on the conserved motif of the nucleotide binding site (NBS) domain from resistance (R) genes were used to isolate RGAs from genomic DNA and cDNA in cassava, wild *Manihot* species, and castor bean (*Ricinus communis*). A total of 552 RGAs sequences were identified and deposited in GenBank. Conserved motifs such as P-loop, Kinase-2a and GLPL were present in the NBS domain. This study sheds light on the nature of NBS- leucine-rich repeat (LRR) R genes in cassava and closely related taxa in the family Euphorbiaceae. These candidate sequences mapped to the draft cassava genome with high sequence similarity to predicted NBS-LRR genes. These novel sequences may serve as a stepping stone for further characterization and experimental validation of predicted R genes in the draft cassava genome, ultimately leading to the development of functional gene-targeted markers that can be used in molecular resistance breeding aimed at combating CBSD and CMD.

Key words: Cassava, resistance, *Manihot*, castor bean, resistance gene analog, nucleotide binding site.

INTRODUCTION

Cassava, *Manihot esculenta*, is one of the major food crops in sub-Saharan Africa (SSA) providing the bulk of dietary calories to hundreds of millions of households. In the highest producing countries of West Africa, the potential of cassava as an industrial crop and its role as a rural income generator is steadily rising. One of the favourable attributes of cassava is its ability to produce reasonable yield under marginal growth conditions including low soil fertility and moisture stress. Two viral diseases, namely cassava mosaic disease (CMD), which is prevalent across all cassava growing regions of SSA, and cassava brown streak disease (CBSD), pandemic to East and Central Africa pose serious threats to cassava production (Hillocks and Jennings, 2003; Legg and Fauquet, 2004; Patil and Fauquet, 2009). Yield loss due to these diseases is exacerbated by the prolonged cropping cycle and clonal propagation which inevitably leads to build up of inoculums.

The battle against CMD (Dixon et al., 2007) and CBSD (Hillocks and Jennings, 2003), has been aimed at breeding for durable host resistance using sources in land races as well as introgression of genes from wild relatives. Breeding cassava for resistance to cassava mosaic virus began several decades ago and has yielded many improved clones such as TMS58308, one of the earliest known resistant clones derived from an interspecific cross between *M. esculenta* and *Manihot glaziovii* (Jennings, 1994). Subsequently, TMS58308 has

*Corresponding author. E-mail: m.gedil@cgiar.org.

been used as donor for several other CMD resistant improved TMS series clones including the popular TMS30572. A number of early studies have confirmed that the nature of resistance conferred by this gene, designated CMD1, is polygenic and recessive (reviewed in Hahn et al., 1989). The collection and evaluation of several African land races unveiled additional sources of durable resistance which were successfully utilized to develop and deploy several new CMD resistant lines in major cassava growing countries such as Nigeria (Dixon et al., 2003; Nassar and Ortiz, 2007). Akano et al. (2002) proposed qualitative inheritance of a new source of resistance from these landraces and named the gene CMD2. Subsequently, markers associated to the trait were used in marker-assisted introgression of the gene to other genotypes (Okogbenin et al., 2007). A study designed to understand the inheritance of the gene and its complementarity with the introgressed *M. glaziovii* gene revealed that the gene in the land races exhibit polygenic inheritance and the genes are non allelic and not linked (Lokko et al., 2005). The paucity of knowledge on the genetics of resistance is further compounded by complex infection pattern with multiple viruses associated in various regions, mixed infection of African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV) as well as satellite-like DNAs speculated to have a role in symptom enhancement and resistance breakdown (Nawaz-Ul-Rehman and Fauquet, 2009; Patil and Fauquet, 2010).

Leveraging its holding of a large number of African landraces, wild species, as well as germplasm introductions from Latin America, International Institute of Tropical Agriculture (IITA) endeavored to tackle the devastating effect of CMV on cassava production. The African germplasm has been used to intercross and select CMD resistant genotypes that have been deployed widely in cassava growing regions of Africa (Dixon et al., 2007, 2003; Hahn et al., 1989). Several CBSD tolerant clones have also been identified and utilized in IITA's and National Agricultural Research Systems (NAR)'s cassava breeding program for East and Southern Africa. While the bulk of the achievement in the past could be attributed to the accelerated conventional breeding scheme implemented in recent years, considerable emphasis has also been placed on the development of molecular tools to aid in shortening the variety development process and reducing population sizes.

In the past decade and half, a number of mapping populations were developed and field evaluated in different environments. A growing number of molecular markers are becoming available to the cassava research community. The increasing number of available expressed sequence tags (ESTs) and the completion of the cassava genome sequencing is poised to spur the discovery and utilization of single nucleotide polymer phism (SNP) markers that are suitable for high throughput genotyping platforms. The use of SNP markers is becoming more and more relevant for cassava as the amounts of publicly available genomic and cDNA sequences continue to grow at an accelerated pace (Ferguson et al., 2011, 2012; Kawuki et al., 2009; Prochnik et al., 2012; Rabbi et al., 2012).

Many genes conferring resistance to different pathogens have been isolated and cloned in plants (Chisholm et al., 2006; Stange, 2006). Sequence comparisons among disease resistance (R) genes from different plants have revealed remarkable similarities in their general structure and in the conservation of specific domains that participate in defense signaling pathways (Hammond-Kosack and Parker, 2003; van Ooijen et al., 2007). Most of the disease R genes in plants encode proteins featuring nucleotide binding site (NBS) and leucine-rich repeats (LRR). As a result, two major subfamilies of the NBS-LRR type R genes were described based on the type of the amino-terminal domains. The NBS-LRR proteins carrying Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) motifs are known as TIR-NBS-LRR and CC-NBS-LRR, respectively (McHale et al., 2006).

The presence of conserved region of R genes provide opportunities for designing degenerate primers and isolating resistance gene analogues (RGAs) by the polymerase chain reaction (PCR) from plant genomes. PCR with degenerate primers led to identification of thousands of partial sequences of NBS-LRR genes in wide array of plant genomes (Azhar et al., 2011; Chen et al., 2007; Tarr and Alexander, 2009). These RGAs often map close to major R genes or quantitative trait loci (QTL) (Gebhardt et al., 2006), and hence may provide either interesting candidate genes or useful markers for marker-assisted selection (MAS). A previous study in cassava, *M. esculenta*, resulted in the isolation of 12 classes of resistance gene candidates (RGCs) of which two full-length protein coding sequences were identified and mapped on the framework cassava linkage map (Lopez et al., 2003). The candidate partial genes identified by this approach also provide information about the organization and evolution of R genes and RGAs in plant genomes (Ballvora et al., 2007; Meyers et al., 2005; Pan et al., 2000). RGAs have been used to develop and map molecular markers co-segregating with disease resistance traits (Calenge et al., 2005; Moroldo et al., 2008; Qiu et al., 2007), clone and characterize full length NBS-LRR gene (Okuyama et al., 2011; Seo et al., 2006). The availability of plant genome sequences prompted the identification and characterization of the complete set of NBS-LRR-encoding R genes in several plant species. A combination of computational and experimental analysis revealed the presence of about 150 NBS-LRR-encoding genes in *Arabidopsis* (Meyers et al., 2003; Tan et al., 2007), about 500 in rice (Zhou et al., 2004), and several hundred in grapevine (Moroldo et al., 2008), and a smaller number in papaya (Porter et al., 2009) and at least 333 in *Medicago truncatula* (Ameline-Torregrosa et

al., 2008). In general, understanding the structure, localization, function, variation, and evolution of R genes will provide the basis for devising an efficient breeding strategy for disease resistance (Hammond-Kosack and Parker, 2003; Lawson et al., 2010; Stange, 2006).

Cassava improvement in Africa exclusively depended on intercrossing and clonal selection of the existing germplasm (Ceballos et al., 2004; Dixon et al., 2003; Nassar and Ortiz, 2007). The first CMD resistance gene was introgressed from *M. glaziovii* into cassava in East Africa (Hahn et al., 1989). This study will provide informative insight into the nature of RGAs in hitherto unexplored, at genomics level, wild *Manihot* species. With respect to integrating molecular marker technology into the breeding scheme, an effort of screening genomic and EST-derived simple sequence repeat (SSR) markers that are associated with R genes has not yielded robust marker for molecular breeding of disease resistant cultivars. This study was, therefore, carried out as an alternative, comparative approach, to isolate and characterize NBS-LRR type R gene-like sequences in cassava and wild relatives in the Euphorbiaceae family. Knowledge obtained will contribute to the ongoing effort in marker development ultimately leading to the development of novel strategies for anticipatory and durable disease control.

MATERIALS AND METHODS

Plant materials and DNA isolation

Plant materials used for DNA isolation were obtained from the cassava breeding program, IITA, Nigeria. These include seven African land races of cassava (TME52, TME56 and TME117 are susceptible to CMD whereas four - TME3, TME6, TME7 and TME279 are resistant to CMD), four wild *Manihot* species, namely, *Manihot brachyandra, Manihot epruinosa*, *M. glaziovii*, *Manihot tripartita* and castor bean (*Ricinus communis*). Genomic DNA was isolated from young leaflets using the procedure optimized for cassava and other crops in IITA (Gedil et al., 2009). Final precipitation and purification of the DNA was performed as previously mentioned and resuspended in ultrapure water (Gibco BRL, Gaithersburg, MD).

cDNA-RGA amplification

Total RNA was isolated from 0.1 g frozen tissue using RNeasy plant mini kit (Qiagen, Valencia, CA) according to the manufacturer's instruction. Following the elimination of DNA, the template RNA was reverse transcribed. A 'minus' reverse transcriptase PCR reaction, where no reverse transcriptase enzyme was added during the cDNA synthesis reaction, was used to test each mRNA sample for genomic DNA contamination. Amplification of cDNA fragments was performed in 25 µl PCR reactions. Each reaction mixture contained 2 µl of first strand cDNA, 2.5 µl of 10x PCR buffer, 1 µl 50 mM $MgCl_2$, 2.5 mM deoxynucleotide triphosphates (dNTPs) mix, 5 mM each of forward and reverse primer, and 0.2 µl of Taq polymerase. Cycling condition was 35 cycles of 15 s at 94°C, 1.0 min at 42°C and 2.0 min at 72°C. A final extension step was performed at 72°C for 10 min.

Primers and PCR amplification

Two pairs of degenerate primers were developed in this study. The first pair was targeting various domains of the *Arabidopsis thaliana* NBS-LRR gene alignment (Meyers et al., 2003). Three forward primers were designed on the P-loop/Kinase1A domain whereas seven reverse primers were placed on the LRR domain. All 21 combinations were tested for patterns of amplifications. Primer pair NL-50-F (GGN GGN STN GGN AAR CAN CAN CTN) and TNL-470-RL (CAT GCA TGY GAD ATN AGN GAY TT) were selected for further assay. The second pair of degenerate primers RGA-F (GGI GGI GTI GGI AAI ACI AC); RGA-R (IAG IGC IAG IGG IAG ICC) were used to target the P-loop/Kinase1A and the GLPL region, respectively (Leister et al., 1996).

Out of the 21 combinations of primer pairs, NL-50-F and TNL-470-RL were used to amplify RGA-like sequences in cassava land races, TME3 and TME7. PCR was performed in volumes of 25 µl solution containing 2.5 µl of 10x PCR buffer, 1.5 µl of 50 mM $MgCl_2$, 2.0 µl, of 2.5 mM dNTPs, 2.0 µl of 5 mM of each primer, 1.0 µl of dimethyl sulfoxide (DMSO), 50 ng of template DNA, and 0.2 -ul *Taq* DNA polymerase (Sigma Aldrich, St. Louis, Mo.). DNA was denatured at 94°C for 2 min and amplification was performed over 34 cycles of 15 s at 94°C, 1.0 min at 42°C and 2.0 min at 72°C. A final extension step was performed at 72°C for 10 min. PCR amplification products were separated by electrophoresis on 1.5% agarose gel and stained with ethidium bromide for visualization.

Cloning of PCR products and sequencing

The amplicon size ranging from 200 to 900 bp were cloned using Qiagen PCR cloning plus kit (Qiagen). For direct colony PCR, white colonies were suspended in 50 µl of ultrapure water, denatured at 99°C for 5 min. After centrifugation of cell debris at 12,000 r.p.m for 1 min, 10 µl of supernatant was used as template for PCR. Amplification was performed in volumes of 25 µl reaction using T7 and SP6 universal primers. Amplification was carried out in BioRad thermocycler with initial denaturation at 94°C for 2 min followed by 30 cycles of denaturation (15 s at 93°C), and annealing (45 s for 50°C) and extension 30 s at 72°C completed with a final extension step of 10 min at 72°C. Clones with insert were purified with ethanol precipitation and used as a template for sequencing.

Sequence analysis

Raw sequences were edited in CodonCode aligner v3.7.1 (CodonCode Corporation, Dedham, MA, USA) for base calling error and other artifacts. Vector and primer sequences were trimmed from the sequences. Different open source computer programs were used for multiple sequence alignment (MSA), identity matrix, similarity search, and phylogentic tree construction including BioEdit (Hall, 1999), MEGA (Tamura et al., 2007) and Jalview (Waterhouse et al., 2009). Similarity search based on BLAST algorithm (Altschul et al., 1997) were performed on local databases made up of cassava and castor bean nucleotide and protein sequences download from the GenBank http://www.ncbi.nlm.nih.gov/) and the Phytozome web site (http://www.phytozome.net/). Further sequence comparative analysis was performed against local database made up of plant disease R genes (Sanseverino et al., 2010). Nucleotide sequence analysis and translation to the corresponding amino acid sequence was performed using CodonCode Aligner or web-based translation tools. Distances for phylogenetic analyses were performed using the MEGA software (version 4.0) based on MSA made with the program MUSCLE (Edgar, 2004). Phylogenetic trees were constructed by the Neighbor–joining method (Saitou and Nei, 1987).

Table 1. Number of RGAs isolated from genomic DNA and cDNA in various cultivars of cassava, wild *Manihot*, and castor bean.

Variable		Genomic		cDNA		All clones		NBS-LRR
		Nt*	AA**	Nt*	AA**	Nt*	AA**	like[£]
Cassava	TME117	16	10	41	32	57	42	35
	TME3	1	1	13	10	14	11	7
	TME7	1	1	-	-	1	1	1
	TME6	4	2	-	-	4	2	2
	TME279	18	11	-	-	18	11	9
	TME52	16	9	-	-	16	9	4
	TME56	27	10	-	-	27	10	8
	Subtotal	83	44	54	42	137	86	66
Wild species	*M .brachyandra*	12	9	52	39	64	48	35
	M .epruinosa	99	74	1	1	100	75	35
	M .glaziovii	38	11	18	17	56	28	19
	M .tripartita	76	40	33	22	109	62	27
	Subtotal	225	134	104	79	329	213	116
Castor bean	Castor bean-1	31	8	-	-	31	8	6
	Castor bean-2	0	0	46	37	46	37	14
	Castor bean-3	0	0	9	9	9	9	4
	Subtotal	31	8	55	46	86	54	24
	Grand Total	339	186	213	167	552	353	206

*Number of nucleotide sequence before translation. **sequences with ORFs. [£]Number of sequences (RGAs) showing significant similarity to known R genes.

Bootstrapping was used to provide an estimate of confidence for each branch point and the final tree.

Local BLAST searching was performed within BioEdit using appropriate BLAST program (BLASTp or BLASTn) depending on the query and subject sequence type. Local database was made for cassava (Cassava Genome Project, 2010; http://www.phytozome.net/cassava) and castor bean (Castor bean Genome Project, http://www.phytozome.net/ricinus) nucleotide and protein sequences. Likewise, nucleotide and protein sequences of plant disease resistance proteins and FASTA files of nucleotide sequences of RGAs were downloaded from PRGdb (Sanseverino et al., 2010) (http://www.prgdb.org).

RESULTS

Identification of NBS containing sequences

Genomic DNA and cDNA from seven land races of cassava and four wild *Manihot* species (*M. brachyandra, M. epruinosa, M. glaziovii* and *M. tripartita*), and castor bean (*R. communis*) were subjected to the identification and characterization of RGAs. Three of the seven cassava land races (TME52, TME56 and TME117) are susceptible to CMD whereas four (TME3, TME6, TME7 and TME279) were resistant to CMD. In the first experiment, degenerate primers NL-50-F and TNL-470-RL were used to amplify RGAs in TME3 and TME7 using genomic DNA as a template. These primers were designed from aligned multiple sequences known to condition resistance in *Arabidopsis* (Meyers et al., 2003). Most focus was made on *Arabidopsis* sequences due to availability of extensive genome annotation and characterization at the time of the study. Several pairs of degenerate primers were tested (data not shown) after which the above pair was selected for further analysis. The forward and reverse primers of the second pair of degenerate primer (Leister et al., 1996) RGA-F and RGA-R were designed to target the P-loop and the GLPL region, respectively. A total of 795 genomic clones and 549 cDNA clones, altogether 1344 clones, were sequenced. A total of 339 genomic and 213 cDNA unique sequences were identified (Table 1). The filtering criteria included removal of the sequences that are shorter than 200 base pairs in compliance with GenBank submission requirement. More than half (353 sequences) of the total sequences (both genomic and cDNA) had open reading frame (ORF) uninterrupted with stop codon (Table 1). The 552 sequences were submitted to GenBank and assigned accession numbers as shown in Table 2.

Multiple sequence alignment

The 352 deduced amino acid sequences were subjected to MSA by using the program MUSCLE (Edgar, 2004).

Table 2. GenBank accession number of isolated cassava, wild *Manihot* and castor bean RGAs.

Species	Genbank accession number	Number
M. esculenta	JF790354–JF790490	137
M. epruinosa	JF790491–JF790498 JF790500–JF790504 JF790506–JF790543 JF790545–JF790546 JF790548–JF790550 JF790552–JF790558 JF790560–JF790596	100
M. tripartita	JF790597–JF790618 JF790620–JF790630 JF790632–JF790652 JF790654–JF790708	109
M. brachyandra	JF790709–JF790754 JF790756–JF790759 JF790761 JF790763–JF790775	64
R. communis	JF790776–JF790861	86
M. glaziovii	JF790862–JF790917	56
Total		552

Table 3. Species composition of selected RGA groups.

Group	Mb	Me	Mep	Mg	Mtr	Rc	Total
G1	30	35	21	13	22	11	132
G4	3	2	11	0	13	0	29
G5	4	3	3	1	1	5	17
G6	0	0	0	0	8	0	8
G7	0	0	0	0	0	9	9
G14	3	18	11	0	2	5	39

Mb, *Manihot brachyandra*; Me, *Manihot esculenta*; Mep, *Manihot epruinosa*; Mg, *Manihot glaziovii*; Rc, *Ricinus communis*.

MSA resulted in 19 groups with 3 to 132 sequences and 13 groups each with two sequences. Thirty-four (34) sequences remained ungrouped. Within group difference was not more than a few nucleotide differences. The largest group (G1, Table 3) contained 132 members derived from all of the studied species suggesting the high level of conservation of the NBS domain across the genus *Manihot* and *Ricinus*. The second largest group was G14 comprising 39 sequences derived from 5 of the 6 studied species. Most of the large groups contained sequences from multiple species (Table 3).

Similarity search

Similarity search was performed on different databases by using representative sequence of the 32 groups as query. Local BLAST was made on nucleotide and protein sequences of plant R genes downloaded from PRGdb (Sanseverino et al., 2010). As of July, 2010, the PRGdb (http://www.prgdb.cbm.fvg.it) had a collection of 93 well characterized plant disease R genes (updated version of plant R protein sequences were kindly provided by Walter Sanseverino, Italy). These were used to create local protein database in BioEdit. Nine (9) out of the 32 groups showed significant similarity to known R genes or partial sequences or RGAs. Likewise, the 34 ungrouped (singletons) were subjected to similarity search of which 9 were found to be NBS-LRR like sequences. Similarity search result of the 18 NBS containing classes of cassava RGAs, obtained from three different databases is shown in Table 4. These 18 classes, representing 206

Table 4. BLAST result of isolated RGAs run against GenBank, reference genes, and RGAs (PRGdb)*.

Group	#seq	Best match	Matching R genes	Matching RGA sequence
G01	132	Gb:XP_002531928	Gro1.4, BS4	49175_*Citrus grandis* × *Poncirus trifoliata* USDA17-47 putative citrus disease resistance protein Pt14 gene (Gb:AY130794)
G02	6	Gb:XP_002531928	Gro1.4, BS4	48776_*Malus x domestica* clone ABHA006074CT putative NBS-LRR disease resistance protein gene (Gb: DQ644292)
G03	4	Gb:XP_002531928	Gro1.4, BS4	49175_*Citrus grandis* × *Poncirus trifoliata* USDA17-47 putative citrus disease resistance protein Pt14 gene (Gb:AY130794)
G11	6	Gb:AAP30049	Rps5, Rps2	47182_*Malus* × *domestica* putative disease resistance gene analog NBS-LRR (HRGA-H40) gene (Gb:AF516624)
G14	39	Gb:AAO38216	Gro1.4, N	58721_*Populus trichocarpa* clone 806367 TIR-NBS-LRR type disease resistance protein mRNA (Gb:DQ513251)
G16	3	Gb:AAO38220	Rps2, Rps5	47583_*Arabidopsis thaliana* disease resistance protein homolog (pNd1) gene (Gb:ATU97217)
G17	3	Gb:AAP30044	Rps5, Rps2	47182_*Malus* × *domestica* putative disease resistance gene analog NBS-LRR (HRGA-H40) gene (Gb:AF516624)
G22	2	Gb:XP_002513539	Bs4, Gro1.4	58754_*Populus trichocarpa* clone 8261519 NBS type disease resistance protein mRNA (Gb:DQ513204)
G23	2	Gb:XP_002328224	I2, Rps1-k-1	47965_*Pyrus communis* clone Pc-Sta-3 nucleotide binding site leucine-rich repeat disease resistance protein gene (Gb:EU939805)
G33	1	Gb:ACV85792.1	Gro1.4, BS4	49599_Ipomoea batatas clone IBRGA-4 NBS-LRR protein gene (Gb:DQ099391)
G34	1	Gb:ABV30884.1	Gro1.4, BS4	50065_*Cucumis melo* clone BAC 13J4
G35	1	Gb:AAO38214.1	Gro1.4, L6	58753_*Populus trichocarpa* clone 1458657 TIR-NBS-LRR type disease resistance protein mRNA (Gb:DQ513205)
G36	1	Ref:XP_002522604.1	Gro1.4, BS4	58754_*Populus trichocarpa* clone 8261519 NBS type disease resistance protein mRNA (Gb:DQ513204)
G37	1	Ref:XP_002517883.1	Gro1.4, BS4	58753_*Populus trichocarpa* clone 1458657 TIR-NBS-LRR type disease resistance protein mRNA (Gb:DQ513205)
G38	1	Gb:AAO38216.1	Gro1.4, N	58721_*Populus trichocarpa* clone 806367 TIR-NBS-LRR type disease resistance protein mRNA (Gb:DQ513251)
G39	1	Gb:ABJ55951.1	RPM1, BS4	58827_*Prunus persica* putative NBS-LRR type disease resistance protein (RPM1) mRNA (Gb:AY59923)
G40	1	Gb:AEB61537.1	RPM1, rpiblb2	58827_*Prunus persica* putative NBS-LRR type disease resistance protein (RPM1) mRNA (Gb:AY59923)
G41	1	Embl:CBI22403.3	Dm3, Rps5	49164_*Citrus grandis* × *Poncirus trifoliata* USDA17-47 putative citrus disease resistance protein Pt19 gene (Gb:AY130805)

*Local BLAST was performed on sequences downloaded from plant resistance genes database (PRGdb) http://www.prgdb.cbm.fvg.it; (Sanseverino et al., 2010).

Table 5. BLASTp result of RGA sequences matching predicted cassava CDS (coding sequence, predicted)*.

RGA representing the group	Best match	Percent identity	Alignment Length (bp)	E-value	Bit score
G1-Me_JF790354	Cassava4.1_031978m	99	156	8e-086	311
G2-Mb_JF790735	Cassava4.1_031978m	99	135	3e-072	266
G3-Rc_JF790834	Cassava4.1_031978m	71	159	9e-061	228
G11-Me_JF790437	Cassava4.1_001427m	100	161	1e-089	325
G14-Me_JF790387	Cassava4.1_022519m	99	161	2e-088	320
G16-Me_JF790440	Cassava4.1_000804m	48	147	1e-032	135
G17-Me_JF790441	Cassava4.1_000804m	100	160	1e-088	321
G22-Mg_JF790880	Cassava4.1_031334m	55	95	3e-024	106
G23-Me_JF790446	Cassava4.1_032094m	99	169	8e-096	345
G33-Me_JF790452	Cassava4.1_000585m	99	157	1e-087	318
G34-Mg_JF790882	Cassava4.1_025042m	99	92	4e-046	179
G35-Mg_JF790883	Cassava4.1_000627m	100	74	4e-036	145
G36-Mg_JF790884	Cassava4.1_030208m	95	92	4e-046	179
G37-Rc_JF790837	Cassava4.1_025042m	64	164	1e-055	211
G38-Rc_JF790821	Cassava4.1_022519m	85	136	4e-060	226
G39-Me_JF790453	Cassava4.1_025665m	98	104	5e-057	215
G40-Mep_JF790573	Cassava4.1_026581m	95	168	1e-091	331
G41-Mtr_JF790668	Cassava4.1_034119m	95	156	2e-080	294

*Cassava Genome Project, 2010; http://www.phytozome.net/cassava.

putative RGA sequences, were used for further analysis. Out of the 32 groups, 10 groups showed very low alignment to R genes, 13 groups did not show significant hits in the reference R gene database and hence discarded.

Similarity search was performed on all non-redundant GenBank CDS by using the BLASTp algorithm at a minimum E-value of 0.0001. As expected, the top matches were the previously published cassava resistant gene candidates (Lopez et al., 2003) and predicted cassava and castor bean genes, in most cases with 98 to 100% identity over the aligned segments. The top matching predicted transcript of cassava (Table 5) and castor bean (Table 6) were identified by local BLAST on databases made up of sequences downloaded from www.phytozome.net/cassava and www.phytozome.-net/ricinus, respectively. All but four groups of RGA showed over 80% identity to matching sequences in the cassava draft genome sequence. However, in the case of castor bean, percentage (%) identity was slightly lower than that for cassava, even though some groups showed over 90% sequence identity.

Motifs and TIR-NBS-LRR (TNL) versus CC-NBS-LRR (CNL)

The NBS-LRR proteins carrying TIR or CC motifs are known as TNL and CNL, respectively. Conserved motifs of the NBS domain including P-loop, Kinase-2 and GLPL were present in all NBS-LRR-like sequences (Figure 1). The last residue of kinase-2 (position 100 in Figure 1) tryptophan (W) was present in seven groups suggesting that the protein is non-TIR (Meyers et al., 1999; Pan et al., 2000). The non-TIR proteins were further divided into two. G11, G16, and G17 were closely related to RPS2/ RPS5/Dm3 whereas, G23, G39, G40 and G41 were closely related to RPM1 and I2 proteins. On the other hand, the remaining 11 groups were found to be TIR-NBS-LRR type of protein due to the absence of W at position 100. While three groups (G14, G33 and G38) had aspartic acid (D), the remaining eight had asparagine (N). The TIR groups were closely related to known TIR-NBS-LRR proteins (Gro1.4/BS4/N) all of which have D at the same position.

Phylogenetic analysis

To elucidate the relationship between the sequences, a representative of each of the groups and along with published RGC and the corresponding region of known R genes were used to construct phylogenetic trees (Figure 2). The method of Neighbour-Joining with 1000 bootstrap values was used. Based on their similarity to plant R genes, these unique sequences can be subdivided into three clusters as shown in Figure 2. Groups 1, 2, 3, 14, 22, 33 to 38 showed similarity to R genes Gro1.4 and BS4 whereas group 11, 16, 17 and 41 showed similarity to RPS2 and RPS5. Gro1.4 is a TNL class *Solanum tuberosum* nematode resistance protein (Gro1-4) and

Table 6. RGA sequences matching predicted castor bean CDS (coding sequence, predicted)*.

RGA representing the group	Best match	Percent identity	Alignment length (bp)	E-value	Bit score
G1-Me_JF790354	29579.m000196	71	155	2e-059	224
G2-Mb_JF790735	29579.m000196	73	135	5e-052	199
G3-Rc_JF790834	29579.m000196	99	158	6e-086	312
G11-Me_JF790437	29841.m002832	73	161	6e-065	242
G14-Me_JF790387	29747.m001050	63	161	2e-052	201
G16-Me_JF790440	29912.m005378	76	154	4e-064	239
G17-Me_JF790441	29841.m002832	60	161	1e-052	201
G22-Mg_JF790880	29910.m000957	59	94	8e-025	108
G23-Me_JF790446	30131.m007215	58	170	8e-052	199
G33-Me_JF790452	29841.m002919	75	160	3e-065	243
G34-Mg_JF790882	28525.m000274	58	92	4e-022	99
G35-Mg_JF790883	28525.m000274	74	74	5e-027	115
G36-Mg_JF790884	29838.m001641	78	92	1e-037	150
G37-Rc_JF790837	29983.m003159	100	123	8e-068	246
G37-Rc_JF790837	29983.m003159	67	18	8e-068	28
G38-Rc_JF790821	29747.m001050	63	136	4e-041	163
G39-Me_JF790453	29757.m000737	48	104	4e-025	109
G40-Mep_JF790573	29757.m000737	50	168	9e-043	169
G41-Mtr_JF790668	30074.m001381	46	156	3e-030	127

*Castor bean Genome Project 2010, http://www.phytozome.net/ricinus Additional files.

BS4 is a TNL class *Lycopersicon esculentum* bacterial spot disease resistance protein 4 (Bs4) gene. RPS2 and RPS5 are a CNL class *A. thaliana* R gene for *Pseudomonas syringae* 2 and 5, respectively. Group 23 is similar to I2, a CNL class *L. esculentum* fusarium wilt resistance protein. While G41 is clustered with G11, G16 and G17, it is more similar to DM3 (RGC2B), a CNL class R gene conferring resistance to *Lactuca sativa* downy mildew disease. Rca11 and Rca12 sequences, TNL-like cassava sequences (Lopez et al., 2003), grouped in a distant branch on the tree.

DISCUSSION

This study reports the identification of RGAs in cultivated cassava and its wild relatives. There are very limited nuclear DNA sequences of the four wild *Manihot* species in public databases. The deposition of these sequences in GenBank will provide data for evolutionary studies in the family. Whereas, the discovery of RGAs was performed in genomic DNA and cDNA, there was no noticeable difference in the RGAs derived from the two sources. Of the 552 final sequences, 64% (353) of the RGAs had ORF. However, only 206 of the predicted proteins were NBS-LRR like sequences based on similarity search (Table 1). As previously reported, some of the sequences cloned might be pseudogenes. The pseudogene ratio was 10% in *Arabidopsis* (Meyers et al., 2003), 11% in soybean (Kanazin et al., 1996) and 9.3% in tomato (Pan et al., 2000).

Similarity search

Extensive similarity search in local and public databases showed that over 200 of these partial sequences are homologues to known plant disease resistance genes. Our local databases were made up of cassava CDSs downloaded from Phytozome v7.0 (Cassava Genome Project 2010, http://www.phytozome.net/cassava). At the time of the manuscript preparation, the latest assembly (cassava4) of the database had a record of 34,151 transcripts/peptides. Even though the peptides are homology-based prediction, most of the genes were supported with EST based evidence. Using the RGA sequences as query, we could identify 172 putative NBS-LRR containing cassava transcripts (loci) at the cut-off of 1e-10 (data not shown).

For instance, the largest group (G1) matches the locus ‘cassava4.1_031978m.g’ at 99% protein sequence identity over an alignment length of 156 amino acids (Table 4). Likewise, similarity search against castor bean genome (www.phtozome.net/ricinus.php) at the same e-value retrieved 117 putative R genes. However, these numbers might change as the assembly and annotation of the two genomes is improved with the availability of more sequence data.

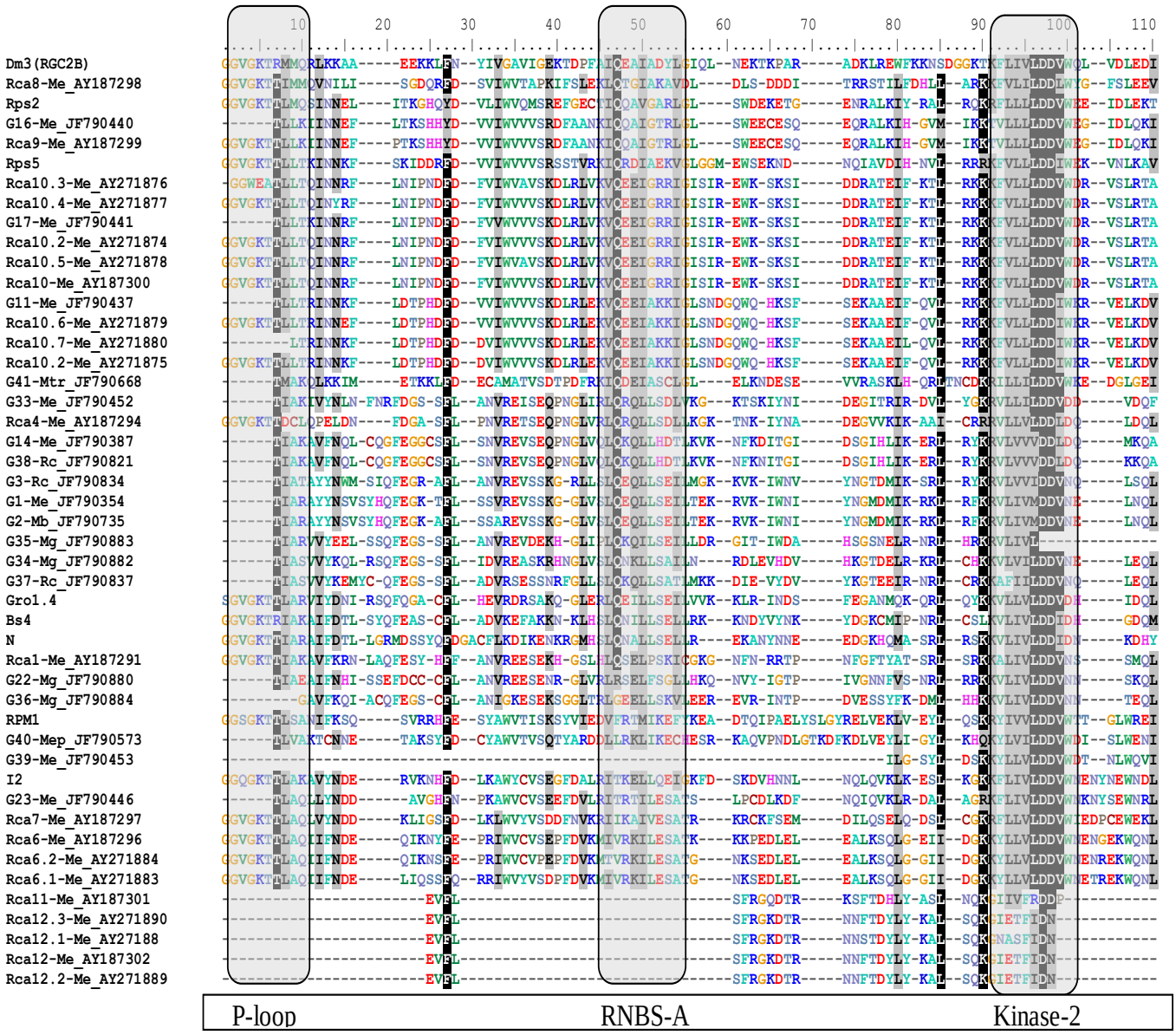

Figure 1. Multiple alignment of deduced protein sequences spanning the P-loop and GLPL motif of NBS domain in 18 novel RGA-like sequences, published cassava RGC sequences, and known R genes. Approximate position of conserved motifs is shown in rectangular boxes. Amino acid residue at position 100 (the end of Kinase-2 motif) is W (tryptophan) in non-TIR sequences and D (aspartic acid) or N (asparagine) in TIR sequences as shown in Figure 2.

Grouping

In plants, NBS-LRR genes are classified depending on whether they code for a TIR domain (having homology to the intracellular domain of the Drosophila Toll and mammalian interleukin-1 receptors in their N-terminus, TIR) or the non-TIR group. The TIR group genes comprise an N-terminal TIR domain, a central NBS domain, and a C-terminal LRR region, commonly observed in dicot plant species (McHale et al., 2006; Tarr and Alexander, 2009). In order to get an insight into the classes of NBS-LRR, we have included known R genes that were homologues to the RGAs identified in this study. Included also were previously reported cassava RGAs (Lopez et al., 2003). Multiple sequence alignment revealed 32 groups. However, only 9 groups and 9 singletons were found to be NBS-LRR R genes (Table 4). Based on BLAST search result, which shows the most similar homologues gene, the candidate NBS-LRR containing sequences could be grouped into three main clusters as shown in Figure 2. According to the tree, while the first and the second cluster comprised sequences closely related to non-TIR type genes, the third cluster was related to a TIR type R gene such as

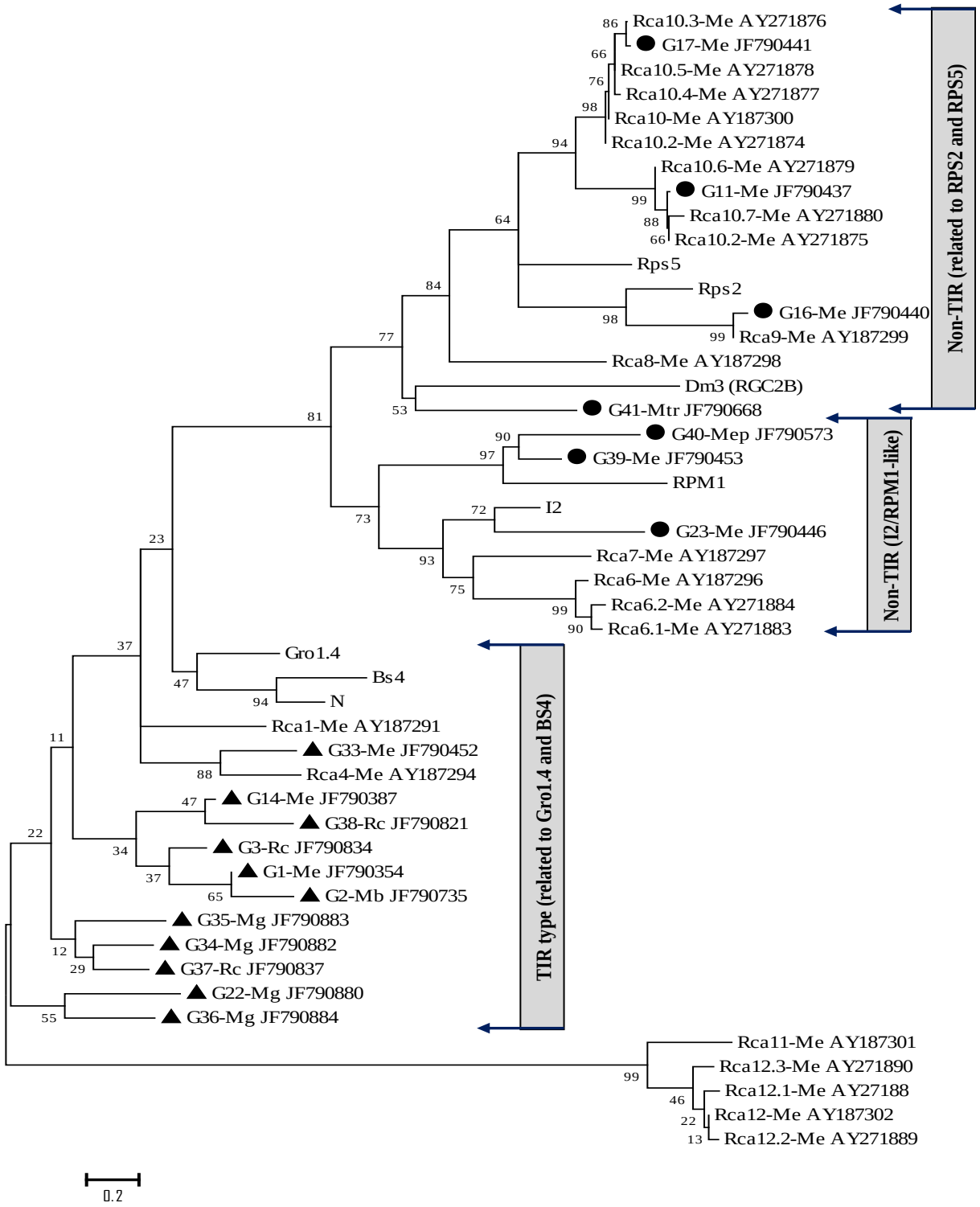

Figure 2. Phylogenetic relationship among 18 novel RGA-like sequences, published cassava RGC sequences, and the NBS domain of known R genes. Muscle alignment was subjected to tree construction based on Neighbor-Joining with 1000 bootstrap.

Gro1.4 and BS4. The amino acid residue that determines whether the gene is TIR or non-TIR (Meyers et al. 1999; Pan et al., 2000) is located at position 100 in the alignment (Figure 1). In the case of the first and the second clusters, position 100 was W consistent with previous findings suggesting that this sequences are non-

TIR NBS-LRR proteins. However, members of the third cluster contained D or N; whereas, literature indicates that the presence of D at the end of the Kinase-2 motif suggests TIR-NBS-LRR type proteins. The significance of having D or N at this position remains to be investigated on the full sequence of the NBS domain. The partial sequences in this study do not include the TIR domain and inference based on the single residue may be inconclusive. The reference genes in the non-TIR clusters (RPS2, RPS5, DM3, RPM1 and I-2) contained W at position 100 of this alignment (Figure 1). Similarly, the three reference genes in the TIR cluster (Gro1.4, BS4 and N) featured D at this position. The non-TIR sequences were divided into two sub-groups. Our observation corresponds to the study of Tarr and Alexander (2009) in which NBS sequences from multiple species exhibited a single TIR clade, while the non-TIR sequences were clustered into multiple clades in phylogenetic tree. Previous analysis of *Arabidopsis* and other monocots have implicated that monocot NBS-LRR genes are more homogenous in their domain architecture, and dicot genes can be divided into two distinct domains, the TIR coding at the N termini, and non-TIR genes (Wan et al., 2010).

Sequence identity and motifs

The observed high degree of protein sequence similarity (up to 99%) of the RGAs identified in this study and majority of the RGAs in previous studies may be the result of a tandem duplication of a common ancestral gene. The fact that duplicated RGA sequences in a given genome arise due to unequal crossing-over mechanism, generating diversity between genes has been established by extensive comparative sequence analysis and supported by large amount of genetic and molecular evidences in numerous species (Ellis et al., 2000; McDowell and Simon, 2006; Meyers et al., 2003). Mapping of these RGAs or the corresponding full-length NBS-LRR genes on the cassava genome will reveal whether NBS-LRR encoding genes are clustered in the genome as observed in other species (McHale et al., 2006). This study identified 11 TIR- and 7 non-TIR-NBS-LRR like groups, representing 189 and 17 RGAs, respectively (Figure 2 and Table 4). This is consistent with the findings in other dicots such as *A. thaliana* genome which contains 94 TIR and 55 non-TIR genes (Meyers et al., 2003).

cDNA versus genomic RGAs

There is no notable difference in RGAs derived from constitutively expressed cDNAs and those derived from genomic DNA with respect to motif structure within the NBS domain or the proportion of ORFs even though the cDNAs have a slightly higher proportion of ORFs. About 199 genomic and cDNA sequences contained multiple stop codons. Hence, these sequences were excluded from further analysis. The lack of ORFs in resistance gene like (RGL) cDNAs from different species have been reported (Budak et al., 2006).

Conclusions

Plants use different types of disease-R genes to detect the presence of pathogens and trigger signal transduction cascades culminating in defense response (Hammond-Kosack and Parker, 2003). The recent advance in genome technology has led to the identification of several classes of plant R genes (Meyers et al., 2005) the most common class being the NBS-LRR proteins which are known to activate a range of plant defense responses (McHale et al., 2006). An understanding of the genetic and genomic mechanism underlying host-pathogen interaction paves the way for efficient and effective crop improvement strategy aimed at resistance breeding. Characterization of NBS-LRR genes has generated profound knowledge about disease R genes with respect to the types and mechanisms of evolution, structure, localization, function, variation, and generation of novel recognition specificity (Hammond-Kosack and Parker, 2003; McDowell and Simon, 2006; McHale et al., 2006; Stange, 2006) in several plant species. Likewise, this study sheds light on the nature of NBS-LRR type R genes in cassava and closely related taxa in the family Euphorbiaceae and boosts the meagre GenBank entries for the four wild *Manihot* species thereby generating data for genomic and evolutionary studies in the family. Despite the partial coverage of the NBS domain, both TIR and non-TIR subfamilies of the NBS-LRR proteins were identified. To capture the various classes of NBS-LRR proteins, it is essential to use multiple pairs of primers targeting the different motifs that could exist in the genome. But the availability of genome sequence avails alternative approaches, computational and experimental, to perform genome-wide survey of R proteins including those that do not belong to the NBS-LRR classes (Meyers et al., 2003; Porter et al., 2009; Tan et al., 2007; Yang et al., 2006; Zhou et al., 2004). These candidate sequences mapped to the draft cassava genome with high sequence similarity to predicted NBS-LRR genes. The result in this study will serve as a stepping stone for further characterization and experimental validation of putative R genes, eventually leading to the development of functional gene-targeted markers that can be used in molecular resistance breeding aimed at combating cassava brown streak and cassava mosaic diseases.

ACKNOWLEDGEMENTS

This research was supported by the core fund provided by the International Institute of Tropical Agriculture (IITA).

REFERENCES

Akano O, Dixon O Mba C Barrera E, Fregene M (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor. Appl. Genet. 105:521-525.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, MillerW Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, 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.

Azhar MT, Amin I, Bashir A, Mansoor S (2011). Characterization of resistance gene analogs from *Gossypium arboreum* and their evolutionary relationships with homologs from tetraploid cottons. Euphytica 178:351-362.

Ballvora A, Jocker A, Viehover P, Ishihara H, Paal J, Meksem K, Bruggmann R, Schoof H, Weisshaar B, Gebhardt C (2007). Comparative sequence analysis of *Solanum* and *Arabidopsis* in a hot spot for pathogen resistance on potato chromosome V reveals a patchwork of conserved and rapidly evolving genome segments. BMC Genomics 8:112.

Budak H, Su S, Ergen N (2006). Revealing constitutively expressed resistance genes in *Agrostis* species using PCR-based motif-directed RNA fingerprinting. Genet. Res. 88:165-175.

Calenge F, van der Linden CG, Van de WE, Schouten HJ, Van AG, Denance C, Durel CE (2005). Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. Theor. Appl. Genet. 110:660-668.

Ceballos H, Iglesias CA, Perez JC, Dixon AG (2004). Cassava breeding: opportunities and challenges. Plant Mol. Biol. 56:503-516.

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.

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006). Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803-814.

Dixon A, Akoroda M, Okechukwu R, Ogbe F, Ilona P, Sanni L, Ezedinma C, Lemchi J, Ssemakula G, Yomeni MO, Okoro E, Tarawali G (2007). Fast track participatory approach to release of elite cassava genotypes for various uses in Nigeria's cassava economy. Euphytica 160:1-13.

Dixon AGO, Bandyopadhyay R, Coyne D, Ferguson M, Shaun R, Ferris B, Hanna R, Hughes J, Ingelbrecht I, Legg J, Mahungu N, Manyong V, Mowbray D, Neuenschwander P, Whyte J, Hartmann P, Ortiz R (2003). Cassava: From poor farmer's crop to pacesetter of African rural development. Chronica Hortic. 43:8-15.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinforma. pp. 5-113.

Ellis J, Dodds P, Pryor T (2000). Structure, function and evolution of plant disease resistance genes 614. Curr. Opin. Plant Biol. 3:278-284.

Ferguson M, Rabbi I, Kim DJ, Gedil M, Lopez-Lavalle LAB, Okogbenin E (2011). Molecular Markers and Their Application to Cassava Breeding: Past, Present and Future. Trop. Plant Biol. 5:95-109.

Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, Town CD, de Young J, Marri PR, Rabbi IY, de Villiers EP (2012). Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor. Appl. Genet. 124:685-695.

Gebhardt C, Bellin D, Henselewski H, Lehmann W, Schwarzfischer J, Valkonen JP (2006). Marker-assisted combination of major genes for pathogen resistance in potato. Theor. Appl. Genet. 112:1458-1464.

Gedil M, Kolade F, Raji A, Ingelbrecht I, Dixon A (2009). Development of molecular genomic tools for verification of intergeneric hybrids between castor bean (*Ricinus communis* L.) and cassava (*Manihot esculenta* Crantz). J. Food Agric. Environ. 7:534-539.

Hahn S, Isoba J, Ikotun T (1989). Resistance breeding in root and tuber crops at the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. Crop Protect. 8:147-168.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41:95-98.

Hammond-Kosack KE, Parker JE (2003). Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. Curr. Opin. Biotechnol. 14:177-193.

Hillocks RJ, Jennings DL (2003). Cassava brown streak disease: a review of present knowledge and research needs. Int. J. Pest Manag. 49:225-234.

Jennings DL (1994). Breeding for resistance to African cassava mosaic geminivirus in East Africa. Trop. Sci. 34:110-122.

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

Kawuki R, Ferguson M, Labuschagne M, Herselman L, Kim DJ (2009). Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). Mol. Breed. 23: 669-684.

Lawson WR, Jan CC, Shatte T, Smith L, Kong GA, Kochman JK (2010). DNA markers linked to the R2 rust resistance gene in sunflower (*Helianthus annuus* L.) facilitate anticipatory breeding for this disease variant. Mol. Breed. 28:569-576.

Legg JP, Fauquet CM (2004). Cassava mosaic geminiviruses in Africa. Plant Mol. Biol. 56:585-599.

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. Nature Gene. 14:421-429.

Lokko Y, Danquah EY, Offei SK, Dixon AGO, Gedil MA (2005). Molecular markers associated with a new source of resistance to the cassava mosaic disease. Afr. J. Biotechnol. 4:873-881.

Lopez CE, Zuluaga AP, Cooke R, Delseny M, Tohme J, Verdier V (2003). Isolation of Resistance Gene Candidates (RGCs) and characterization of an RGC cluster in cassava. Mol. Genet. Genomics 269:658-671.

McDowell JM, Simon SA (2006). Recent insights into R gene evolution. Mol Plant Pathol. 7:437-448.

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 within the nucleotide-binding superfamily. Plant J. 20:317-332.

Meyers BC, Kaushik S, Nandety RS (2005). Evolving disease resistance genes. Curr. Opin. Plant Biol. 8:129-134.

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.

Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di Gaspero G, Morgante M, Adam-Blondon AF (2008). A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. BMC Plant Biol. pp. 8-66.

Nassar N, Ortiz R (2007). Cassava Improvement: challenges and impacts. J. Agric. Sci. 145:165-171.

Nawaz-Ul-Rehman MS, Fauquet CM (2009). Evolution of geminiviruses and their satellites. FEBS Lett. 583:1825-1832.

Okogbenin E, Porto MCM, Egesi C, Mba C, Espinosa E, Guillermo Santos L, Ospina C, Marin J, Barera E, Gutierrez J, Ekanayake I, Iglesias I, Fregene M (2007). Marker assisted introgression of CMD resistance in Latin American Germplasm for genetic improvement of cassava in Africa. Crop Sci. 47:1895-1904.

Okuyama Y, Kanzaki H, Abe A, Yoshida K, Tamiru M, Saitoh H, Fujibe T, Matsumura H, Shenton M, Galam DC, Undan J, Ito A, Sone T, Terauchi R (2011). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. Plant J. 66:467-479.

Pan Q, Wendel J, Fluhr R (2000). Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. J. Mol. Evol. 50:203-213.

Patil BL, Fauquet CM (2010). Differential interaction between cassava mosaic geminiviruses and geminivirus satellites. J. Gen. Virol.

91:1871-1882.
Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009). Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Mol. Genet. Genomics 281:609-626.
Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2012). The Cassava Genome: Current Progress, Future Directions. Trop. Plant Biol. 5:88-94.
Qiu JW, Schurch AC, Yahiaoui N, Dong LL, Fan HJ, Zhang ZJ, Keller B, Ling HQ (2007). Physical mapping and identification of a candidate for the leaf rust resistance gene Lr1 of wheat. Theor. Appl. Genet. 115:159-168.
Rabbi IY, Kulembeka HP, Masumba E, Marri PR, Ferguson M (2012). An EST-derived SNP and SSR genetic linkage map of cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 125:329-342.
Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.
Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res. 38:D814-D821.
Seo YS, Rojas MR, Lee JY, Lee SW, Jeon JS, Ronald P, Lucas WJ, Gilbertson RL (2006). A viral resistance gene from common bean functions across plant families and is up-regulated in a non-virus-specific manner. Proc. Natl. Acad. Sci. USA 103:11856-11861.
Stange C (2006). Plant-virus interactions during the infective process. Cien. Inv. Agr. 33:1-18.
Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596-1599.
Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, Bent AF, Michelmore RW (2007). Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol. 7:56.
Tarr DE, Alexander HM (2009). TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res. Notes 2-197.
van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007). Structure and function of resistance proteins in solanaceous plants. Annu. Rev. Phytopathol. 45:43-72.
Wan H, ZhaoZ, Malik A, Qian C, Chen J (2010). Identification and characterization of potential NBS-encoding resistance genes and induction kinetics of a putative candidate gene associated with downy mildew resistance in Cucumis. BMC Plant Biol. 10:186.
Waterhouse AM, Procter JB, Martin DMA, Clamp Ml, Barton GJ (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189-1191.
Yang S, Feng Z, Zhang X, Jiang K, Jin X, Hang Y, Chen JQ, Tian D (2006). Genome-wide investigation on the genetic variations of rice disease resistance genes. Plant Mol. Biol. 62:181-193.
Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004). Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. Mol. Genet. Genomics 271:402-415.