IDENTIFICATION AND CHARACTERIZATION OF A RESISTANCE GENE ANALOG (RGA) FROM THE Caricaceae DUMORT FAMILY

PAULO DE PAIVA ROSA AMARAL1, PAULO CÉSAR MARTINS ALVES2, NATÁLIA FLORÊNCIO MARTINS3, FELIPE RODRIGUES DA SILVA4, GUY DE CAPDEVILLE5, MANOEL TEIXEIRA SOUZA JÚNIOR6.

ABSTRACT - The majority of cloned resistance (R) genes characterized so far contain a nucleotide-binding site (NBS) and a leucine-rich repeat (LRR) domain, where highly conserved motifs are found. Resistance genes analogs (RGAs) are genetic markers obtained by a PCR-based strategy using degenerated oligonucleotide primers drawn from these highly conserved “motifs”. This strategy has the advantage of the high degree of structural and amino acid sequence conservation that is observed in R genes. The objective of the present study was to search for RGAs in Carica papaya L. and Vasconcellea cauliflora Jacq. A. DC. Out of three combinations of primers tested, only one resulted in amplification. The amplified product was cloned in pCR2.1TOPO and than sequenced using M13 forward and reverse primers. Forty-eight clones were sequenced from each species. The 96 sequences generated for each species were cleaned of vector sequences and clustered using CAP3 assembler. From the GENEBANK, one RGA was identified in the sequence of Carica papaya. DNA fragments of 589 and 1127 bp covering 1.2 and 2.4 kb of the genomic sequence, respectively, were amplified using degenerate primers designed from conserved motifs in the NBS and LRR sites. These fragments were cloned in pCR2.1TOPO and sequenced. A putative disease resistant protein RGA3 (Solanum bulbocastanum). To the extent of our knowledge this is the first report of a RGA in the Caricaceae Dumort family. Preliminary structural studies were performed to further characterize this putative NBS-LRR type protein. Efforts to search for other RGAs in papaya should continue, mostly to provide basis for the development of transgenic papaya with resistance to diseases.

Index Terms: Caricaceae, nucleotide-binding site (NBS), V. cauliflora, C. papaya, papaya.

IDENTIFICAÇÃO E CARACTERIZAÇÃO DE UM ANÁLOGO DE GENE DE RESISTÊNCIA (AGR) DA FAMÍLIA DE CARICACEAE DUMORT.

RESUMO – A maioria dos genes de resistência (R) clonados e caracterizados até o momento contém domínios NBS (nucleotide binding site) e LRR (leucine-rich repeat). Destes domínios, encontram-se “motifs” altamente conservados. Análogos de genes de resistência (RGAs) são marcadores genéticos obtidos por uma estratégia, baseada em PCR, que usa primers degenerados desenhados a partir desses “motifs” altamente conservados dos genes R. Esta estratégia possui a vantagem do elevado grau de conservação da estrutura e sequência dos aminóacidos observados nos genes R. O objetivo do presente estudo foi realizar uma busca por RGAs em Carica papaya L. e Vasconcellea cauliflora Jacq. A. DC. De três combinações de primers avaliadas, somente uma obteve sucesso na amplificação. O produto da amplificação foi então clonado em pCR2.1TOPO e sequenciado utilizando os primers universais M13 forward e reverse. Quarenta e oito clones foram sequenciados de cada espécie vegetal. As 96 sequências geradas para cada espécie, retiraram-se as sequências do vetor e, em seguida, as mesmas foram agrupadas utilizando o programa “CAP3 assembler”. A partir do GENEBANK, foi identificado um RGA em C. papaya apresentando um BlastX e-value de 2x10-61 com o “gb|AAP45165.1| putative disease resistant protein RGA3 (Solanum bulbocastanum)”. Na extensão do nosso conhecimento, este é o primeiro relato de um RGA na família Caricaceae Dumort. Estudos preliminares de estrutura foram realizados visando à maior caracterização deste potencial “NBS-LRR type protein”. Esforços para encontrar novos análogos de gen de resistência devem continuar, principalmente para fornecer bases para o desenvolvimento de plantas de mamão transgênicas com resistência a doenças.

Termos para indexação: Caricaceae, sítios de ligação de nucleotídeos (SLN), V. cauliflora, C. papaya, mamão.

INTRODUCTION

Carica papaya L. is the most well-known and cultivated species in the Caricaceae Dumort family. Cylicomorpha, Jacarataia, Horovitzia, Vasconcellea and Jarilla are the other five genera of this family. The Carica genus contains only one species, C. papaya, while the genera Vasconcellea contains several species, which were originally classified as belonging to the Carica genus, including Vasconcellea cauliflora Jacq. A. DC. This species, known as “tapaculo”, “papayo de montaña” or “zonzapote” is found from the South of Mexico to the North of South America, as well as in Trinidad. Such species is a well-known source of natural resistance to Papaya Ringspot Virus (PRSV), the main virus attacking papaya worldwide (Badillo, 1993, 2000; 2001; Manshardt & Wenslaff, 1989).

The majority of the resistance genes (R genes) cloned and sequenced until now are part of the nucleotide binding site -leucine-rich repeat (NBS-LRR) gene family (Rommens & Kishore, 2000). The NBS-LRR gene products are generally composed of three main domains: a) a variable N-terminal domain of approximately 200 amino acids; b) a NBS domain of 300 amino acids, and c) a more variable tandem array of approximately 10 to 40 short LRR (leucine-rich-repeat) motifs (Cannon et al., 2002). The NBS domain is believed to participate in signal transduction, while the LRR domain is thought to be involved in ligand binding and pathogen recognition (Young, 2000). P-loop, RNBS-A, kinase 2, RNBS-B, RNBS-C, GLPL, and RNBS-D are also highly conserved motifs generally present in the NBS domain of the R genes (Lee et al., 2003).

A new PCR-based strategy, using degenerated primers designed from these conserved motifs, has resulted in the isolation of numerous resistance gene analogs (RGAs) from a variety of plant species such as potato (Leister et al., 1996), bean (Ferrier-Cana et al., 2003), rice (Leister & Katagiri, 2000) and several others (for review see Chelkowski & Koczky, 2003). Once found, a series of uses can be assigned to this type of marker: a) as a probe to screen BAC or...
cDNA libraries, in the process of searching for R genes; b) as a marker to be applied in marker assisted selection; and c) to obtain resistance by its overexpression in the plant genome. The objective of the present study was to search for RGAs in *C. papaya* and *V. cauliflora* using primers designed from the P-loop and RNBS-D motifs. To the best of our knowledge, no RGA has so far been described for *Caricaceae*.

**MATERIAL AND METHODS**

Plant genomic DNA from the transgenic papaya Embrapa PTP18 (Souza Jr. et al., 2005) and from *V. cauliflora* plants was extracted as described in Souza Jr. et al. (2005). The DNA was quantified and stored at -20°C until use.

PCR was performed in a total volume of 30 µl containing 5.0 µl of DNA (10 ng/µl), 0.25 µl dNTPs (25 mM), 3.0 µl 10X Taq DNA Polymerase Buffer, 1.2 µl MgCl₂ (50 mM), 3.0 µl of each primer (5 pmol/µl), and 0.25 µl of Recombinant Taq DNA Polymerase (5 µl) (Invitrogen Life Technologies). Amplification of NBS domains were performed using the following sense primers: a) P1a (5'-ggiatgcciggiiiiggiaaraciac-3'), b) P1b (5'-ggiatgggiggiiiiggiaaraciac-3') (Bertioli et al., 2003), and c) LM638 (5'-ggiggigtiggiaaiaciac-3') (Kanazin et al., 1996), all of them designed from the consensus motif P-loop. The antisense primer used was RNBS-D (5'-ggraaiarishrcartaiviraarc-3') (Peñuela et al., 2002). Cycling conditions consisted of a 3 min initial denaturation step at 96°C followed by 30 cycles at 96°C for 1 min, a 1 min annealing at 40°C and a 1 min elongation at 72°C. Ten microliters of the reaction mix were loaded onto a 1% agarose gel to visualize the PCR products.

DNA products from the PCR reaction were cloned into the pCR2.1-TOPO vector (TOPO TA Cloning Kit - Invitrogen Life Technologies). Forty-eight white TOP10 *E. coli* colonies were randomly selected per species, and sequenced using M13 forward and reverse primers at the Embrapa Genetic Resources and Biotechnology DNA Sequencing Platform (http://www.cenargen.embrapa.br/laboratorios/laboratorios.html#dna).

The software PHRED (Ewing et al., 1998) was used to base call and to estimate error probability in the 192 chromatograms. After trimming (Telles & da Silva, 2001) to remove off artifacts, low quality sequences, vector and primer regions, the 123 remaining sequences were clustered using the CAP3 assembler (Huang & Madan, 1999). BLAST (Altschul et al., 1997) was used to identify similarities between the resulting 76 clusters (43 singlets and 33 contigs) and sequences in the NCBI’s nr database (Benson et al., 2002). Sequence alignment was performed using CLUSTALW (Thompson et al., 2000). To further characterize the translated protein we submitted the amino acid sequence to a secondary structure prediction server. The PSIPRED method was used (Jones, 1999). The preliminary homology molecular modeling was performed in the Meta-Server from CBS-CNRS site (http://bioserv.cbs.cnr.fr/HTML_BIO/tito.html).

**RESULTS AND DISCUSSION**

The primer combination P1b and RNBS-D was the only one that successfully amplified DNA. The profiles generated after running on an agarose gel were different for the two *Carica* species, with the number of strong and well-defined DNA bands being higher in *V. cauliflora*. DNA bands of expected size (~700 bp) were observed in both species (Figure 1).

Multiple alignment of the *C. papaya* sequence with retrieved RGAs demonstrates that all of them share the same conserved sequences described in other plant resistance genes (Pan et al., 2000). The alignment is shown in Figure 2. This finding indicates that in *C. papaya* might have members of the family of NBS-LRR disease resistance genes. In particular interest for comparison are sequences conserved in plants NBS region. This region is a tripartite...
conserved motif considered to be involved in the nucleotide binding (Holt III et al., 2003). Between the NBS and the LRR regions there is a hydrophobic domain (RNBS), whose three motifs (RNBS-A, B and C) are conserved in the majority of the R genes of the NBS-LRR class (Chelkowski & Koczyn, 2003). The presence of two conserved phenylalanine residues, separated by four amino acids, comprises the kinase 2 motif, which is a characteristic feature of NBS-LRR proteins (Mago et al., 2002). In the *C. papaya* sequence the majority of conserved feature were identified. At the N-terminus of the NBS domain, the conserved region denominated RNBS-A motif, may play a signal transduction function, and the consensus G-X-X-G-X-G-K-T-T appears as the P-loop or kinase 1 motif with an essential function for the orientation of the phosphate group (Moffett et al., 2002). Other important features are the hydrophobic residues usually present in RNBS-B and C, which are also shown in the *C. papaya* sequence. In addition, the consensus amino acid domain GLPL is present and it is functionally associated with structural stability of domains adjacent to the NBS complex in protein-protein interactions (Shirasu & Schulze-Lefert, 2003).

Intracellular R proteins can be divided into subfamilies with members that have either a coiled-coiled (CC) structure or a motif TIR. The TIR sequences contain domains in their amino terminus very similar to the Drosophila Toll or human interleukin receptor-like (TIR) region (Meyers et al., 2002). The TIR domain seems to be found only in dicotyledonous plans, whereas CC are found in monocots and dicots (Pan et al., 2000). For the TIR classification the characteristic amino acid is the Aspartic acid (D) at the end of the kinase-2 domain (position 91 - underlined in Figure 2). When there is a replacement for tryptophan (W), the sequence can be classified as non-TIR group. Consequently, as the *C. papaya* candidate shows the W residue at position 91, it should be classified as a non-TIR resistance gene analog.

A phylogenic tree (Figure 3) of RGAs from *C. papaya* and nine other plant species, based on alignment of amino acid sequences using CLUSTALW, shows an evolutionary proximity of the papaya sequence with RGAs from *Solanum bulbocastanum* (AAP45165), *Vitis vinifera* (AAM21291), and *Glycine max* (AAL50031).

![FIGURE 3](image_url) - Phylogenic tree based on alignment of amino acid sequences of RGAs from *Carica papaya* (bankit558486), *Vitis vinifera* (AAM21291), *Oryza sativa* (BAC45162), *Hordeum vulgare* (AAQ16121), *Glycine max* (AAL50031), *Solanum bulbocastanum* (AAP45165), *Arabidopsis thaliana* (Q9LRR4), *Coffea arabica* (CAC82607), *Gossypium barbadense* (AAP93894) and *Malus baccata* (AAM77271) using CLUSTALW.
Pairwise comparisons of the NBS candidate with the Protein DataBank showed the highest similarity between the GLPL sequence domain and (1Ko9) 8-oxoguanine DNA glycosylase (42% identity). The proposed model for the common core domain is shown in the Figure 4B. The common core GLPL structure is organized by a α-strand followed by a coil. The hydrogen bond network is organized and the portion of the structure seems to be stable. Preliminary structural evaluation indicates that the sidechains of the conserved leucine residues might have a similar tridimensional structure to leucine-zipper domains and the binding helices might have the same motif. To determine the precise three dimensional structure it would be necessary to investigate other templates in the NBS region in order to stabilish further structural features.

CONCLUSIONS

At the extent of our knowledge, no partial or complete sequence of an NBS-LRR type protein has so far been described for any member of the Carica genus. The sequence described here is restricted to the NBS domain, and is the first one ever described for C. papaya. At our lab new degenerated primers were designed from conserved motifs present in the NBS, as well as in the N-terminal and LRR domains of resistance genes described in the literature and at the NCBI database. Seven new primers combinations were tested and all of them were successful in amplifying DNA from C. papaya and V. cauliflora (data not shown). Cloning and sequencing of these PCR products are in progress, and new RGAs are expected to be identified from this work. Efforts to search for other RGAs in papaya should continue, mostly to provide basis for the development of transgenic papaya with resistance to diseases.

ACKNOWLEDGMENTS

We wish to thank Drs. David and Soraya Bertioli for providing the degenerated primers, as well as Dr. Dulcinea Fogel for providing the V. cauliflora population used in this study.

LITERATURE


