Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (*Ipomoea batatas* Lam.)

Guanshui Chen[A], Daren Pan[1,*], Yifei Zhou[B], Sheng Lin[2] and Xiangde Ke[B]

[1]*College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350002, the People's Republic of China*

[2]*College of Crop Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian, 350002, the People's Republic of China*

**Corresponding author (Fax, 86-591-83769584; 86-591-83778870; Email, pandaren@yahoo.com.cn)*

Most plant disease-resistance genes (*R*-genes) isolated so far encode proteins with a nucleotide binding site (NBS) domain and belong to a superfamily. NBS domains related to *R*-genes show a highly conserved backbone of an amino acid motif, which makes it possible to isolate resistance gene analogues (RGAs) by degenerate primers. Degenerate primers based on the conserved motif (P-loop and GLPL) of the NBS domain from *R*-genes were used to isolate RGAs from the genomic DNA of sweet potato cultivar Qingnong no. 2. Five distinct clusters of RGAs (22 sequences) with the characteristic NBS representing a highly diverse sample were identified in sweet potato genomic DNA. Sequence identity among the 22 RGA nucleotide sequences ranged from 41.2% to 99.4%, while the deduced amino acid sequence identity from the 22 RGAs ranged from 20.6% to 100%. The analysis of sweet potato RGA sequences suggested mutation as the primary source of diversity. The phylogenetic analyses for RGA nucleotide sequences and deduced amino acids showed that RGAs from sweet potato were classified into two distinct groups—toll and interleukin receptor-1 (TIR)-NBS-LRR and non-TIR-NBS-LRR. The high degree of similarity between sweet potato RGAs and NBS sequences derived from *R*-genes cloned from tomato, tobacco, flax and potato suggest an ancestral relationship. Further studies showed that the ratio of non-synonymous to synonymous substitution within families was low. These data obtained from sweet potato suggest that the evolution of NBS-encoding sequences in sweet potato occur by the gradual accumulation of mutations leading to purifying selection and slow rates of divergence within distinct *R*-gene families.

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

1. Introduction

Recently, more than 50 disease-resistance genes (*R*-genes) that confer resistance to diverse pathogens have been isolated from a wide range of plant species by map-based cloning or transposon tagging (Dong *et al* 2001; Yi *et al* 2002). Amino acid sequence comparisons among these *R*-genes have revealed striking structural similarities, and conservation of specific domains that participate in protein–protein interactions and signal transduction (Dong *et al* 2001; Xu *et al* 2005). So far, these *R*-genes have been grouped into five classes based on the structure of

Keywords. Diversity; nucleotide-binding site; resistance gene analogue; sweet potato

Abbreviations used: RGA, resistance gene analogue; NBS, nucleotide-binding site; LRR, leucine-rich repeat; eLRR, extracellular leucine-rich repeat; TM, transmembrane; TIR, toll and interleukin receptor; CC, coiled-coil; LZ, leucine zipper; PCR, polymerase chain reaction; ORF, open reading frame; CTAB, cetyl trimethyl ammonium bromide

their predicted protein products: the nucleotide-binding site (NBS)-leucine-rich repeat (LRR), protein kinase, extracellular leucine-rich repeat transmembrane (eLRR-TM), LRR and toxin reductase (Dong *et al* 2001; Yi *et al* 2002; Hammond-Kosack *et al* 1997; Luo *et al* 2000). Among these, the NBS-LRR class occurs in large numbers in the plant genome. Depending on the presence or absence of a toll and interleukin-1 receptor (TIR) domain at the N-terminal domain of the protein, this class of *R*-genes can be subdivided into TIR and non-TIR classes. Non-TIR NBS-LRR proteins contain N-terminal coiled-coil (CC) or leucine zipper (LZ) domains, corresponding to helical structures that are thought to play an important role in interactions with other molecules in the signal transduction pathway. TIR and non-TIR *R*-genes can be anticipated on the basis of characteristic domains present in the NBS–LRR domain. For example, RNBS-A-TIR (LQKKLLSKLL) and RNBS-D-TIR (FLHIACFF) are found exclusively in the TIR subclass, whereas RNAS-A-non-TIR (FDLxAWCCVSQxF) and RNBS-D-non-TIR (CFLYCELPED) occur only in the non-TIR subclass. In addition, the single amino acid residue at the final position of the kinase-2 motif can also be used to distinguish the two classes (TIR or non-TIR): tryptophan (W) and aspartic acid (D) are present in the non-TIR and TIR-type proteins, respectively (Meyers *et al* 1999; Pan *et al* 2000; Kanazin *et al* 1996; Lee *et al* 1996; Wang *et al* 2005; Noir *et al* 2001; Bayram *et al* 2005; Martinez *et al* 2004; Xu *et al* 2005; Irigoyen *et al* 2004; Yaish *et al* 2004; Leeuwen *et al* 2005; Leister *et al* 1996). Although the NBS-LRR class *R*-genes contain many conserved domains, the whole sequence similarity among the NBS-LRR class of *R*-genes is so low that it is very difficult to detect by cross-hybridization.

The conserved domains of *R*-genes provide opportunities for designing degenerate primers and isolating resistance gene analogues (RGAs) by the polymerase chain reaction (PCR) strategy from plant genomes. At present, NBS-LRR type RGAs have been isolated from very diverse species of plants by this approach (Hinchliffe *et al* 2005; Deng *et al* 2000; Tu *et al* 2003; Yaish *et al* 2004).

A high mutation rate is present in many plant pathogens which makes individual *R*-genes ineffective. Therefore, the ability of plant species to survive over evolutionary time might depend on their ability to maintain and generate useful levels of diversity at resistance loci. Functionally diverse *R*-genes are often clustered in plant genomes. Previous studies have shown that DNA rearrangements play a key role in the evolution of *R*-gene loci, allowing plants to generate novel resistance to match the changing patterns of pathogen virulence. It was reported that point mutations and in-dels are also reasons for the diversity of *R*-genes (Meyers *et al* 1999; Pan *et al* 2000; Noir *et al* 2001). The evolution of *R*-genes remains largely unexplored, but useful information has recently been gained from molecular genetic analyses.

Sweet potato (*Ipomoea batatas* Lam.) is a member of the family Convolvulaceae, Genus *Ipomoea*, section *batatas*. It is the only hexaploid (2n=6x=90) in this section. Sweet potato is China's fourth most important food crop after rice, wheat and maize. It produces a stable yield under a wide range of environmental conditions and is cultivated in more than 100 countries. China accounts for roughly 80% of the world's production. Sweet potato is widely used as food and animal feed, and is processed into starch, liquor and a variety of other industrial products, but yields are strongly affected by various disease-causing organisms such as fungi, nematodes, viruses, etc. (FAO 2002). Although some molecular markers of disease resistance in sweet potato have been reported, the complex genetic background of sweet potato makes it difficult to associate molecular markers with disease-resistance genes. Since there are as yet no reports of the molecular cloning of putative resistance genes from sweet potato, the main aims of the present study were the isolation and characterization of RGAs from sweet potato and the assessment of their evolutionary relationship.

2. Materials and methods

2.1 *Plant material and DNA extract*

Sweet potato cultivar 'Qingnong no. 2', which is highly resistant to the knot nematode was used. DNA was extracted from young leaflets using the modified 2.0% (w/v) cetyl trimethyl ammonium bromide (CTAB) protocol, as described by Guo *et al* (2002).

2.2 *Primers and PCR conditions*

A large set of degenerate primers was designed according to the P-loop and GLPL motifs of the *R*-genes cloned. These primers are listed in table 1.

PCR was carried out in a total volume of 20 μl containing 50 ng of template DNA, 10 mM TRIS-HCl (pH 8. 3), 50 mM KCl, 2.0 mM $MgCl_2$, each dNTP at 100 μM, each primer at 0.3 μM, and 1.0 U of *Taq* DNA polymerase (TAKARA Bio Inc.). PCR was performed in an Eppendorf Mastercycler gradient thermal cycler. PCR cycling conditions were: 5 min at 94°C, 35 cycles (45 s at 94°C, 50 s at 55 °C, 1min 30 s at 72°C), 10 min at 72°C.

2.3 *Cloning and sequencing of PCR products*

PCR amplification products were separated by electrophoresis on 1.5% TAE–agarose gels and stained with

Table 1. Degenerate primers used in amplification

Primer name	Conserved amino acid domain	Sequence (5′ to 3′)
F1	P-loop (GGV/I/MGKTT)	GGHDYVGGKAARACWAC
F2	P-loop (GMGGV/I/SGKTT)	GGDATGVSVGGHDYVGGKAARAC
F3	P-loop (GGV/I/MGKTT)	GGNGGNRTHGGNAARACHAC
F4	P-loop (GGV/I/MGKTT)	GGNGTNGGNAARACNAC
F5	P-loop (GGV/I/MGKTT)	GGNGGNGTNGGNAANACNAC
F6	P-loop (GGV/I/MGKTT)	GGNGGNRTNGGNAARACNAC
R1	hydrophobic domain (GL/FPL/FA/VL)	AKWGCYARRGGDARYCC
R2	hydrophobic domain (GL/FPL/FA/VL)	GCMRCCARAGGMARYCC
R3	hydrophobic domain (GL/FPL/FA/VL)	AGNGCHAGNGGYAANCC
R4	hydrophobic domain (GL/FPL/FA/VL)	AANGCHAGNGGYAANCC
R5	hydrophobic domain (GL/FPL/FA/VL)	AGIGCHAGNGGNAGNCC
R6	hydrophobic domain (GL/FPL/FA/VL)	ARNWYYTTVARDGCVARWGGVARWCC

Fx, forward primer; Rx, reverse primer; I, deoxyinosine ; M, A/C; W, A/T; S, C/G; Y, C/T; K, G/T; V, A/C/G; H, A/C/T; D, A/G/T; B, C/G/T; R, A/G

ethidium bromide for visualization. Bands of the expected size (approximately 500 bp) were excised from the gel and purified using the PCR Fragment Recovery Kit (TAKARA Bio Inc.). The obtained DNA was cloned using the pMD-18T cloning kit (TAKARA Bio Inc.) and transformed into *Escherichia coli* DH5α. Recombinant plasmid DNA was extracted using alkaline lysis, and digested with double digestion to verify the presence of the expected insert.

Unique clones were determined using an ABI PRISM 377 DNA sequencer (TAKARA Bio Inc.). The nucleotide sequences of sweet potato RGAs have been deposited in the GenBank Database under Accession Nos DQ251184, DQ272296, DQ272297, DQ303440–DQ303442, DQ341401–DQ341403, DQ341405–DQ341 408, DQ377949–DQ377957.

2.4 *Sequence analysis*

Nucleotide sequence analysis and translation to the corresponding amino acid sequence were performed using DNAMAN software (version 6.0.) package (*www.lynnon.com*). Similarity searches were performed with the BLASTX and BLASTP programs on the website of the National Centre for Biotechnology Information (NCBI). Multiple alignments were performed with the ClustalW1.8 software (Thompson *et al* 1994) and DNAMAN software (version 6.0) packages. Phylogenetic and molecular evolutionary analyses were conducted using the MEGA software (version 3.1) (Kumar *et al* 2004). The bootstrap value was used to evaluate the reliability.

3. Results

3.1 *Isolation of RGAs from sweet potato by PCR using degenerate primers*

Thirty-six primer combinations (table 1) were used to amplify sweet potato genomic DNA. However, only three primer combinations (F2–R2, F3–R2, and F6–R1 generated a band about 500 bp in size from the sweet potato genomic DNA. The approximately 500 bp band from each primer combination was close to the fragment size expected based on the sequences of the *N*, *Gro-4* and *RPS2* genes; therefore, these bands were cloned. Restriction analyses of inserts of randomly selected clones showed that the 500 bp DNA band from all three primer combinations contained target fragments. Cloning and characterization of the amplicons obtained revealed that each band was made up of many RGA sequences. A total of 40 unique clones (10 from the F2–R2, 15 from the F3–R2 and 15 from F6–R1 combinations) were sequenced. In all cases, amplified bands were made up of many sequences of different sizes. Analysis of the 40 sequences using the BLAST algorithm demonstrated that 22 of the sequences (6 from the F2–R2, 8 from the F3–R2 and 8 from F6–R1 combinations; designated RGA-1 to RGA-22, respectively) had marked similarity to the NBS domains encoded by known plant *R*-genes or RGAs isolated from other species in the GenBank database. The other clones showed a similarity to retrotransposon sequences or gave no significant matches. Sequence analysis of the NBS clones showed that each of the 22 RGA sequences contained a single uninterrupted open reading frame (ORF). Their

Table 2. The percentage of similarity among nucleotide acid sequences and amino acid sequences of sweet potato RGAs

	RGA1	RGA9	RGA8	RGA11	RGA12	RGA19	RGA4	RGA18	RGA14	RGA17	RGA13	RGA5	RGA15	RGA16	RGA22	RGA20	RGA21	RGA3	RGA6	RGA7	RGA10	RGA2
RGA1		99.3	98.7	99.3	63.2	65.4	65.8	65.4	66.0	67.3	62.1	67.5	66.4	58.9	57.9	47.0	37.3	28.5	26.6	21.3	20.0	23.1
RGA9	99.0		99.4	100.0	61.0	65.9	63.0	63.0	65.8	65.9	60.7	65.1	65.7	59.4	56.1	44.7	35.5	26.5	24.4	19.4	18.2	23.1
RGA8	98.8	99.0		99.4	60.5	65.3	62.4	62.4	65.2	65.3	60.1	64.5	65.1	58.8	55.6	45.3	34.9	25.9	23.8	19.4	18.2	23.8
RGA11	98.7	99.2	99.4		61.0	65.9	63.0	63.0	65.8	65.9	60.7	65.1	65.7	59.4	56.1	44.7	35.5	26.5	24.4	19.4	18.2	23.1
RGA12	74.5	74.9	74.7	75.3		75.0	72.1	72.1	69.4	71.5	71.5	71.3	74.4	66.5	56.5	45.6	45.6	28.0	27.0	20.8	17.6	23.3
RGA19	76.2	76.2	76.6	76.8	82.4		73.4	71.1	72.8	72.8	71.1	70.3	73.8	67.3	57.3	47.1	41.4	26.5	25.0	20.6	18.2	24.4
RGA4	76.3	75.8	75.8	76.0	80.3	81.0		93.6	84.2	85.0	83.2	73.6	71.5	66.1	57.3	45.9	40.8	28.0	27.2	21.2	14.9	25.9
RGA18	74.1	74.1	73.9	74.1	79.3	80.2	95.6		82.9	84.4	82.7	73.8	71.5	63.6	59.1	46.5	40.2	25.9	25.0	20.6	14.9	24.4
RGA14	77.1	77.3	77.5	77.7	79.7	82.8	90.3	90.1		97.5	90.5	78.3	74.5	64.2	59.2	49.4	43.2	25.2	24.7	22.6	17.1	24.0
RGA17	76.4	76.8	77.2	77.2	80.3	82.1	90.6	90.6	97.9		92.5	77.9	75.0	64.8	57.9	47.1	40.2	24.7	23.8	21.2	16.2	23.1
RGA13	74.9	75.4	75.2	75.4	80.5	79.7	90.0	88.9	91.6	93.1		75.0	72.7	62.4	56.1	44.1	39.1	24.1	23.1	21.2	16.2	23.8
RGA5	76.5	76.1	76.3	76.4	81.0	79.3	82.8	81.3	84.1	83.6	81.9		82.5	63.4	56.5	45.6	39.3	23.3	22.4	19.5	14.3	22.4
RGA15	77.8	78.0	78.0	78.0	83.0	82.6	81.5	81.3	82.2	82.2	79.9	89.5		65.2	59.4	46.7	37.5	26.1	25.2	20.1	15.4	23.3
RGA16	72.2	72.6	72.6	72.8	76.3	77.3	75.7	75.5	74.7	75.1	72.6	73.9	75.5		57.0	40.9	38.0	27.5	26.6	17.6	15.0	22.0
RGA22	68.9	68.7	68.9	68.9	68.6	70.9	69.3	70.1	69.8	69.3	67.6	67.8	71.1	70.0		43.8	37.5	23.6	22.6	21.1	14.9	25.6
RGA20	55.8	55.6	56.2	56.0	56.7	57.1	57.9	57.9	60.8	58.5	56.6	57.9	58.7	54.1	58.8		48.5	25.5	24.5	18.2	14.4	24.5
RGA21	52.6	52.6	52.8	53.0	51.9	52.0	52.6	52.0	55.2	53.9	53.0	52.9	52.1	52.8	52.1	63.8		23.0	21.4	17.7	18.8	22.6
RGA3	44.7	43.9	43.9	43.7	40.8	41.2	42.6	41.0	42.8	41.7	42.1	42.9	41.6	40.7	40.3	41.1	39.0		93.8	23.5	22.3	27.2
RGA6	43.8	43.0	43.0	42.8	40.5	40.6	42.5	41.4	42.2	41.6	42.0	42.5	41.3	40.3	40.3	40.4	38.8	97.0		23.1	20.0	25.0
RGA7	39.7	39.5	39.7	39.7	39.5	38.3	38.9	38.3	39.2	39.7	39.5	37.9	39.1	38.1	37.7	37.9	39.0	51.3	52.1		30.8	21.7
RGA10	42.4	42.4	42.6	42.8	41.2	39.3	39.5	37.7	37.9	38.5	38.7	40.2	38.9	40.3	37.3	37.7	37.2	52.3	52.6	57.2		23.1
RGA2	44.0	43.3	43.5	43.3	46.1	43.3	48.3	45.8	47.3	48.5	47.4	48.4	45.2	45.7	43.9	43.5	42.9	44.9	44.7	44.3	44.5	

Note: The percentage of similarity among amino acid sequences and nucleotide acid sequences are given above and below the diagonal, respectively.

deduced amino acid sequences showed the internal motifs characteristic of the NBS-LRR gene class. Kinase-2 and kinase-3a motifs were present in all NBS sequences from sweet potato, although in some sequences there were minor amino acid substitutions. No putative splicing sites were detected in these clones.

3.2 *Diversity analysis of sweet potato RGAs*

The diversity of the NBS-type RGAs isolated from sweet potato was analysed, and 22 nucleotide sequences obtained using three degenerate primer combinations were included in the analysis. Sequence identity among the 22 RGA nucleotide sequences ranged from 37.2% to 99.4%, while the identity of the deduced amino acid sequences from 22 RGAs ranged from 14.3% to 100% (table 2).

A neighbour-joining phylogenetic tree based on these RGA nucleotide acid sequences is shown in figure 1. The 22 sweet potato RGA sequences were grouped into five clusters or families (A–E), all of which were highly supported by bootstrap values. Cluster A was composed of 15 members, while the other clusters consisted of only one or two members (figure 1).

The ratio of non-synonymous to synonymous substitutions (K_A/K_S ratio) reveals whether sequence polymorphism supports diversifying ($K_A/K_S>1$) or purifying ($K_A/K_S<1$) selection. The K_A/K_S ratio among five distinct clusters (A, B, C, D, E) of triplets encoding amino acid sequences in sweet potato ranged from 0.19 to 0.77.

3.3 *Comparative analysis of sweet potato RGAs and cloned R-genes*

To establish the origin and elucidate the evolution of sweet potato RGAs, phylogenetic relationships between amino acid sequences deduced for sweet potato RGAs and cloned *R*-gene products from other species were analysed. The analysis was conducted using the NBS region sequences of the following *R* genes in the GenBank database: *N* (U15606) and *M* (U73916) from tobacco, *Mi-1.2* (AF039682) and *Prf* (U65391) from tomato, *Gpa2* (AF195939) and *Gro-4* (AY196151) from potato, *L11* (AF093641) from flax, *RPP8* (AF089710) and *RPM1* (AAF27008) from *Arabidopsis*. A multiple alignment of amino acid sequences deduced from 22 sweet potato RGAs and NBS region sequences of the *R*-genes mentioned above is shown in figure 2. From the alignment of amino acid sequences, a neighbour-joining

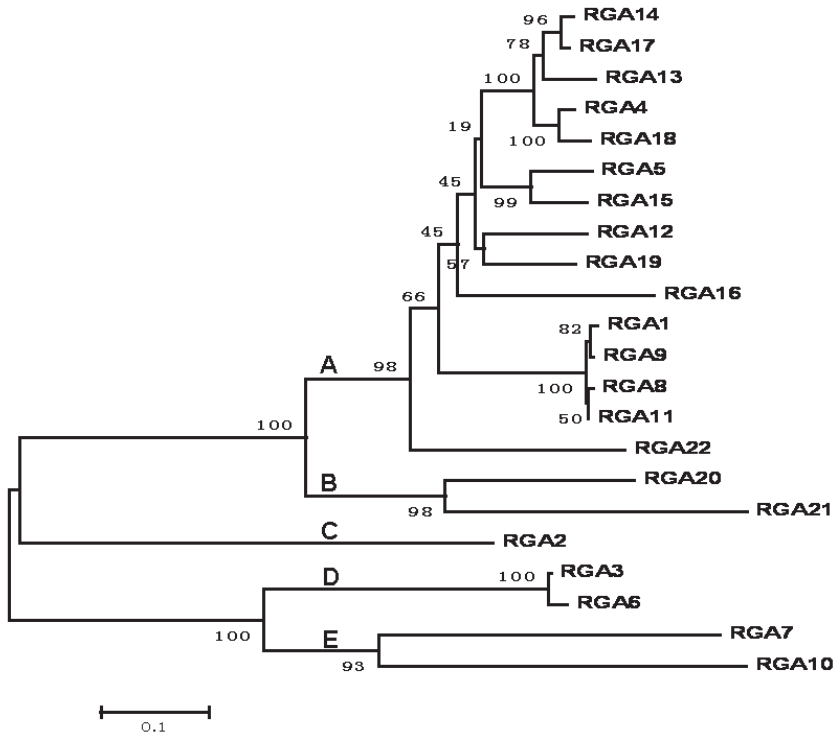

Figure 1. Neighbour-joining phylogenetic tree for RGA nucleotide sequences isolated from sweet potato. Branch labels from A to E correspond to families of closely related RGA sequences. The numbers on the branches indicate bootstrap values (1000 iterations).

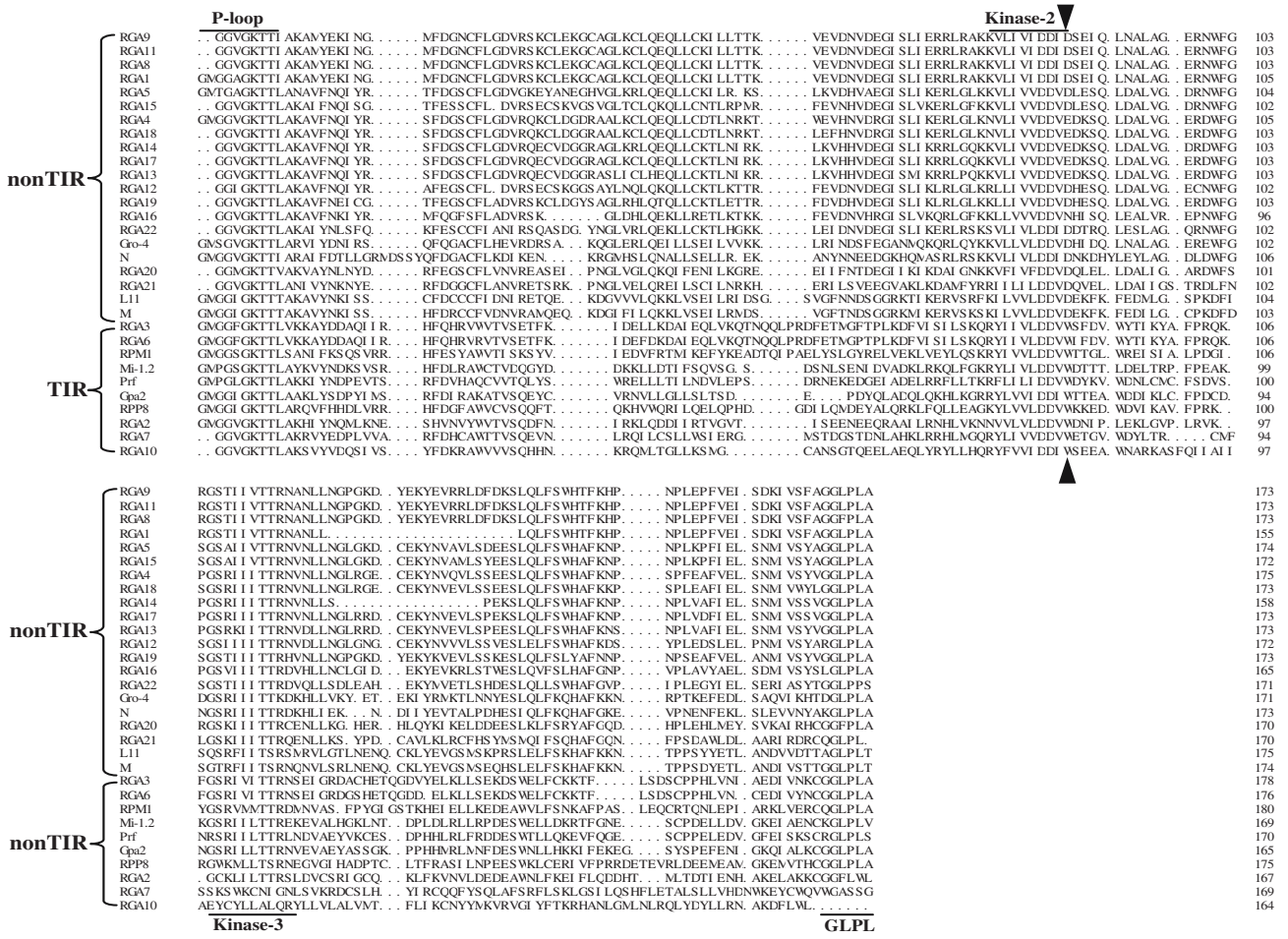

Figure 2. Multiple alignment of amino acid sequences of sweet potato RGAs with the NBS domains of other plant-resistance genes. The positions of the aspartic acid (D) and tryptophan (W) residues characteristic of TIR and non-TIR sequences in the kinase-2, respectively, are indicated with an arrow head. Mutual domains are denoted by the line below. The alignment was constructed with DNAMAN 6.0.

phylogenetic tree was constructed which separates the five families into two distinct groups (figure 3). The first group contains RGA2, RGA3, RGA6, RGA7, RGA10 and non-TIR R proteins RPP8, RPM1, Mi-1.2, Prf and Gpa2. In addition, the kinase-2 motif (LLVLDDVX) in the non-TIR-type sequences has the characteristic amino acid residue tryptophan (LLVLDDVW) which was absent in TIR-NBS-type sequences (Meyers *et al* 1999). The second group includes RGA1, RGA4, RGA5, RGA9, RGA11-RGA22 and other known TIR-NBS-LRR proteins Gro-4, N, L11, M. Most of these have aspartic acid (D) as the final amino acid in the kinase-2 motif, corresponding to TIR-NBS sequences. In the deduced amino acid sequences of sweet potato RGAs, P-loop, kinase-2, kinase-3a and GLPL motifs were found.

In addition to the phylogenetic analysis, searches in the GenBank database using the BLAST algorithm confirmed that sweet potato RGA sequences are closely related to the NBS regions of *R*-genes cloned in other species.

4. Discussion

4.1 *Isolation of sweet potato NBS-encoding RGAs by different PCR strategies*

By using degenerate primer combinations (F2–R2, F3–R2, and F6–R1) from NBS-conserved motifs of cloned plant disease-resistance genes, five distinct families of RGAs representing a wide range of diversity were isolated from the sweet potato genome. In addition to the NBS sequences found in this study, the amplification of retrotransposon-related sequences is likely to result from the high copy number of such sequences in plant genomes. For example, previous research showed that retrotransposons constitute at

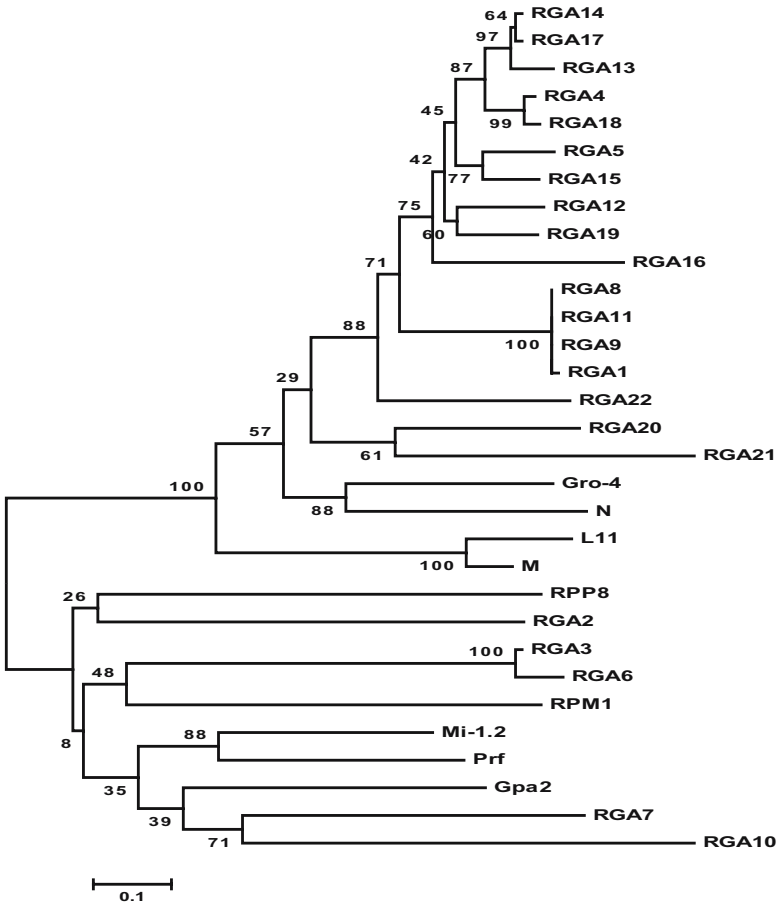

Figure 3. Neighbour-joining tree based on alignment of amino acid sequences of sweet potato RGAs and the NBS region of isolated disease-resistance genes. The numbers on the branches indicate bootstrap values (1000 iterations).

least 50% of the maize genome. The PCR-derived sequences were identified as RGAs or fragments of *R*-genes based on conserved motifs characteristic of the NBS-type *R*-gene. A single uninterrupted ORF and conserved motifs such as P-loop, kinase-2, kinase-3a, GLPL were found among 22 sweet potato RGA sequences. Most of the sweet potato RGAs were also closely related to the NBS sequences of *Arabidopsis*.*RPP8* and *RPM1*, tomato *Mi-1.2,* and tobacco *N*. Thus, some of them may encode resistance gene products of unknown specificity.

NBS sequences are abundant in plant genomes. Genomic DNA sequence analyses reveal that there are about 660 putative NBS-LRR type disease-resistance genes in the rice (*Oryza sativa*) genome, and about 200 in the *Arabidopsis* genome (Meyers *et al* 1999; Han *et al* 2005) and our results reveal that a number of NBS-type sequences are also present in the sweet potato genome. Whether these sequences originate from different alleles spread over the hexaploid genome or from paralogous genes in each haploid genome set is not known. Future research will be directed towards elucidating this issue.

4.2 *Diversity and evolutionary relationship of sweet potato NBS-encoding RGAs*

RGAs encoding an NBS motif have been successfully isolated from a broad range of plant species. To the best of our knowledge, results from this research provide the first view of the complexity and diversity of RGAs in sweet potato. Sequence identity among the 22 RGA nucleotide sequences ranged from 37.2% to 99.4%, with the highest value for RGA11 and RGA8, and the lowest for RGA10 and RGA21. It is likely that RGA11 and RGA8 might have arisen from a recent duplication of a common ancestor gene, and that RGA10 and RGA21 may represent divergent origins. It is believed that duplicated RGA sequences in a given plant genome are probably subject to unequal crossing-over

events for generation of diversity. However, so far, clear evidence is lacking for the genetic events or mechanisms responsible for the complexity and diversity of RGAs in the plant genome. The high degree of similarity between sweet potato RGAs and NBS sequences derived from *R*-genes cloned from tomato, tobacco, flax and potato suggest an ancestral relationship. Furthermore, all of the sweet potato RGAs isolated were grouped into two types: TIR-NBS-LRR and non-TIR-NBS-LRR. Seventeen of the 22 sweet potato RGAs belong to the TIR-NBS-LRR type. This also supports the view that TIR-NBS-LRR and non-TIR-NBS-LRR *R*-genes occur in dicot species (Meyers *et al* 1999).

Molecular data are increasingly consistent with the idea that plants have *R*-genes arrayed in complex clusters. Indeed, clustering of *R*-genes and homologous sequences may facilitate the generation of diversity and new resistance specificities (Noir *et al* 2001) The source of genetic variation within *R*-genes has been the subject of much discussion (Meyers *et al* 1999; Pan *et al* 2000; Kanazin *et al* 1996; Lee *et al* 1996; Wang *et al* 2005; Noir *et al* 2001; Bayram *et al* 2005; Martinez *et al* 2004; Xu *et al* 2005; Irigoyen *et al* 2004; Yaish *et al* 2004; Leeuwen *et al* 2005; Leister *et al* 1996; Kumar *et al* 2004). Two main important genetic mechanisms including a slowly evolving divergence hypothesis and a rapidly evolving process have been proposed to account for the evolution of *R*-genes (Xu *et al* 2005). The results also provide clues to the putative mechanisms that may have contributed to the evolution of *R*-genes in sweet potato. Point mutations, insertions or deletions in the regions between the conserved motifs were observed and these types of genetic variations might constitute the primary source for RGA cluster divergence in sweet potato. Taking these into consideration, the evolution of NBS-type genes in sweet potato appears to favour a slowly evolving divergence mechanism rather than a rapidly evolving one (Xu *et al* 2005).

Analyses of the ratio of synonymous and non-synonymous nucleotide acid substitutions (K_a/K_s) (i.e. 0.19–0.77) for all sweet potato RGA families show that these are similar to NBS domains from *R*-genes and suggest that the sweet potato NBS domain is subject to purifying rather than diversifying selection, which is consistent with earlier proposals (Meyers *et al* 1999; Noir *et al* 2001; Xu *et al* 2005).

This is the first study aimed at analysing the diversity and evolutionary relationship of RGAs within the genome of sweet potato to facilitate an understanding of the role of potential *R*-genes, and to isolate functional *R*-genes from sweet potato.

Acknowledgements

This research was supported by the Natural Science Foundation of Fujian Province of China (No.2006J0059), National Natural Science Foundation of China (No.30370900), and Key Technologies R&D Program of Fujian Province of China (No.2003N001).

References

Bayram Yuksel, Estill JC, Schulze S R and Paterson A H 2005 Organization and evolution of resistance gene analogs in peanut; *Mol. Gen. Genomics* **274** 248–263

Deng Z, Huang S, Ling P, Chen C, Yu C, Webber C A, Moore G A and Gmitter F G 2000 Cloning and characterization of NBS-LRR class resistance-gene candidate sequence in citrus; *Theor. Appl. Genet.* **101** 814–822

Dong J X, Dong H T and Li D B 2001 Recent advances in plant disease resistance genes; *Acta Phytopathol. Sinica* **31** 1–9

Food and Agriculture Organization 2002 *Production year book 2002*, *vol. 54* (Rome: Food and Agriculture Organization of the United Nations)

Guo J P and Pan D R 2002 PCR Detection of nematode resistance in sweet potato; *Acta Agronomica Sinica* **28** 167–169

Hammond-Kosack K E and Jones J D G 1997 Plant disease resistance gene; *Annu. Rev. Plant Mol. Biol.* **48** 575–607

Han D J, Cao L, Chen Y F and Li Z Q 2005 Molecular basis of interaction between disease resistance gene and avirulence gene; *Acta Genet. Sinica* **32** 1319–1326

Hinchliffe D J, Lu Y Z, Potenza C, Segupta-Gopalan C and Cantrell R G 2005 Resistance gene analogue markers are mapped to homologous chromosomes in cultivated tetraploid cotton; *Theor. Appl. Genet.* **110** 1074–1085

Irigoyen M L, Loarce Y, Fominaya A and Ferrer E 2004 Isolation and mapping of resistance gene analogs from *Avena strigosa* genome; *Theor. Appl. Genet.* **109** 713–724

Kanazin V, Marek L F and Shoemaker R C 1996 Resistance gene analogs are conserved and clustered in soybean; *Proc. Natl. Acad. Sci. USA* **93** 11746–11750

Kumar S, Tamura K and Nei M 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment; *Brief. Bioinform.* **5** 150–163

Lee S Y, Seo J S, Rodriguez-Lanetty M and Lee D H 2003 Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species; *Mol. Gen. Genomics* **269** 101–108

Leeuwen H V, Garcia-Mas J, Coca M, Puigdomenech P and Monfort A 2005 Analysis of the melon genome in regions encompassing TIR-NBS-LRR resistance genes; *Mol. Gen. Genomics* **273** 240–251

Leister D, Ballvora A, Salamini F and 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

Luo M, Zhu Y L, Yu C, Wu Z W and Zou Y 2000 Cloning of the plant resistance genes and their structure and function; *Hereditas* **22** 429–433

Martinez Zamora M G, Castagnaro A P and Diaz Ricci J C 2004 Isolation and diversity analysis of resistance gene analogues (RGAs) from cultivated and wild strawberries; *Mol. Gen. Genomics* **272** 480–487

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

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

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

Thompson J D, Higgins D G and Gibson T J 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice; *Nucleic Acids Res.* **22** 4673–4680

Tu L L, Zhang X L, Zhu L F, Nie Y C and Guo X P 2003 Origin, diversity and evolution of NBS-type disease-resistance gene analogues in sea-island cotton; *Acta Genet. Sinica* **30** 1071–1077

Wang S Q, Zhang D C, Li P, Wang X D, Li S G, Zhu L H and Zhai W X 2005 Cloning and analysis of a new NBS-LRR resistance gene family in rice; *Acta Genet. Sinica* **32** 704–711

Xu Q, Wen X P and Deng X X 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

Yaish M W F, Saenz de Miera L E and Perez de la Vega M 2004 Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (NBS) type from *Lens* species; *Genome* **47** 650–659

Yi T Y, Xie B Y, Zhang B X and Gao B D 2002 Application of plant resistance gene analogs in cloning and mapping resistance genes; *Biotechnol. Bull.* **2** 16–20

Yu Y G, Buss G R, Saghai and Maroof M A 1996 Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site; *Proc. Natl. Acad. Sci. USA* **93** 11751–11756

MS received 6 June 2006; accepted 3 January 2007

*e*Publication: 4 April 2007

Corresponding editor: Imran Siddiqi