GMR
The Original by FUNPEC-RP

Identification, Characterization, and Phylogenetic analysis of Pigeonpea (Cajanus cajan L. Millsp.) Resistance Gene Analogs using PCR cloning and in silico methods

I.O. Agbagwa[1, 2], P. G. Patil[1], A. Das[1], K. R. Soren[1], I. P. Singh[1], S. K. Chaturvedi[1], N. Nadarajan[1]

[1]Crop Improvement Division, Indian Institute of Pulses Research (IIPR), Kanpur, India
[2]Department of Plant Science & Biotechnology, Faculty of Science, University of Port Harcourt, Port Harcourt, Nigeria

Corresponding author: I.O. Agbagwa
E-mail: ikechukwu.agbagwa@uniport.edu,ng

Genet. Mol. Res. 17 (3): gmr16039895
Received June 03, 2018
Accepted July 06, 2018
Published August 01, 2018
DOI http://dx.doi.org/10.4238/gmr16039895

ABSTRACT. Pigeonpea (*Cajanus cajan*), an important grain legume, is susceptible to *Fusarium* wilt (FW), sterility mosaic disease (SMD), and *Phytophthora* blight. Identification of resistance gene analogs (RGAs) is important for development of resistant varieties. In this study, degenerate primers targeting nucleotide binding sites (NBS) of known resistance (R) genes were used to amplify RGAs from two pigeonpea genotypes with differing disease resistance profiles. The translated cloned RGAs had high amino acid identity (68–71%) with putative disease resistance proteins in *Glycine max*. Five RGA open reading frames were found in the whole pigeonpea genome after BLASTN analysis with the cloned sequences. Translated RGA proteins contained several characteristic features such as the NB-ARC domain (characteristic of death-related disease resistance genes) and four NBS motifs. A tryptophan residue at the kinase-2 motif was indicative of the non-TIR-NBS class of proteins. Phylogenetic analysis revealed two major clusters. The seven pigeonpea RGAs were in a non-TIR group alongside wilt resistance proteins from tomato. Specific primers were designed against the RGAs identified by

BLASTN, and these successfully amplified sequences from all eight pigeonpea genotypes. The 40 resultant sequences were combined according to genotype and subjected to phylogenetic analysis. Genotypes clustered according to breeding pedigree. Multiple alignment of the 40 sequences revealed a number of single nucleotide polymorphisms (SNPs) that are useful in identifying candidate resistance genes associated with FW and SMD.

Key words: NB-ARC, NBS-LRR, Phylogenetic analysis, Pigeonpea, Resistance Gene Analog (RGA)

INTRODUCTION

Pigeonpea (*Cajanus cajan* L. Millsp.) is one of the most important grain legume crops worldwide. It is mainly cultivated as an annual crop in Asia, Africa, the Caribbean, and Latin America (Saxena *et al.*, 2010). India is one of the largest pigeonpea producers, where the crop occupies an area of about 3.4 million hectares and produces an average annual yield of 2.89 million tons (DAC, 2011). However, expected pigeonpea productivity levels have not yet been achieved. This is largely due to substantial yield losses caused by major biotic stresses like *Fusarium* wilt (FW), sterility mosaic disease (SMD), and *Phytophthora* blight. Exploitation of host resistance mechanisms through development of resistant varieties is one strategy to address these issues.

Plants have an innate immune system and encode a large family of dominant resistance (R) proteins. NBS-LRR proteins, which are particularly widely distributed in plants, have several conserved features, including an N-terminal variable region, central nucleotide binding site (NBS) region, and C-terminal LRR motifs (Sharma *et al.*, 2009). The NBS-LRR gene family is classified into two subfamilies according to sequence variation at the N-terminus: TIR-NBS-LRR (similar to *Drosophila Toll*) and non-TIR-NBS-LRR (similar to mammalian *Interleukin-1*) (Hulbert *et al.*, 2001). A single amino acid difference (W/D) at the kinase-2 motif in resistance genes can be used to classify TIR (W) and non-TIR (D) proteins (Pei *et al.*, 2007). The NBS domain consists largely of four motifs: P-loop/kinase-1a [GGV(I/M)GKTT], kinase-2 [LVLDDVW(D)], kinase-3a (GSRIIITTRD), and hydrophobic domain [GL(F)PL(F)AL]. The P-loop and kinase-2 domains have ATP and GTP binding sites (Meyers *et al.*, 1999), and the kinase-3a and hydrophobic domains have putative membrane spanning domains (Baldi *et al.*, 2004). In plants, the domains in NBS-LRR proteins co-operate to mediate elicitor recognition and activate downstream signal pathways to stimulate the disease resistance response.

Overall sequence similarities between plant NBS-LRR genes vary, but sequences encoding several functional motifs, such as the LRR motifs and motifs in the NBS region, are highly conserved. These conserved sequences enabled rapid isolation of NBS-LRR genes by degenerate polymerase chain reaction (PCR) in a number of crops such as soybean (Yu *et al.*, 1996), lettuce (Shen *et al.*, 1998), rice (Mago *et al.*, 1999), wheat (Seah *et al.*, 2000), chickpea (Huettel *et al.*, 2002), and *Medicago truncatula* (Zhu *et al.*, 2002). These resistance gene analog (RGA) fragments were used as molecular markers for tagging disease resistance loci

in several species, including *Arabidopsis* (Aarts *et al.*, 1998), rice (Ilag *et al.*, 2000), cowpea (Gowda *et al.*, 2002), tomato (Zhang *et al.*, 2002), and common bean (Lopez *et al.*, 2003).

The pigeonpea genome sequence and concomitant genetic resources are now available (Singh *et al.*, 2012; Varshney *et al.*, 2012). However, no candidate resistance genes have been identified to date. In pigeonpea, heritable resistance is generally found against FW and SMD. The aims of the present study were as follows: i) cloning of RGAs from pigeonpea plants exhibiting susceptibility or resistance to FW and SMD, ii) *in silico* analysis and identification of RGAs from the pigeonpea genome sequence, and iii) determination of phylogenetic relationships between pigeonpea genotypes and identification of RGA-based markers. Polymorphisms revealed in the study will be useful in developing resistance gene-based markers such as cleaved amplified polymorphic sequences (CAPS) and targeted region amplification polymorphisms (TRAP).

MATERIALS AND METHODS

Plant material and dna isolation

Eight pigeonpea genotypes (seven cultivars and one wild relative, *C. cajanifolius*) were used in this study (Table 1). Seeds were provided by the Crop Improvement Division, Indian Institute of Pulses Research (IIPR), Kanpur, Uttar Pradesh, India. Genotypes were selected largely on their responses to two major pigeonpea diseases, FW and SMD (Table 1). Genomic DNA was extracted from 15-day-old seedlings using a modified CTAB method (Agbagwa et al., 2012). The resultant DNA samples were treated with RNase, purified using phenol-chloroform-isoamyl alcohol, and finally ethanol precipitated at -20˚C. DNA samples were analyzed using 0.8% agarose gel electrophoresis.

Table 1. Eight pigeonpea genotypes used in the present study

Serial Number	Pigeonpea genotypes	Species	Pedigree/Origin	Disease reaction	
				FW	SMD
1	UPAS 120	*C. cajan*	Selection from P4768	S	S
2	ICPL 87119	*C. cajan*	C11 x ICPL6	R	R
3	BSMR 853	*C. cajan*	(ICPL7336 x BDN1) x BDN2	R	R
4	ICP 8863	*C. cajan*	Selection from landrace of Mharashtra, India	R	S
5	BDN 2	*C. cajan*	Local selection from Borigermplasm	R	S
6	PUSA 9	*C. cajan*	UPAS 120 x 3673	R	R
7	DA 11	*C. cajan*	[Bahar x NP (WR)15] x PS16	S	R
8	ICP–1629 -1	*C. cajanifolious*	ICRISAT collection	S	S

FW, *Fusarium* wilt; SMD, Sterility Mosaic Disease; S, Susceptible; R, Resistant

PCR amplification of resistance gene analogs (RGAs)

Two pigeonpea genotypes, ICPL 87119 and UPAS 120, which were highly resistant and susceptible to FW and SMD, respectively, were selected for initial RGA amplification. Two degenerate primers were used: B1(f) (common bean), which targeted P-loop/kinase-1a (GGVGKTT), and B2(r) (soybean), which targeted the hydrophobic domain (GLPLAL) of NBS (Table 2). PCR amplification was performed in 25 µl reaction volumes. Each reaction contained 40 ng of genomic DNA, 20 pmol of each primer, and 2.5 µl of 10× PCR buffer containing 15 mM $MgCl_2$ (TaKaRa, India), 2 mM dNTP mix, and 3U *Taq* DNA polymerase

(TaKaRa). PCR amplification was performed using a thermal cycler (G-STORM, India) as follows: initial denaturation at 94ºC for 3 min, followed by 35 cycles of 30 sec at 94ºC, 1 min at 52ºC, and 1 min at 72ºC, and a final extension at 72ºC for 5 min. Amplification products were analyzed using a 1.5% agarose gel stained with ethidium bromide.

Table 2. Degenerate primers used for amplification of RGA sequences from pigeonpea genotypes ICPL 87119 and UPAS 120

Primer name	Primer sequence (5′–3′)	Conserved motif	Amplification product size and eluted (bp)	Species	Reference
B1-F	GG**I**GG**I**RTIGGIAARAC**I**AC	P-loop (GGVGKT)	520	Common bean	Rivkin et al. (1999)
B2-R	AR**I**GCTAR**I**GG**I**AR**I**CC	GLPLAL		Soybean	Kanazin et al. (1996)

Codes for degenerate positions: I, inosine; R, A/G.

Sequence diversity in the eight pigeonpea genotypes was examined with reference to FW and SMD susceptibility. Five primer pairs were designed for specific amplification of NBS domains of RGAs identified through *in silico* analysis (Table 3). PCR components and conditions used for specific amplification were as described above, except that an annealing temperature of 60ºC was used. Product sizes from the five reactions were 691–829 bp.

Table 3. Specific primers used for PCR amplification of a set of eight pigeonpea genotypes

Primer pair	Primer sequence (forward)	Primer sequence (reverse)	Expected size (bp)
PP_ICPL87119	CTT CCC TTC TTT CCA CTG GT	CAA CTT GGG CAT GGG TAT CT	233
AFS_226 G1	GCA AGG GAG GAT GAC AAA GA	CGT CAT GTG CCC ATA AGT TG	694
AFS_282 G2	CAT GTG CCC ATA AGT TGC TG	TGT TGC TAG GGA GGA TGA CA	695
AFS_884 G3	CAC GTG CCC ATA AAT TGC	GCA AGG GAG GAT GAC AAA GA	691
AFS_586 G4	AGA GAA GCT CCA GAG AAT GC	AAG AAA GCC TTC AGC CAT CC	820
AFS_519 G5	ACA TTT GCT GCC GAG GTT AG	GCA AAA GGC CTC CCA ATG T	829
Actin gene	CCT TGG TTA CGA GCA AGA GC	TTC CTC CTG AAA GGA CGA TG	250

Expression of the actin gene was used as an internal control in RT-PCR experiments.

Sequencing and in silico based identification of RGAs

Amplification products (520 bp) from PCR of pigenonpea ICPL 87119 and UPAS 120 genomic DNA with degenerate RGA primers (B1f and B2r) were separated using 1% agarose gel electrophoresis. Bands of the expected size were excised and purified by gel extraction (Qiagen Gel Extraction Kit, India). Purified amplicons were sequenced (MWG, India) and the sequences were used for *in silico* analysis. BLASTX and BLASTN searches were performed against the NCBI protein and genomic databases, respectively. Matching contigs were searched for open reading frames (ORFs) similar to those of RGAs using an ORF-finder tool. Primer3 (Untergasser et al., 2012) was used to design six specific primer pairs from NBS domain nucleotide sequences (Table 3). Five of the six specific pairs targeted

all four internal NBS domain motifs (G1_WGS, G2_WGS, G3_WGS, G4_WGS, and G_WGS). One primer pair (PP_ICPL 87119) targeted two internal motifs (kinase-2 and kinase-3a).

cDNA preparation and amplification

Reverse transcriptase PCR (RT-PCR) of genotype ICPL 87119 was used to test whether the pigeonpea RGAs were functional. Total RNA was extracted from young leaves of pigeonpea seedlings using a modified CTAB method (Mehar *et al.,* 2000). RT-PCR using the six primer pairs (Table 3) was performed using a one-step RT-PCR Kit (MMLV) (Qiagen, India) according to the manufacturer's instructions. Actin expression was used as an internal control.

Amplification and sequencing of specific RGA products

Five specific primer pairs (Table 3) were used to amplify RGA sequences from eight pigeonpea genotypes. PCR products were separated on a 1% agarose gel stained with ethidium bromide. Specific amplicon sizes with different primer pairs were in the range of 691–829 bp. Amplicons were gel purified (Qiagen Gel Extraction Kit, India) and sequenced. In total, 40 PCR products were sequenced (MWG, India).

Sequence alignment and phylogenetic analysis

Pigeonpea RGA sequences were compared to the NCBI database using BLASTN and BLASTX. Nucleotide sequences were translated using the ExPASy algorithm (Gasteiger et.al, 2003). Using ClustalW, NBS domains from seven pigeonpea RGAs were aligned with twelve previously reported resistance (R) proteins from other plant species: tomato 12C-1 and 12C-2; tobacco N; maize RP1-D; rice Xa1; wheat Yr10; barley Mla6; and *Arabidopsis* RPM1, RPP1, RPP5, RPP8, and RPS2. Phylogenetic analysis and construction of neighbor-joining trees were performed using the Poisson Correction (PC) Distance Model of MEGA5 software (Tamura *et al.,* 2011). The robustness and reliability of tree branches were evaluated using bootstrap analysis with 1000 iterations. Nucleotide sequences were also compared using ClustalW. Sequences were edited using Bio-Edit software (Hall, 1999) and genetic relationships were determined by constructing neighbor-joining trees using the PC Distance Model of MEGA5 software (Tamura *et al.,* 2011). The reliability of tree branches was tested by bootstrap analysis using 1000 iterations.

RESULTS

RGA cloning and in silico based identification of RGAs

Degenerate primers B1(f) and B2(r), which aligned to the P-loop and GLPLAL regions of the NBS domain, respectively, were used to amplify a 520 bp DNA fragment from two pigeonpea genotypes with differing disease resistance profiles (ICPL 87119 and UPAS 120) (Fig. 1A). The translated amplicon sequences were used to search the NCBI protein

database (BLASTX) to identify potential RGAs (Table 4). Numerous sequences with high similarity to NBS-LRR-, CC-NBS-, and CC-NBS-LRR-type resistance (R) proteins were identified from *Glycine max*, *Medicago truncatula*, *Cicer arietinum*, *Vicia faba*, and other species. The highest identity match (87%) was to a partial *Cajanus* sequence of unknown function. The next-highest identities were seen with putative disease resistance proteins At3g14460, At3g14460-like, and RPP13 from *Glycine max* (68–71% identity).

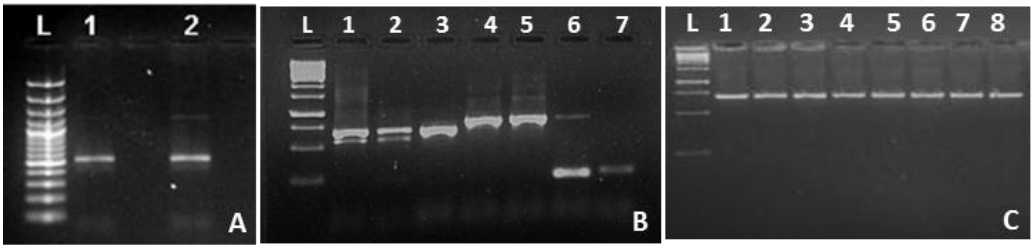

Figure 1. PCR amplification of RGA sequences. **A)** Amplification from UPAS120 (lane 1) and ICPL 87119 (lane 2) pigeonpea genomes using degenerate primers. **B)** RT-PCR of ICPL 87119 using six RGA-specific primer pairs (lanes 1–6) and an actin internal control (lane 7). **C)** Amplification of eight pigeonpea genomes (Table 1) using a specific RGA primer pair.

Table 4. Putative RGAs identified by BLASTX search of the NCBI database with RGA sequences from pigenonpea genotypes UPAS 120 and ICPL 87119

NCBI protein accession	Accession number	Amino acid identity (%)	*E-value
Unknown partial sequence (*Cajanus cajan*)	AAF36345	87	4e-71
Putative disease resistance protein At3g14460 like (*Glycine max)*	XP 003543829	71	3e-49
Putative disease resistance RPP13 like protein 1 (*Glycine max*)	XP 003556801	70	1e-47
Putative disease resistance protein At3g14460 (*Glycine max*)	XP 003556802	68	6e-46
NBS-LRR type disease resistance protein (*Medicago truncatula*)	XP 003599594	57	3e-39

*E-value refers to the number of matches expected by chance. Lower E values indicate higher similarity.

The two cloned RGA sequences isolated from the ICPL 87119 and UPAS 120 pigeonpea genotypes were used to search the whole pigeonpea shotgun genome. Five sequence contigs (19925, 43339, 18318, 27641, and 47635) were identified with sequence identities of 82–96% (Table 5). Five RGA-like ORFS were identified in these contigs using ORFfinder. All seven pigeonpea RGA sequences were translated using the ExPASy algorithm (Gasteiger et.al, 2003).

Table 5. Putative RGA sequences identified in the pigeonpea shotgun genome by BLASTN analysis using pigeonpea RGA sequences from ICPL 87119 and UPAS120

GeneBank accession with highest similarity	GeneBank ID	Maximum identity (%)	E-value
Cajanus cajan contig19925, whole genome shotgun sequence	AFSP01019884	96	0.0
Cajanus cajan contig43339, whole genome shotgun sequence	AFSP01043226	96	0.0
Cajanus cajan contig18318, whole genome shotgun sequence	AFSP01018282	95	0.0
Cajanus cajan contig27641, whole genome shotgun sequence	AFSP01027586	95	3e-178
Cajanus cajan contig47635, whole genome shotgun sequence	AFSP01045301	82	7e-85

Multiple sequence alignment of pigeonpea RGAs with other known R genes

A conserved domain search of the NCBI database indicated that all seven pigeonpea sequences belonged to the NB-ARC domain family, which was characteristic of death-related disease resistance genes. ClustalW was used to align the seven pigeonpea RGA consensus sequences with NBS sequences of 12 known R genes: tomato *12C-1* and *12C-2*; tobacco *N*; maize *RP1-D*; rice *Xa1*; wheat *Yr10*; barley *Mla6*; and *Arabidopsis RPM1*, *RPP1*, *RPP5*, *RPP8*, and *RPS2*. Three of these proteins (*N*, *RPP1*, and *RPP5*) had an aspartic acid (D) residue at the final position in the kinase-2 motif; this was indicative of the TIR-NBS subfamily. All the other R proteins, including the seven pigeonpea RGAs, had tryptophan (W) at this position, indicative of the non-TIR-NBS subfamily. Sequence comparison with known R proteins revealed four crucial motifs (P-loop/kinase-1a, kinase-2, kinase-3a, and hydrophobic domain) in the five pigeonpea RGAs identified using ORFfinder. The two PCR-cloned RGAs had only two motifs (kinase-2 and kinase-3a) (Fig. 2)

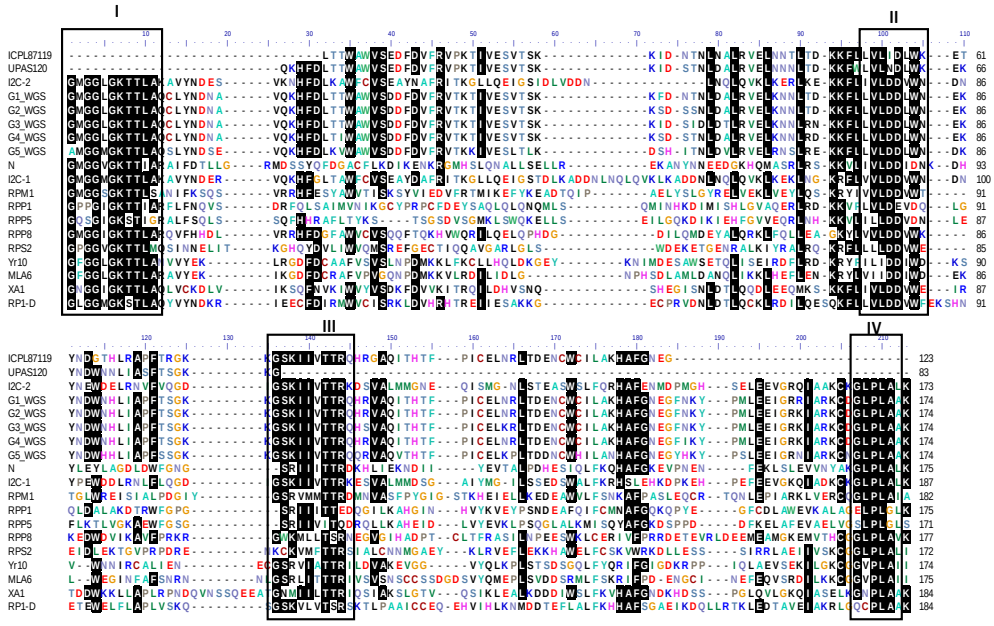

Figure 2. ClustalW alignment of sequences from seven pigeonpea RGAs with NBS domain sequences from other plant R proteins: tomato *I2C-1* (Accession No. AF004878) and *I2C-2* (Accession No. AF004879); maize *RP1-D* (Accession No. AF107294); rice *Xa1* (Accession No. BAA25068); wheat *Yr10* (Accession No. AF149114); barley *Mla6* (Accession No. AJ302292); and *Arabidopsis RPM1* (Accession No. AAF27008), *RPP1* (Accession No. AF098962), *RPP5* (Accession No. AAF08790), *RPP8* (Accession No. AF089710), and *RPS2* (Accession No. U14158).

Phylogenetic analysis of pigeonpea RGAs and other known plant R proteins

Phylogenetic analysis was performed using pigeonpea RGA protein sequences and sequences of 12 known *R* proteins from other plants. Neighbor-joining phylogenetic analysis yielded two major clusters, I and II (Fig. 3A). Cluster I contained 16 non-TIR subfamily proteins and cluster II contained three TIR subfamily sequences. The seven pigeonpea RGAs were assigned to the non-TIR subfamily. Phylogenetic trees had high bootstrap values at most of the nodes (Fig. 3A). Pairwise distances, as determined by maximum composite likelihood, were 0.03–2.47% (Table 6). Pigeonpea RGA sequence divergence was 0.03–0.94%, indicating high sequence divergence. The pigeonpea RGAs were most similar to wilt resistance genes in tomato (I2C-1 and I2C-2), with mean genetic divergence of 0.76%; however, the bootstrap value was 45 (Figure 3A). The two pigeonpea RGAs initially amplified using the degenerate primers had a sequence divergence of ~0.19%.

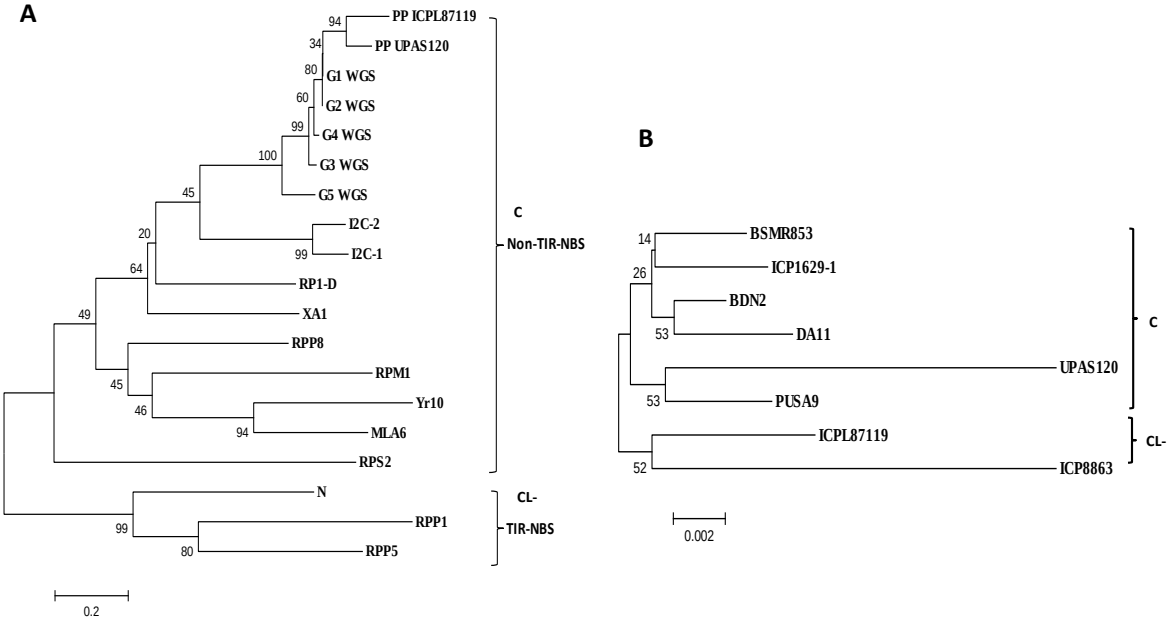

Figure 3. Phylogenetic analysis of **A)** amino acid sequences of NBS domains of pigeonpea RGAs and other known plant R proteins; and **B)** nucleotide sequences of RGA genes from eight pigeonpea genotypes. Multiple sequences (40) amplified by five different primer sets were combined for analysis. CL-I, Cluster I; CL-II, Cluster II.

Table 6. Maximum composite likelihood pairwise distance values for pigeonpea RGA amino acid sequences and sequences of other plant R genes.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
PP_ICPL 87119	0																		
PP_UPA S120	0.19	0																	
I2C-2	0.86	0.94	0																
G1_WGS	0.19	0.13	0.75	0															
G2_WGS	0.19	0.13	0.75	0.00	0														
G3_WGS	0.25	0.19	0.75	0.05	0.05	0													
G4_WGS	0.23	0.17	0.75	0.03	0.03	0.05	0												
G5_WGS	0.36	0.34	0.71	0.21	0.21	0.19	0.19	0											
N	1.88	2.00	1.44	1.68	1.68	1.59	1.59	1.51	0										
I2C-1	0.82	0.90	0.19	0.71	0.71	0.71	0.71	0.78	1.68	0									
RPM1	1.59	1.51	1.19	1.44	1.44	1.44	1.44	1.51	2.13	1.33	0								
RPP1	2.13	2.13	1.88	2.00	2.00	2.00	1.88	1.88	1.13	2.13	2.13	0							
RPP5	2.00	2.13	1.59	1.77	1.77	1.77	1.77	1.77	1.24	1.77	2.00	1.03	0						
RPP8	1.19	1.24	1.90	1.19	1.19	1.24	1.19	1.19	1.77	0.94	1.08	2.00	1.44	0					
RPS2	1.77	2.00	1.51	1.68	1.68	1.59	1.51	1.51	1.88	1.44	1.51	1.88	2.00	1.51	0				
Yr10	1.68	1.51	1.37	1.51	1.51	1.51	1.44	1.44	1.59	1.30	1.37	2.00	2.29	1.37	1.68	0			
MLA6	1.51	1.59	1.19	1.44	1.44	1.37	1.44	1.30	2.00	1.03	1.13	2.47	2.47	0.99	1.51	0.75	0		
XA1	1.08	1.08	0.86	0.90	0.90	0.86	0.86	0.90	1.77	0.99	1.24	1.13	1.77	1.19	1.30	1.44	1.30	0	
RP1-D	1.03	0.99	0.99	0.82	0.82	0.78	0.82	0.75	2.00	1.08	1.19	2.00	1.88	1.08	1.37	1.68	1.37	0.78	0

Nucleotide sequencing-based diversity analysis

Sequences were generated by amplification of two pigeonpea genomes with degenerate primers. Five specific primer pairs were designed from these sequences and used for PCR and RT-PCR analysis (Fig. 1A and 1B). The eight pigeonpea genotypes were successfully amplified using these primers and 691–829 bp PCR products were produced (Fig. 1C). Amplicons were sequenced and, in total, forty sequences were obtained. Multiple sequence alignments revealed several single nucleotide variations (Fig. 4). Nucleotide sequences amplified by the five primer pairs were combined for each genotype and used to construct a neighbor-joining tree (Fig. 3B). Cluster I contained six genotypes (BSMR 853, ICP1629-1, BDN 2, DA 1, UPAS 120, and PUSA 9) and cluster II contained two genotypes (ICPL 87119 and ICP 8863). Bootstrap values indicated the reliability of the tree. Expression

of the seven RGAs in pigeonpea was confirmed by RT-PCR of ICPL 87119 with six primer sets (Table 3, Fig. 1B).

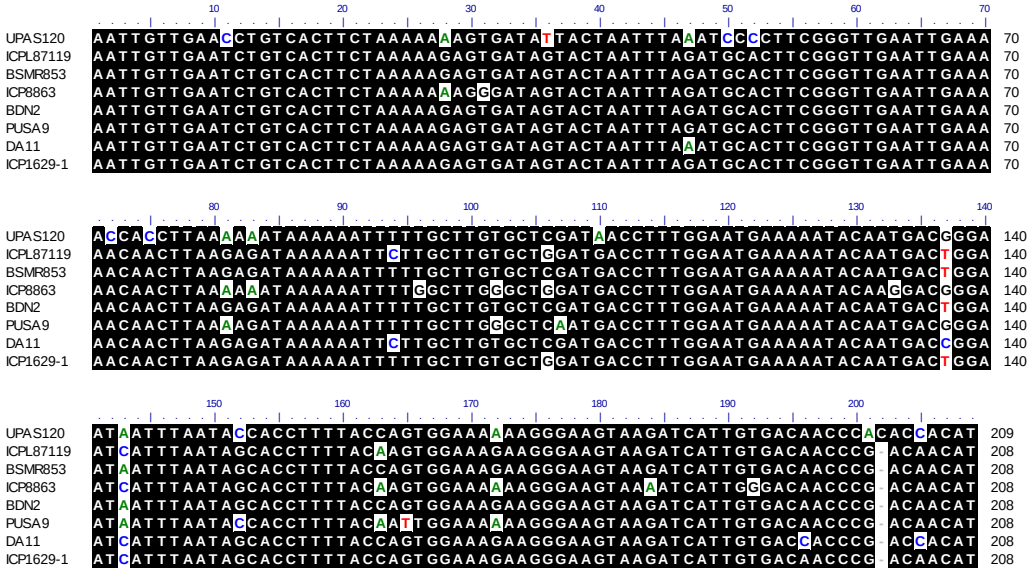

Figure 4. Multiple sequence alignment of RGA nucleotide sequences from eight pigeonpea genotypes. Single nucleotide polymorphisms are highlighted in white.

DISCUSSION

Plant resistance genes play a crucial role in imparting resistance against many diseases caused by fungal, bacterial, and viral agents. As an alternative to transposon tagging and map-based cloning (Seah *et al.*, 1998), PCR-based cloning using degenerate primers for RGA genes has allowed the isolation and characterization of resistance gene analogs in many plant species. In crops with available whole genome sequence information, a candidate gene approach can be used to efficiently investigate associations between candidate genes and their functions. If candidate genes are themselves involved in disease resistance, they can be used for marker-assisted selection. This approach was accelerated with the publication of the draft rice genome (Goff *et al.*, 2002). Disease resistance genes share common sequence features, such as the CC, NBS, LRR, and kinase domains (Hulbert *et al.*, 2001). Genome scanning for conserved motifs allowed the detection of many resistance genes in several crops, such as rice (Monosi *et al.*, 2004), *Medicago truncatula* (Ameline-Torregrosa *et al.*, 2008), *Brassica rapa* (Mun *et al.*, 2009), and *Lotus japonicas* (Li *et al.*, 2010). The NBS and LRR regions have been widely used for the design of degenerate primers for PCR cloning of RGAs (Garcia-Mas *et al.*, 2001).

In this study, we isolated and characterized RGAs from the pigeonpea genome using PCR and *in silico* analysis. Two degenerate primers from species phylogenetically related to pigeonpea were used: forward primer B(f) from common bean (Rivkin *et al.*, 1999) and reverse primer B (r) from soybean (Kanazin *et al.*, 1996). A PCR product of expected size

(520 bp) was amplified from two pigeonpea genotypes (ICPL 87119 and UPAS 120). The sequences of these cloned RGAs facilitated the *in silico* identification of similar genes in the pigeonpea genome.

Identification of RGA sequences is more reliable when comparison is carried out at the amino acid rather than the nucleotide level (Totad *et al.,* 2005). The cloned pigeonpea RGA amino acid sequences were therefore used for BLASTX analysis. The RGA sequences were highly similar (68–71%) to putative disease resistance-like proteins At3g14460 and RPP13 from *Glycine max*, a close relative of pigeonpea. In addition, the translated amino acid sequences had high sequence identity with NBS-LRR, CC-NBS, and CC-NBS-LRR resistance proteins from other legumes such as *Medicago truncatula*, *Cicer arietinum*, and *Vicia faba*. Conserved domain searching revealed the presence of NB-ARC domains in the two pigeonpea RGA sequences. Proteins containing the NB-ARC domain are often involved in regulation of cell death in animals and resistance in plants. More than 70% of the approximately 40 resistance proteins that have been characterized to date possess an NB-ARC domain (Selvaraj *et al.,* 2011).

Deduced amino acid sequences of seven pigeonpea RGAs were aligned with twelve known plant R proteins. Alignments revealed substantial similarity at the four motifs (P-loop/kinase-1a, kinase-2, kinase-3a, and hydrophobic domain) of the NBS domain. As was evident from the consensus amino acid (tryptophan/W) observed at the kinase 2 motif, all seven pigeonpea RGAs belonged to the non-TIR subfamily of R proteins. NBS-LRR proteins were grouped mainly into two subfamilies: subfamily I (dicots only) and subfamily II (dicots and monocots). Subfamily I proteins contained the TIR element, whereas subfamily II proteins contained only non-TIR elements (Meyers *et al.,* 1999; Pan *et al.,* 2000). Previous research suggested that, whereas loci encoding TIR-NBS-LRR R proteins may have been lost, non-TIR resistance genes have expanded and diversified throughout evolution (Seah *et al.,* 1998; Leister, 2004; McHale *et al.,* 2006).

Phylogenetic analysis of the seven pigeonpea RGAs and twelve known plant R proteins revealed two major clusters: cluster I and cluster II. Cluster I contained non-TIR subfamily R proteins (12C-1, 12C-2, RP1-D, Xa1, Yr10, Mla6, RPM1, RPP8, and RPS2). Cluster II contained TIR subfamily R proteins (N, RPP1, and RPP5). Similar groupings were observed in other crops (Pei *et al.,* 2007; Thirumalaiandi *et al.,* 2008; Lu *et al.,* 2011). All the pigeonpea RGAs grouped with the non-TIR subfamily R proteins. Pairwise comparison values for RGA sequences were 0.03–2.47, indicating a wide range of sequence divergence. The seven pigeonpea RGAs grouped together (with bootstrap values of 100%) and had pairwise distance values of 0.03–0.94%. These results reflected the substantial sequence differences between the seven pigeonpea RGAs. The two PCR-cloned pigeonpea RGAs exhibited 0.19% sequence variation. These sequence variations may be of use for identification of candidate genes influencing disease resistance in pigeonpea.

All seven pigeonpea RGAs were in the same phylogenetic group, alongside two wilt resistant proteins from tomato (I2C-1 and I2C-2), with bootstrap values of 45%. Mean sequence divergence between pigeonpea RGAs and tomato wilt resistance proteins was minimal (<0.76%), indicating strong similarity between the sequences. In banana, some RGAs were identified with 28–54% similarity to known wilt resistance proteins such as Fom-2, I2C-1, I2C-2, and I2 (Sun *et al.,* 2009). The RGAs identified in this study might therefore be candidates for wilt resistance in pigeonpea.

The relationships between RGA gene nucleotide sequences were also examined. Using five RGA-specific primer sets, 40 sequences were amplified and sequenced from the eight pigeonpea genotypes. Phylogenetic analysis confirmed the existing relationships of these genotypes. Genotypes BDN2, BSMR 853, ICP 1629-1, and DA 11 were grouped together. This reflected their breeding relationships and suggested that some common ancestral genes were shared between the genotypes. The two genotypes UPAS 120 and PUSA 9 were grouped together but had substantial sequence divergence, as is evident from the tree. This may reflect their different resistance traits (UPAS 120 is highly susceptible to both FW and SMD, but PUSA 9 is highly resistant to both). Finally, genotypes ICPL 87119 and ICP 8863 grouped together in major cluster II, with a bootstrap value of 52%. Although these two genotypes share common resistance genes for FW, wide sequence divergence was evident from the tree. Branch length indicated that two genotypes (UPAS 120 and ICP 8863) were particularly divergent. These two genotypes have contrasting resistance characteristics with respect to *Fusarium* wilt and SMD.

Several single nucleotide polymorphisms (SNPs) were found when nucleotide sequences were aligned. These nucleotide variations will be helpful in developing cleaved amplified polymorphic markers (CAPS) that can be used for the discovery of candidate genes for resistance to FW and SMD in pigeonpea. Specific primers for each RGA class, in combination with a number of restriction enzymes, are used to detect polymorphic CAPS markers to map resistance genes for different diseases in crops such as maize (Quint *et al.*, 2002), chickpea (Palomino *et al.*, 2009), and common bean (Liu *et al.*, 2012). The specific primers developed in this study will facilitate the development of targeted region amplification polymorphism (TRAP) markers that can be used for trait mapping in pigeonpea.

In summary, PCR cloning of RGA sequences allowed the identification of five similar RGA types from the whole pigeonpea genome. Gene sequences were determined after amplification by specific primers, allowing the genetic relationships between eight pigeonpea genotypes to be determined. The sequence variation observed between genotypes for these genes will facilitate the development of CAPS markers, which can in turn be used to identify candidate resistance genes associated with FW and SMD in pigeonpea.

ACKNOWLEDGEMENTS

This work was supported by the Department of Biotechnology (DBT) India and the Third World Academy of Science (TWAS) Trieste Italy (DBT-TWAS) Postdoctoral Fellowship in Plant Biotechnology/Genomics.

REFERENCES

Aarts MGM, Hekkert BTL, Holub EB, Beynon JL, Stiekema WJ, Pereira A. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Mol. Plant-Microbe Interact. 11: 251-258.

Ameline-Torregrosa C, Wang BB, Obleness 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.

Baldi P, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M. 2004. Cloning and linkage mapping of resistance gene homologues in apple. Theor. Appl. Genet. 109: 231-239.

DAC. 2009. Fourth Advance Estimate of Production of Food grains for 2010-11, Agricultural Statistic Division, Department of Agriculture and Cooperation, Government of India, New Delhi http://eands.dacnet.nic.in/Advance_Estimate/ 3rd Advance_Estimates_2010-11.

Garcia-Mas J, Van Leeuwen H, Monfort A, Carmen V, Nech PP, Aru P. 2001. Cloning and mapping of resistance gene homologues in melon. Plant Sci. 161: 165-172.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31:3784-3788.

Goff SA, Ricke D, Lan TH, Presting G, Wang R et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296: 92 - 100.

Gowda BS, Miller JL, Rubin SS, Sharma DR, Timko MP. 2002. Isolation, sequence analysis, and linkage mapping of resistance-gene analogs in cowpea (*Vigna unguiculata* L. Walp). Euphytica 126: 365-377.

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.

Huettel B, Santra D, Muehlbauer F, Kahl G. 2002. Resistance gene analogues of chickpea (*Cicer arietinum*L.): isolation, genetic mapping and association with a *Fusarium* resistance gene cluster. Theor. Appl. Genet. 105: 479-490.

Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene cornplexes: evolution and utilization. Annu. Rev. Phytopathol. 39: 285-312.

Ilag LL, Yadav RC, Huang N, Ronald PC, Ausubel FM. 2000. Isolation and characterization of disease resistance gene homologues from rice cultivar IR64. Gene 255: 245-255.

Jain PK, Bhattacharya R, Gaikwad K, Mohapatra T, Srinivasan R, Sharma TR. 2012. The first draft of the pigeonpea genome sequence. J. Plant Biochem. Biotechnol. 21: 98-112.

Kanazin V, Frederick ML, Shoemaker RC. 1996. Resistance gene analogs are conserved and clustered in soybean. Proceedings of the National Academy of Sciences of the United States of America 93: 11746-11750.

Leister D. 2004. Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. Trends Genet. 3: 116-122

Li X, Cheng Y, Ma W, Zhao Y, Jiang H. 2010. Identification and characterization of NBS-encoding disease resistance genes in Lotus japonicus. Plant Syst. Evol. 289: 101–110.

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.

Lopez CE, Acosta IF, Jara C, Pedraza F, Gaitan-Solis E, Gallego G, Beebe S, Tohme J. 2003. Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathol. 93: 88-95.

Lu Y, Xu WH, Xie YX, Zhang X, Pu JJ, Qi YX, Li HP. 2011. Isolation and characterization of nucleotide-binding site and C-terminal leucine-rich repeat-resistance gene candidates in bananas. Genet Mol. Res. 10: 3098-3108.

Mago R, Nair S, Mohan M. 1999. Resistance gene analogues from rice: cloning, sequencing and mapping. Theor. Appl. Genet. 99: 50-57.

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

Mehar H. Puneet D, Pravendra N. 2002. A simple procedure for the isolation of high quality RNA from banana ripening fruit. Plant Mol. Biol. Rep. 18: 109-115.

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.

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

Mun JH, Yu HJ, Park S, Park BS. 2009. Genome-wide identification of NBS encoding resistance genes in *Brassica rapa.* Mol. Genet. Genomics 282: 617–631.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8: 4321-4325.

Palomino C, Fermandez-Romero MD, Rubio J, Torres A, Moreno MT, Millan T. 2009. Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance. Theor. Appl. Genet. 118: 671-682.

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.

Quint M, Mihaljevic R, Dussle M, Xu L, Melchinger E, Lubberstedt T. 2002. Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. Theor. Appl. Genet. 105: 355-363.

Rivkin MI, Vallejos CE, Mcclean PE. 1999. Disease-resistance related sequences in common bean. Genome 42: 41-47.

Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK. 2010. Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breed. 129: 142–148.

Seah S, Spielmeyer W, Jahier J, Sivvvasithamparam K, Lagudah ES. 2000. Resistnce gene analogs within an introgressed chromosomal segment derived from *Triticum ventricosum* that confers resistnce to nematodes and rust pathogens in wheat. Mol. Plant-Microbe Interact. 13: 334-341.

Seah S, Sivasithamparam K, Karakousis A, Lagudah ES. 1998. Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. Theor. Appl. Genet. 97: 937-945.

Selvaraj AI, Babu S, Sankar PD, Nagarajan P, Sabesan M. 2011. Genome scanning for identification of resistance gene analogs (RGAs) in a highly durable blast resistance rice (*Oryza sativa* L.) cultivar, Moroberekan. Afr. J. Biotechnol. 10: 6418-6433.

Sharma TR, Das A, Kumar SP, Lodha ML. 2009. Resistance Gene Analogues as a Tool for rapid Identification and Cloning of Disease Resistance in Plants – A Review. J. Plant Biochem. Biotechnol. 18: 01-11

Shen KA, Meyers BC, Islam-Faridi MN, Chin D, Stelly DM, Michelmore RW. 1998. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol. Plant-Microbe Interact. 11: 815-823.

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S et al. 2012. The first draft of the pigeonpea genome sequence. J. Plant Biochem. Biotechnol. 21: 98-112.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Mol. Biol. Evol.** 28:2731-2739.

Thirumalaiandi R, Selvaraj MG, Rajasekaran R, Subbarayalu M. 2008. Cloning and characterization of resistance gene analogs from underexploited plant species. Electronic J. Biotechnol.11 DOI: 10.2225/vol11-issue4-fulltext-11

Sun D, Hu Y, Zhang L, Mo Y, Xie J. 2009. Cloning and analysis of Fusarium wilt resistance gene analogs in 'Goldfinger' banana. Mol. Plant Breed. 7: 1215-1222.

Totad AS, Fakrudin B, Kuruvinashetti MS. 2005. Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). Euphytica 143: 179-188.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. Nucleic Acids Res. 40:e115.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK et al. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnol. 30: 83-89.

Yu YG, Buss GR, Maroof MAS. 1996. Isolation of a super- family of candidate disease resistance genes in soybean based on a conserved nucleotide- binding site. Proc. Natl. Acad. Sci. USA 93: 11751-11756.

Zhang LP, Khan A, Nino-Liu D and Foolad MR (2002) A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum* x *Lycopersicon hirsutum* cross. Genome 45: 133-146.

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 trauncatula*. Mol. Plant-Microbe Interact. 15: 529-539.

