## IDENTIFICATION AND CHARACTERIZATION OF A RESISTANCE GENE ANALOG (RGA) FROM THE *Caricaceae* DUMORT FAMILY<sup>1</sup>

# PAULO DE PAIVA ROSA AMARAL<sup>2</sup>, PAULO CÉSAR MARTINS ALVES<sup>3</sup>, NATÁLIA FLORÊNCIO MARTINS<sup>4</sup>, FELIPE RODRIGUES DA SILVA<sup>5</sup>, GUY DE CAPDEVILLE<sup>6</sup>, MANOEL TEIXEIRA SOUZA JÚNIOR<sup>7\*</sup>.

**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 *C. papaya* showing a BlastX e-value of 2x10<sup>-61</sup> to the gb|AAP45165.1| 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). Dentro 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 seqüência dos aminoácidos 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 seqüenciado utilizando os primers universais M13 forward e reverse. Quarenta e oito clones foram seqüenciados de cada espécie vegetal. Das 96 seqüências geradas para cada espécie, retiraram-se as seqüências 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<sup>-61</sup> 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 genes 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, Jacaratia, 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 e 2001; Manshardt & Wenslaff, 1989).

The majority of the resistance genes (R genes) cloned and sequenced until now are part of the nucleotide binding site-leucinerich 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-richrepeat) 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 & Koczyk, 2003). Once found, a series of uses can be assigned to this type of marker: a) as a probe to screen BAC or

<sup>&</sup>lt;sup>1</sup> (Trabalho 211-2005). Recebido: 21-12-2006. Aceito para publicação: 24-07-2006.

<sup>&</sup>lt;sup>2</sup> Aluno de Graduação, Universidade de Brasília, ppamaral@iq.usp.br

<sup>&</sup>lt;sup>3</sup> Aluno de Graduação, Universidade de Brasília, cesar\_imperio@yahoo.com.br

<sup>&</sup>lt;sup>4</sup> Pesquisador III, Embrapa Recursos Genéticos e Biotecnologia, natalia@cenargen.embrapa.br

<sup>&</sup>lt;sup>5</sup>Pesquisador III, Embrapa Recursos Genéticos e Biotecnologia, felipe@cenargen.embrapa.br

<sup>&</sup>lt;sup>6</sup>Pesquisador III, Embrapa Recursos Genéticos e Biotecnologia, guy@cenargen.embrapa.br

<sup>&</sup>lt;sup>7</sup>Pesquisador III, Embrapa Recursos Genéticos e Biotecnologia, msouza@cenargen.embrapa.br \* Embrapa Genetic Resources and Biotechnology, CP 02372, 70.770-900, Brasilia, DF, Brazil. \*Corresponding author (msouza@cenargen.embrapa.br)

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  $\mu$ l containing 5,0  $\mu$ l of DNA (10 ng/ $\mu$ l), 0,25  $\mu$ l dNTPs (25 mM), 3,0  $\mu$ l 10X *Taq* DNA Polymerase Buffer, 1,2  $\mu$ l MgCl<sub>2</sub> (50 mM), 3,0  $\mu$ l of each primer (5 pmol/ $\mu$ l), and 0,25  $\mu$ l of Recombinant *Taq* DNA Polymerase (5 u/ $\mu$ l) (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'-ggiggiggiagiaaiaciac-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 (<u>http://</u>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.cnrs.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



FIGURE 1 - Gel image of the PCR amplified RGAs fragments (arrow) with the size of approximately 700 bp for both species. (M) marker, (P) *Carica papaya* and (C) *V. cauliflora*.

	RNBS-A						
	P-loop	P-loop Non-TIR					
					- 0		
<u>L</u>					50		
Solanum	ILGMGGLGKT	TLAQMVFNDQ	RITEH. FNLK	IWVCVSDDFD	EKRLIKAIVE		
Vitis							
Carica	DGGGGKT	TLAQLAFNDK	GVKEH.FEKR	IWVCVSDPFD	EIRIAKAILE		
Giycine		LAQHVENDP	RIQEARFOVK	AWVCVSDDFD	AFRVIRTILE		
Oryza	GVGKT	TLAKMVYNDT	RAKDH . POTK	MWHCVSENFE	AVPLLKSIVE		
Hordeum							
A.thallana	IVGIGGVGKT	TLSQLLYNDQ	HVRSY FGTK	VWAHVSEEFD	VERTTERVIE		
Correa	GVGKT	TLAKKVYNDS	SVICN. FHIR	LWCTVSPEFN	TRSLLIQILC		
G.barbadense	GVGKT	TIMKHIHNDL	LKEQR.FERV	IWVTISKEFD	IVKLQDDIAS		
Maius	GMGGIGKT	TLARAVESKY	HHSFNGQC	ILEEVSKKKN	WAGTŐRŐTTK		
				Kinase 2			
					<u>10</u> 0		
Solanum	SIEGKSL	GDMDLAPLQK	KLQELLNGK.	.RYFLVLDDV	WNEDQEKWDN		
Vitis					WNRDREKWLK		
Carica	SVLGSTP	NLVELENLLQ	AIQQYIKNK.	. KFLLVMDDV	WNEDSLKWAC		
Glycine	AITKSTD	DSRDLEMVHG	RLKEKLTGK.	.RFLLVLDDV	<u>WNENRLKWEA</u>		
Oryza	LATNRRCQVP	DKDTIELLRR	QLEGAIGSR.	. RFLLVLDDV	WNEDENKWKD		
Hordeum				LLVLDDV	WNEDERKWED		
A.thaliana	SVTSRPC	EFTDLDVLQV	KLKERLTGTG	LPFLLVLDDL	WNENFADWDL		
Coffea	SNGKQSRMDE	ELKNLNEHEL	LHKLYQRLKT	KRYLVVFDDV	WD. IKVWNE		
G.barbadense	ALNGYM	PKEGNKVRRA	AILSELLKKV	GKHVLILDDV	WDKVSLEEVG		
Malus	DILKRP	DIKVSSVAEG	TKEIGKRLGS	VKVLIVVDDI	DADQLDELA		
		nyna n					
RNBS-B				RNBS-C			
101					150		
Solanum	LRAVLKIGAS	GASILIT	TRLEKIGS	. IMG TLQL	YQLSNLSQED		
Vitis	LRALLMGGGH	GSKIIVT	TRKIGVGP	.IMGTIQT	YVLSPLPPEE		
Carica	LEHSLKCGLQ	GSKILVT	TRKESVAN	. IMGSRSTHM	YFLEKLSEEE		
Glycine	VLKHLVFGAQ	GSRIIAT	TRSKEVASTM	RSRE	HLLEQLQEDH		
Oryza	ELRPLLCSAA	GGHGSVVVVT	TRSQQVAS	IMGTMRS	HELACLNDDD		
Hordeum	VLKPLLCSVG	G.PGSVIVVT	TRSQKVAS	IMQTLGT	HKLACLSEQD		
A.thaliana	LRQPFIHAAQ	GSQILVT	TRSQRVAS	IMCAVHV	HNLQPLSDGD		
Coffea	LRISFPDEKK	GSRIIFT	SRSSNVAS	QVEFGGKP	HNLVPLSEKE		
G.barbadense	IPEPSSSN	GCKLVLT	TRVEQVCK	YMECKV	IKVKPLSKEE		
Malus	IEHESFGPGS	RIIIT	TRDEQVLN	IHKVDKR	YKAQEMTNEE		
	DNDS C			CL NI			
	RIVDS-C		_	GLPL			
151					200		
Solanum	CWLLFKQRAF	CHQTET . SPK	LMEIGKEIVK	KCGGVPLAAK	TLGGLLRFKR		
Vitis	SRSLFLKHAC	VERVEGESSN	LMEFGYQVVE	KCGGIPIQVR	MSGNLMYSAK		
Carica	CWSVFSNIAF	FGRTDEECKH	LEDIGRKIAR	KCNGLPLAAK	FLGGLLRFRK		
Glycine	CWKLFAKHAF	QDDNIQPNPD	CKEIGTKIVE	KCKGLPLALK	TMGSLLHDKS		
Oryza	SWELFSKKAF	SEEVR.ETAE	LVTIGRLIVK	KCKGLPIALN	AMGGLMSSKQ		
Hordeum	SWKLFAQKAY	SNGVEQNQAE	LVSIGKRIAN	KCRGLPLALK	TMGGLLSSFQ		
A.thaliana	CWSLFMKTVF	GNQEPCLNRE	IGDLAERIVH	KCRGLPLAVK	TLGGVLRFEG		
Coffea	SFELLLKKVF	G NEDCPQA	LHGLGMEIAK	KCRGFPFA			
G.barbadense	ALTLFLNKVG	PNILQSPT	LMPTLRLVVK	ECAGLPLAL.			
Malus	AFELLSWHAF	R NPCPDKE	YIELARDVVD	YRGGLPFALQ	ND		
				_	RNBS-D		
201					250		
Solanum	EESEWEHVRD	SEIWNLPQDE	NSVLPALRLS	YHHLPLDLRQ	CFAYCAVFPK		
VITIS	ETEDWISMRD	NGIWSSEH	LPALKLS	TEKLPSHLKP	CFT		
Carica	TREQWESVLD	NQIWELKEAE	QELFPHLLLS	YYDLPSEVRQ	CFPYCCFF		
Glycine	SVTEWKSILQ	SEIWEFSTER	SDIVPALALS	YHHLP			
oryza	QLHEWKAIAD	SARDK	DEILSMLKLS	YKHLPSEMKQ	CFAFCSIFPR		
Hordeum	QVQEWKAIEE	NDIGDSVRGK	DEIMSILKLS	YRHLSSEMKQ	CFAF		
A.thaliana	KVIEWERVLS	SRIWDLPADK	SNLLPVLRVS	YYYLPAHLKR	CFAYCSIFPK		
Coffea							
G.barbadense		• • • • • • • • • • •	• • • • • • • • • • •				
Malus		• • • • • • • • • • •			• • • • • • • • • • •		

FIGURE 2 - 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. Residues in red bold are internal conserved motifs as determined by Meyers et al (1998). The underlined residue at the end of the kinase–2 motif can be used to predict the presence of the Toll/Interleukin-1 receptor-like domain (TIR). The blue arrow indicates primers region. 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 & Koczyk, 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 - 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.

Conf:	נכבכנכנ	ככככככ	[[[[[[[[	בכבכבבבנ	ובכככנ		
Pred:				$\rightarrow$	Ð		
Pred: AA:	CCCCCHH	HHHHHHH TLAQLAF	HHHHHHHCC CEE	EEEEECCCCC	CHHHH DEIRI		
		10	20	30	40		
Conf:	:כככככככ	בככברבים		בכבבכבנ	]{		
Pred:		)	-0		<b>→</b>		
Pred: AA:	HHHHHHH AKAILES	LOSTPI	CCCHHHHHHHHH LVELENLLQAIC	HHHCCCCCEE QYIKIKKFLI	EEEC		
		50	60	70	80		
Conf:	בכבבבבנ	בכבכבב	בכבבבבבבב	בכבכבכבכ	בכבכן		
Pred:	-0		C	⇒—0			
Pred: AA:	VWIIEDSL	HHHHHCC KWACLEH	CCCCCCCCEEEE SLKCGLQGSKIL	EECCCHHHHH	INGS		
		90	100	110	120		
Conf:	בכבבבנננ	כבבבבי	בכבכבכבכב	בכבבבבבנ	בכככנ		
Pred:	$\rightarrow$			0			
Pred: AA:	CCC CEEE	LEKLSEE	HHHHHHHHHHCCC ECWSVFSIIIAFF	GR TDEECKHI	EDIG		
		130	140	150	160		
Conf:	ככככככנ	נכביככב	בכבבבבבבבב	בככככככככ	ןכככנ		
Pred:		-0	)—(		)		
Pred: AA:	RKIARKCI	CCCHHHH	HHHHHHHHCCCC KFLGGLLRFRKT	REQWESVLDI	IN WE		
		170	180	iso	200		
Conf:	בכככככנ	ככבכבב		]בכבבבבנ			
Pred:			<u> </u>				
Pred: AA:	CCC CHHH	HHHHHHH LFPHLLL	HHHCCCHHHHHH SYYDLPSEVROC	HHHHCCC			
		210	220	230			
[	Legend :						
	g	helis	Conf: 3.00	][ = confidence	e of predicti		
	- stead Pred: predicted secondary structure						
l	- doil AA: target dequende						



FIGURE 4- (A) Secondary structure prediction from PSIPRED server (http://bioinf. cs.ucl.ac.uk/psiout/ 106200617814952.pdf) for the *C. papaya* NBS type protein. (B) Partial three-dimensional structure prediction for the common core GLPL domain. The model was built using the CBS/CNRS Meta server (http://bioserv.cbs.cnrs.fr/ HTML\_BIO/tito.html). The target has 42% identity with the template (1Ko9). The β-strand is colored in yellow as the coil is drawn in white. Small dots indicate the CPK region for the amino acids side chains. Green dotted lines represent the hydrogen bonding inside the β-strand.

The secondary structure prediction for the putative resistance gene from *C. papaya* PSI-BLAST algorithm to detect distant homologues reveals four potential â-strand and 11 á-helices which is similar to a predicted plant disease resistance gene product (Rigden et al., 2000). The secondary prediction is shown in Figure 4A.

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 â-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.

#### ACKNOWLEDGEMENTS

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

#### LITERATURE

- ALTSCHUL S. F.; MADDEN, T.L., SCHAFFER, A.A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, Oxford, v.25, p.3389-3402,1997.
- BADILLO, V. M. Caricacea. Revista de la Facultad de Agronomía, Maracay, v.43, p.111, 1993.
- BADILLO, V. M. Carica L. Vs. Vasconcella St. Hil. (Caricaceae) Con La Rehabilitación de Este Ultimo. ERNSTIA, Macaray, v.10, n.2, p.74-79, 2000.
- BADILLO, V. M. Nota Corretiva Vasconcellea St. Hil. Y No Vasconcella (Caricaceae). ERNSTIA, Macaray, v.11, n.1, p.75-76, 2001.
- BENSON D.A.; KARSCH-MIZRACHI, I.; LIPMAN, D.J.. OSTELL, J.; RAPP, B.A.; WHEELER, D.L.. GenBank. Nucleic Acids Research, Oxford, v.30, p.17-20, 2002.
- BERTIOLI D. J., LEAL-BERTIOLI, S.C.; LION, M. B.; SANTOS, V. L.; PAPPAS JR., G; CANNON, S. B.; GUIMARAES, P.M. A large scale analysis of resistance gene homologues in Arachis. Molecular Genetics and Genomics, Heidelberg, v.270, p.34-45, 2003.
- CANNON S.B., ZHU, H.; BAUMGARTEN, A.M.; SPANGLER, R.; MAY, G.; COOK, D.R.; YOUNG, N.D. Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. Journal of Molecular Evolution, New York, v.54, n.4, p.548-62, 2002.

- CHELKOWSKI, J.; KOCZYK. G. Resistance gene analogues of *Arabidopsis thaliana*: recognition by structure. **Journal of Applied Genetics**, Poznan, v.44, n.3, p.311-21, 2003.
- EWING B.; HILLIER, L.; WENDL, M.C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research, New York, v.8, p.175-185, 1998.
- FERRIER-CANA, E., GEFFROY, V.; MACADRE, C.; CREUSOT, F.; IMBERT-BOLLORE, P.; SEVIGNAC, M.; LANGIN, T. Characterization of expressed NBS-LRR resistance gene candidates from common bean. Theoretical and Applied Genetics, Berlin, v.106, n.2, p.251-61, 2003.
- HOLT III, B.F.; HUBERT, D.A.; DANGL, J.L.. Resistance gene signaling in plants complex similarities to animal innate immunity. Current Opinion Immunology, London, v.15, p.20-25, 2003.
- HUANG X.; MADAN, A. CAP3: A DNA sequence assembly program. Genome Research, New York, v.9, p.868-877, 1999.
- JONES D.T. Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology, London, v.292, n.2, p.195-202. 1999.
- KANAZIN, V., MAREK, L. F.; SHOEMAKER, R.C. Resistance gene analogs are conserved and clustered in soybean. Proceedings National Academy Sciences, Washington, v.93, p.11746-11750, 1996.
- LEE, Y., SEO, J.S.; RODRIGUEZ-LANETTY, M.; LEE, D.H. Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. Molecular Genetics and Genomics, Heidelberg, v.269, n.1, p.101-08, 2003.
- LEISTER, R.T., AUSUBEL, F.M.; KATAGIRI, F. Molecular recognition of pathogen attack occurs inside of plant cells in plant disease resistance specified by the *Arabidopsis* genes RPS2 and RPM1. **Proceedings or the National Academy Sciences,** Washington, v.93, p.15497-15502, 1996.
- LEISTER, R.T.; KATAGIRI, F. A resistance gene product of the nucleotide binding site — leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo. Plant Journal, Oxford, v.22, n.4, p.345-354, 2000.
- MAGO R.; SPIELMEYER, W.; LAWRENCE, J.; LAGUDAH, S.; ELLIS, G.; PRYOR, A. Identification and mapping of molecular markers linked to rust resistance genes located on chromosome 1RS of rye using wheat-rye translocation lines. **Theoretical Applied Genetics**, Berlin, v.104, n.8, p.1317-1324, 2002.
- MANSHARDT, R.; WENSLAFF, T. Zygotic polyembryony in interspecific hybrids of *Carica papaya* and *C. cauliflora*. **Journal of the American Society for Horticultural Science**, Alexandria, v.114, p.684-689, 1989.
- MEYERS, B.C.; MORGANTE, M.; MICHELMORE, R.W. TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. **Plant Journal**, Oxford, v.32, n.1, p.77-92, 2002.
- MOFFETT P., FARNHAM, G.; PEART, J.; BAULCOMBE, D.C. Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. **EMBO Journal**, Oxford, v.21, n.17, p.4511-4519, 2002.
- PAN, Q.; WENDEL, J.; FLUHR, R. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. Journal of Molecular Evolution, New York, v.50. n.3, p.203-213, 2000.
- PEÑUELA, S., DANESH, D.; YOUNG, N.D. Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean. Theoretical Applied Genetics, Berlin, v.104, n.2-3, p.261-272, 2002.
- RIGDEN, D.J.; MELLO, L.V.; BERTIOLI, D.J. Structural modeling of a plant disease resistance gene product domain. **Proteins**, New York, v.41, n.1, p.133-43, 2000.

- ROMMENS, C.M.; KISHORE, G.M. Exploiting the full potential of disease-resistance genes for agricultural use. **Current Opinion Biotechnology**, London, v.11, n.2, p.120-25, 2000.
- SHIRASU, K.; SCHULZE-LEFERT, P. Complex formation, promiscuity and multi-functionality: protein interactions in diseaseresistance pathways. Trends Plant Science, London, v.8, n.6, p.252-258, 2003.
- TELLES G.P.; SILVA, F.R. Trimming and clustering sugarcane ESTs. Genetics Molecular Biology, Riberião Preto, v.24, p.17-23, 2001.
- THOMPSON, J.D.; PLEWNIAK, F.; THIERRY, J.; POCH, O.. DbClustal: rapid and reliable global multiple alignments of protein sequences detected by database searches. **Nucleic Acids Research**, Oxford, v.28, n.15, p.2919-2926, 2000.
- YOUNG, N. D. The genetic architecture of resistance. **Current Opinion Plant Biology,** Reino Unido, v.3, p.285-29'0, 2000.