

LUANA MAHÉ COSTA GOMES

**CARACTERIZAÇÃO PARCIAL DE ANÁLOGOS DE
GENES DE RESISTÊNCIA EM DUAS
ESPÉCIES DE EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento para obtenção do título de *Magister Scientiae*.

**VIÇOSA
MINAS GERAIS – BRASIL
2005**

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

G633c
2005

Gomes, Luana Mahé Costa, 1970-
Caracterização parcial de análogos de genes de resis-
tência em duas espécies de eucalipto / Luana Mahé Costa
Gomes. – Viçosa : UFV, 2005.
viii, 69f. : il. ; 29cm.

Inclui anexo.

Orientador: Sérgio Hermínio Brommonschenkel.
Dissertação (mestrado) - Universidade Federal de
Viçosa.

Referências bibliográficas: f. 43-51.

1. Genética vegetal. 2. Genética molecular. 3. Eucalipto - Resistência a doenças e pragas. 4. Cromossomos de bactéria artificial. 5. Clonagem molecular.
I. Universidade Federal de Viçosa. II. Título.

CDD 22.ed. 581.35

LUANA MAHÉ COSTA GOMES

**CARACTERIZAÇÃO PARCIAL DE ANÁLOGOS DE
GENES DE RESISTÊNCIA EM DUAS
ESPÉCIES DE EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento para obtenção do título de *Magister Scientiae*.

APROVADA: 10 de agosto de 2005.

Prof. Francisco Murilo Zerbini Júnior
(Conselheiro)

Prof. Júlio César de Mattos
Cascardo
(Conselheiro)

Profa. Marisa Vieira de Queiroz

Dr. Douglas Lau

Prof. Sérgio Hermínio Brommonschenkel
(Orientador)

A Deus.
Aos meus filhos Vítor e Júlia,
por serem frutos de meu amor e razão do meu viver.
Ao meu esposo Fábio,
pelo eterno e pleno amor!

AGRADECIMENTOS

À Universidade Federal de Viçosa, ao Núcleo de Biotecnologia Aplicada (BIOAGRO) e à coordenação do Curso de Pós-Graduação em Genética e Melhoramento, pela formação científica e profissional.

À CAPES, pela concessão da bolsa de estudo.

Ao professor Sérgio Hermínio Brommonschenkel, pela orientação e oportunidade de aprendizagem.

Aos professores Francisco Murilo Zerbini Júnior, Marisa Vieira de Queiroz e em especial ao Prof. Júlio César de Mattos Cascardo e ao Dr. Douglas Lau pelas colaborações e sugestões neste trabalho.

Aos amigos do Laboratório de Genômica Cynthia, Edgar, Evandro, Gustavo, João Paulo, Klaus, Leonardo, Lúcio, Luiz e Tadeu pelos bons momentos de convivência, amizade, cumplicidade, além do apoio em momentos oportunos. Em especial Bruno e Janaína, pela dedicação na realização deste trabalho, além dos novos amigos Miki e Reginaldo que compartilharam longas conversas produtivas.

Ao meu marido Fábio pelo amor incondicional e por sua presença e dedicação constante. Aos meus filhos, Vítor e Júlia, que são razão de nossas vidas.

A todos que contribuíram de maneira direta ou indireta na realização deste trabalho.

ÍNDICE

RESUMO.....	v
ABSTRACT.....	vii
1. INTRODUÇÃO.....	01
2. MATERIAL E MÉTODOS.....	08
2.1. Material vegetal e extração de DNA.....	08
2.2. Clones de cDNA selecionados para o sequenciamento.....	08
2.3. Sequenciamento dos cDNAs de <i>Eucalyptus</i>	11
2.3.1. Sequenciamento.....	11
2.3.2. Subclonagens.....	11
2.3.3. Análise das sequências dos clones cDNAs.....	14
2.4. Triagem da biblioteca BAC de <i>Eucalyptus grandis</i>	14
2.5. Caracterização dos clones BACs identificados.....	17
2.6. Construção da biblioteca <i>shotgun</i> do BAC 143-F12.....	17
2.7. Sequenciamento da biblioteca <i>shotgun</i>	18
2.8. Análise de sequências para identificação de RGAs.....	18
3. RESULTADOS.....	20
3.1. Caracterização dos cDNAs de <i>Eucalyptus</i> sp.....	20
3.2. Identificação de clones BACs correspondentes aos cDNAs.....	27
3.3. Sequenciamento da biblioteca <i>shotgun</i> do BAC 143-F12 e análise das sequências obtidas.....	29
4. DISCUSSÃO.....	39
REFERÊNCIAS BIBLIOGRÁFICAS.....	43
ANEXOS.....	52

RESUMO

GOMES, Luana Mahé Costa, M.S., Universidade Federal de Viçosa, agosto de 2005. **Caracterização parcial de análogos de genes de resistência em duas espécies de eucalipto.** Orientador: Sérgio Hermínio Brommonschenkel. Conselheiros: Francisco Murilo Zerbini Júnior e Júlio César de Mattos Cascardo.

A maioria dos genes de resistência (genes *R*) clonados codifica para proteínas *R* que contêm um domínio C-terminal rico em repetições do aminoácido leucina (domínio LRR), um domínio central conservado contendo sítios de ligação a nucleotídeos trifosfatados (domínio NBS) e um domínio N-terminal variável (TIR ou não-TIR). O domínio LRR provavelmente está envolvido no reconhecimento de proteínas do patógeno produzidas durante o processo de infecção. O domínio NBS está relacionado com a hidrólise de nucleotídeos trifosfatados e participa de uma sinalização celular necessária à resposta de resistência da planta. A região N-terminal variável pode conter um domínio com similaridade com a proteína ao *Toll* de *Drosophila* e Interleucina de mamíferos (TIR) ou ainda um domínio não-TIR, como *coiled-coil* (CC), ambos aparentemente envolvidos com sinalização celular. Ao contrário do observado em culturas agrônomicas, existem poucos trabalhos de caracterização de genes de resistência em árvores, especialmente em *Eucalyptus*. Estudos de caracterização do transcriptoma desta planta, como parte do projeto Genolyptus (<http://genolyptus.ucb.br>), resultaram na identificação de várias etiquetas de sequências de genes expressos (ESTs) com similaridade a genes *R* e genes envolvidos em resposta de defesa. Desta forma, este trabalho teve por objetivos: 1) sequenciar completamente oito cDNAs de eucalyptus pré-selecionados por possuírem similaridade com genes envolvidos em resistência a doenças em diferentes espécies vegetais, como linho e *Populus* e ao gene *Ra1* envolvido em resposta de defesa; 2) identificar clones BACs de *E. grandis*

correspondentes a cada um desses cDNAs, visando a futura determinação desses genes; e 3) sequenciar parcialmente pelo menos um clone BAC visando determinar a estrutura completa de pelo menos um gene similar a gene *R*. A caracterização dos cDNAs revelou que a maioria dos clones possuem seqüências incompletas de genes *R*. Foram identificados onze clones BACs (insertos variando de 20 a 280 kb) correspondentes a esses genes e um deles foi selecionado (BAC 143 F-12, inserto de 90 kb) para construção de uma biblioteca *shotgun*, visando a caracterização parcial de seu inserto. As seqüências parciais de 917 subclones desta biblioteca foram agrupadas em 14 contíguos sendo que nove apresentaram similaridade com genes conhecidos. Dois contíguos mostraram similaridade com genes que codificam proteínas da classe TIR-NBS-LRR e os demais a genes que codificam para fosfatidilinositol 3-quinases putativas, poligalacturonase e a poliproteínas putativas (retrotransposons do grupo *Ty1-copia*). Análises mais detalhadas dos contíguos com seqüências similares a genes *R* revelaram que um deles contém a ORF completa de um gene da classe TIR-NBS-LRR e o outro contém um gene truncado com domínios TIR-NBS e parte de um domínio LRR. Esses resultados sugerem que este BAC seja derivado de uma região genômica de *E. grandis* que contém um agrupamento de genes TIR-NBS-LRR. Futuros estudos de mapeamento genético poderão determinar se a seqüência desse BAC co-localiza com genes de resistência em *E. grandis*.

ABSTRACT

GOMES, Luana Mahé Costa, M.S., Universidade Federal de Viçosa, August 2005. **Partial characterization of resistance gene analogs from two *Eucalyptus* species.** Adviser: Sérgio Hermínio Brommonschenkel. Committee members: Francisco Murilo Zerbini Júnior and Júlio César de Mattos Cascardo.

Several resistance genes (*R* genes) have been isolated from various plant species in the last decade. These *R* genes can be categorized into several distinct classes based on their predicted protein structures. The largest class falls into the NBS-LRR class, which encodes a nucleotide site domain and a leucine-rich repeat (LRR) domain. NBS-LRR *R* genes can be further subdivided into toll and interleukin-1 (TIR) and non-TIR classes based on the presence of a TIR domain at the N-terminus of the protein. In contrast to agronomic cultures, there are few resistance genes characterized in woody plants, especially in *Eucalyptus*. Transcript characterization studies of this tree, as part of Genolyptus Project (<http://genolyptus.ucb.br>), have identified several expressed sequence tags (ESTs) similar to *R* genes and disease resistance response genes. Therefore, this work aimed to: (1) completely sequence eight pre-selected eucalyptus cDNAs with similarity to *R* genes from flax and *Populus* and to resistance response *Rar1* gene from barley; (2) identify *E. grandis* BAC clones corresponding to these cDNAs, seeking future *R* gene genomic organization studies; and (3) partially sequence at least one BAC clone to determine the complete structure of at least one resistance gene analog. The sequencing results showed that the majority of the cDNAs analyzed are truncated, representing incomplete sequences of *R* genes analogs. Eleven BAC clones were identified (inserts varying from 20 to 280 kb) that correspond to these cDNAs and one (BAC 143F12, 90 kb insert) was characterized by shotgun sequencing. Partial sequences obtained from 917 BAC subclones were

grouped in 14 contigs. Nine contigs showed similarity to known genes. Two contigs were similar to TIR-NBS-LRR genes and the others to putative phosphatidylinositol 3-kinase, polygalacturonase and putative polyproteins (Ty1-copia group retrotransposons). Detailed analysis of the contigs similar to *R* genes demonstrated that one has a complete ORF of a TIR-NBS-LRR gene and the other has one TIR-NBS pseudogene with truncated LRR domain. These results suggest that the insert of this BAC is derived from a *Eucalyptus grandis* genomic region that contains a cluster of TIR-NBS-LRR genes. Future genetic mapping studies will elucidate if these BAC inserts co-localize with known *R* genes in *Eucalyptus*.

1. INTRODUÇÃO

As plantas se defendem constantemente contra ataques de patógenos como bactérias, vírus, fungos, invertebrados ou até mesmo de outras plantas. A maioria destes ataques resulta em resistência da planta sendo a ocorrência de doença uma exceção. Existem barreiras naturais, como parede celular, camadas de cera e compostos químicos que conferem resistência de amplo espectro às plantas em resposta ao ataque inicial de uma grande variedade de patógenos. Em adição a esta primeira linha de defesa constitutiva, cada célula vegetal possui mecanismos de defesa induzíveis e, em caso de suplantação da barreira pré-formada, há uma reorganização metabólica na planta visando conter o organismo invasor no sítio de infecção (Taken & Joosten, 2000; Hammond-Kosack & Jones, 2000; Martin et al., 2003; Hammond-Kosack & Parker, 2003). Esta rápida e efetiva reorganização metabólica é desencadeada por proteínas codificadas por genes de resistência (genes *R*). Estas proteínas são capazes de, direta ou indiretamente, reconhecer proteínas elicitoras produzidas pelos patógenos e ativar as respostas de defesa (Takken & Joosten, 2000; Hammond-Kosack & Jones, 2000; Martin et al., 2003; Hammond-Kosack & Parker, 2003, Belkhadir et al., 2004).

A resposta de defesa caracteriza-se pela rápida ativação de vias de transdução de sinais, de forma coordenada e complexa, que resulta na alteração de fluxo de íons, geração de espécies reativas de oxigênio (*reactive oxygen species*, ROS) e produção de óxido nítrico (NO), fortalecimento da parede celular e morte celular programada das células inicialmente infectadas, evento este denominado resposta de hipersensibilidade (*hypersensitive reaction*, HR) (Hammond-Kosack & Jones, 1997; Martin et al., 2003; Belkhadir et al., 2004). Devido à ação de moléculas sinalizadoras, também ocorre nas células adjacentes àquelas que sofreram HR a ativação de genes de defesa que codificam proteínas de parede celular, peroxidases, proteinases, enzimas hidrolíticas, proteínas relacionadas à patogênese (proteínas PR), como quitinases e glucanases, e enzimas envolvidas no metabolismo secundário

(Takken & Joosten, 2000). Finalmente, ocorre uma sinalização sistêmica que leva a ativação de genes de defesa em locais distintos do sítio de infecção. Esta resposta conhecida como resistência sistêmica adquirida (SAR) confere resistência de amplo espectro efetiva contra um grande número de patógenos (Hammond-Kosack & Parker, 2003; Durrant & Dowg, 2004).

A caracterização molecular dos genes de resistência, em mono e dicotiledôneas, revelou que estes genes codificam proteínas com motivos estruturais comuns, indicando que o modo de reconhecimento de diferentes patógenos e a sinalização de defesa da planta foram conservados durante sua evolução e diversificação (Figura 1). A maioria das proteínas de resistência (proteínas R) contém domínios C-terminais com repetições ricas em aminoácido leucina (*leucine-rich repeats*, LRR), um domínio central com sítios de ligação a nucleotídeos trifosfatados (*nucleotide binding site*, NBS) e domínios N-terminais variáveis.

O domínio LRR está presente em muitas proteínas com diferentes funções, mas sempre envolvido na mediação de interações proteína-proteína. O número de repetições com intervalos regulares de leucina ou outros resíduos hidrofóbicos pode variar, mas comumente são 24. Podem conter, ainda, resíduos de prolina e asparagina espaçados regularmente. Em proteínas R, o domínio LRR pode estar envolvido no reconhecimento direto ou indireto do produto do gene de avirulência do patógeno (proteína *Avr*) (Kobe & Kajava, 2001; Dangl & Jones, 2001; Martin et al., 2003). O reconhecimento direto foi demonstrado para a interação arroz-*Magnaporthe grisea* na qual a proteína NBS-LRR, codificada pelo gene *Pi-ta* de arroz, interage com a proteína codificada pelo gene *Avr-Pita* do patógeno (Jia et al., 2000). A ação do domínio LRR pode estar associada a interações intramoleculares com outros domínios da proteína R conforme sugerem os resultados de análise de quimeras entre os genes *Mi-1.2* e *Mi-1.1* de tomateiro. Quando houve a troca da região LRR entre os os genes *Mi-1.2* e *Mi-1.1* ocorreu a perda de função do gene *Mi-1.2* e um fenótipo letal nas plantas transformadas, indicando desencadeamento constitutivo de respostas de defesa (Hwang et al., 2000).

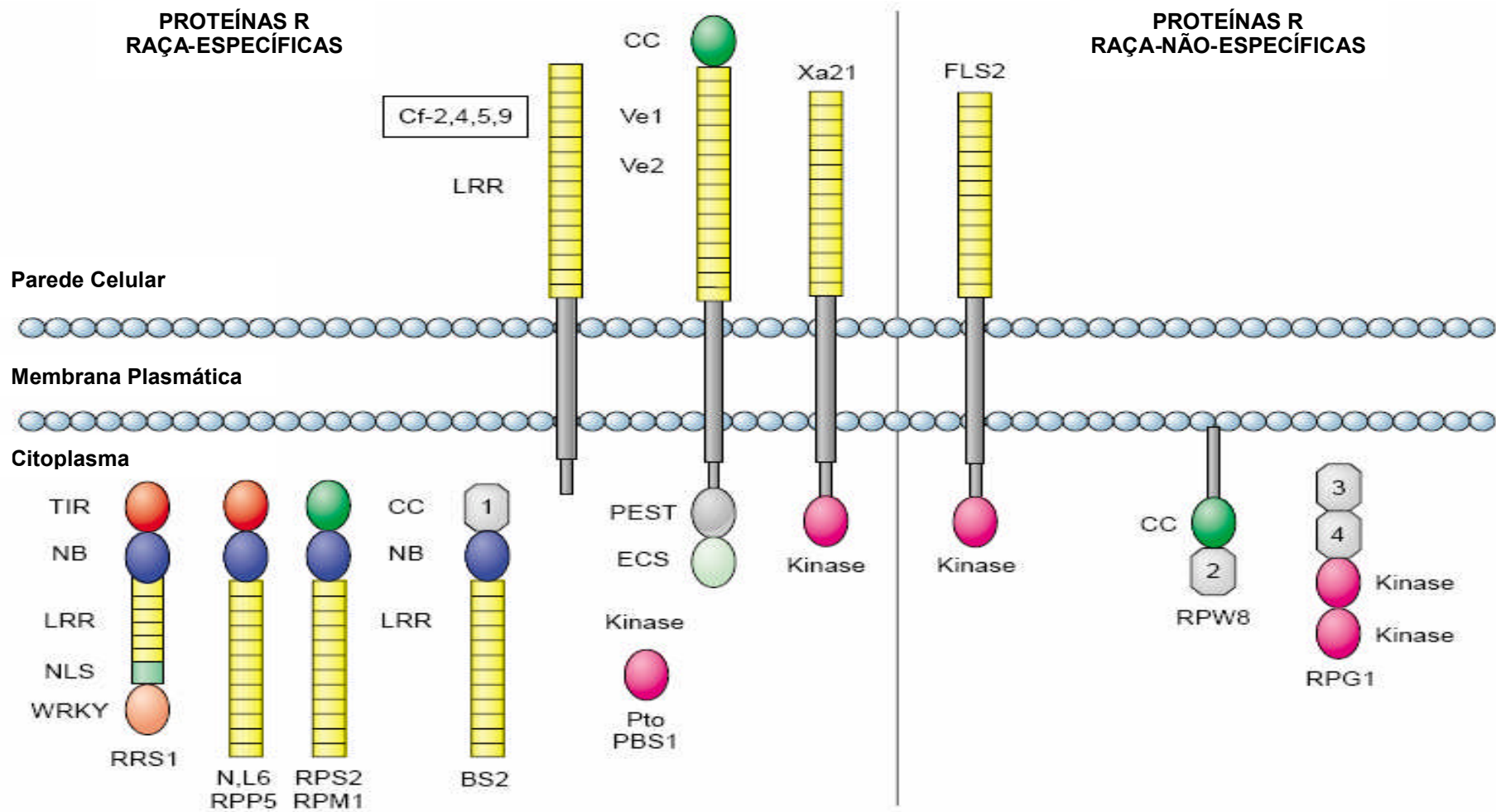


Figura 1 – Representação esquemática de proteínas R codificadas por genes *R* clonados de diferentes espécies de plantas (Hammond-Kosack & Parker (2003)).

O domínio NBS está presente em várias famílias de proteínas como proteínas G e ATPases, promovendo hidrólise de nucleotídeos trifosfatados (ATP e GTP) em proteínas que atuam como sinalizadores celulares, e parece iniciar a ativação da resposta de resistência em plantas (Young, 2000; Pan et al., 2000; Hammond-Kosack & Parker, 2003; Belkhadir et al., 2004). A região NBS inclui motivos quinases 1a (*P-loop*), 2 e 3a que auxiliam na ligação e hidrólise dos nucleotídeos trifosfatados, e outros pequenos motivos com função desconhecida. Uma região do domínio NBS de produtos de genes de resistência (**R**) de plantas é similar ao mesmo domínio das proteínas **APAF-1** de humanos e **CED-4** de nematóides envolvidas na regulação da morte celular programada nestes organismos. Por esta razão, este domínio é atualmente denominado NB-ARC (van der Biezen & Jones, 1998).

As proteínas R com domínios NBS-LRR podem ser divididas em duas subclasses: uma com domínio TIR N-terminal, similar ao domínio citoplasmático *Toll* de *Drosophila* e ao *Receptor Interleucina* (IL-1R) de mamíferos (proteínas conhecidas como TNL ou TIR-NBS-LRR); e uma subclasse com domínio não-TIR N-terminal que possui, na maioria das proteínas R, um motivo *coiled-coil* (CC) (proteínas conhecidas como CNL ou CC-TIR-NBS).

O motivo CC, formado por hélices que contêm cadeias laterais com resíduos de aminoácidos hidrofóbicos, é encontrado em uma variedade de proteínas relacionadas com diversos processos biológicos. Mostra-se envolvido em interações proteína-proteína, mas seu papel na resistência a doenças ainda está por ser elucidado. Em *Arabidopsis*, proteínas R com motivo CC, como RPS2, RPS5 and RPM1, requerem diferentes componentes das vias de transdução de sinais que culminam na resistência daqueles requeridos por proteínas TIR-NBS-LRR, como RPP1, RPP5 e RPS4, sugerindo o envolvimento desse motivo com a sinalização e não com o reconhecimento (van der Biezen et al., 2002; Martin et al., 2003). O domínio TIR apresenta similaridade com domínios de proteínas presentes em animais (*Toll* e IL-1R) e requer componentes de sinalização diferentes dos utilizados pelas proteínas CC-NBS-LRR no processo de defesa. Este domínio parece ter função de transdução de sinal, como demonstrado para o gene *N*, onde deleções e mutações pontuais nessa região resultaram em alelos parcialmente funcionais ou com dominância negativa (Dinesh-Kumar et al., 2000). Além do papel na

sinalização celular, o domínio TIR pode ter ainda a função de reconhecimento do patógeno, como demonstrado no estudo de comparação e troca de domínios TIR entre alelos do loco L (Luck et al., 2000; Young, 2000; Martin et al., 2003; Nimchuk et al., 2003).

Além de proteínas com domínios NBS-LRR, existem outras proteínas R pertencentes a classes menos comuns, mas não menos importantes tais como as proteínas da família serina-treonina-quinase (STK), como a codificada pelo gene *Pto* de tomateiro (Martin et al., 1993), e a codificada pelo gene *Rpg1* de cevada (Brueggeman et al., 2002); proteínas com domínio LRR extracitoplasmático (eLRR) ancorado a um domínio transmembrana (TM), como a codificada pelo gene *Cf-9* de tomateiro (Jones et al., 1994); e proteínas da família eLRR-TM-quinase, como o gene *Xa21* de arroz (Song et al., 1995). Esta última apresenta uma estrutura que contém todos os domínios capazes de exercer as funções que se hipotetizava para as proteínas R (reconhecimento e transdução de sinais). Assim, o domínio LRR extracitoplasmático estaria apto a reconhecer a presença de elicitores produzidos pelo patógenos e o domínio quinase localizado no citoplasma seria o ativador de cascatas de transdução de sinais. Domínios quinases estão envolvidos com a via de transdução de sinal através de fosforilações, sendo extremamente importantes na ativação dos mecanismos de defesa da planta (Hammond-Kosack & Jones, 1997; Romeis, 2001; Martin et al., 2003).

Existem ainda proteínas R que não pertencem às classes anteriores como a HC-toxina redutase, que é uma enzima codificada pelo gene *Hm1* de milho capaz de inativar a toxina produzida pelo patógeno (Johal & Briggs, 1992; Hammond-Kosack & Jones, 1997); as glicoproteínas de superfície celular que contêm receptor sinal de endocitose (ECS), como a codificada pelo gene *Ve* de tomateiro (Kawchuk et al., 2001); e proteínas TIR-NBS-LRR, como a codificada pelo gene *RRS1* de *Arabidopsis*, que contém na porção C-terminal, domínios de sinalização nuclear (NLS) junto a um motivo característico de proteínas ativadoras de transcrição da família WRKY (Deslandes et al., 2002).

A maioria dos genes *R* é encontrada em múltiplas cópias agrupadas no genoma de forma justaposta formando locos complexos, como o loco N de linho (Islam & Shepherd, 1991) e o loco *Pto* de tomateiro (Martin et al., 1993; Chang et al., 2002). Existem ainda genes *R* que possuem variações alélicas em um mesmo loco como gene *L* de linho com 13 alelos (Islam & Shepherd,

1991). Acredita-se que a estrutura física dos agrupamentos (*clusters*) auxilia na geração e manutenção da diversidade dos genes *R* por duplicação e *crossing-over* desigual (Meyers et al., 2005), como exemplificado para os locos *Xa21* de arroz (Song et al., 1997) e *Cf-9* de tomateiro (Parniske & Jones, 1999). Assim, recombinações entre alelos ou entre sequências relacionadas contribuem para amplificação da variabilidade genética dos genes *R*, gerando novos alelos capazes de ativar resposta de defesa às novas raças do patógeno ou sequências análogas a genes de resistência (*resistance gene analogs*, RGAs) (Hammond-Kosack et al., 2000; Young, 2000; Hulbert et al., 2001; Zhu et al., 2002; Meyers et al., 2003; Leister, 2004).

Os RGAs são sequências similares às sequências codificadas por genes *R*, mas sem funções na resistência ainda estabelecidas. Estudos de mapeamento de RGAs no genoma de plantas comprovam sua ligação com locos que contêm genes *R* conhecidos (Yu et al., 1996; Collins et al., 1998; Seah et al., 1998; Shen et al., 1998; Chen et al., 1998; Mago et al., 1999; Gowda et al., 2000; Collins et al., 2001). Esta associação com genes *R* mostra que os RGAs são um reservatório de variação genética, podendo contribuir para a evolução de novos genes *R* e sua dispersão pelo genoma (Michelmore & Meyers, 1998; Young, 2000; Ellis et al., 2000; Leister, 2004)

RGAs são valiosos não somente como fontes potenciais para formação de genes *R* ativos no genoma de plantas, mas também como marcadores em programas de melhoramento de plantas. A identificação e mapeamento de RGAs em cevada, por exemplo, demonstrou a sua ligação a genes *R* e a importância para clonagem de genes *R* (Brueggeman et al., 2001). Oligonucleotídeos degenerados, baseados nos domínios conservados NBS-LRR, têm sido utilizados para amplificação e clonagem de RGAs, utilizando a reação em cadeia da polimerase (PCR), em diferentes espécies de plantas, como em soja, cevada e alface (Kanazin et al., 1996; Shen et al., 1998; Yu et al., 2000; Brueggeman et al., 2001). Essa estratégia foi utilizada por Laia (2001) na tentativa de identificar RGAs que pudessem estar ligados ao loco *Ppr1* de *Eucalyptus*, que contém o gene *Ppr1*, que confere resistência à ferrugem (*Puccinia psidii*) (Junghans et al., 2003). Nesse estudo, foram caracterizados parcialmente vários RGAs de *Eucalyptus grandis* que apresentaram similaridade a genes *R* já caracterizados em outras espécies vegetais, como os

genes *L*⁶ e *M* em linho. Todavia, não foram encontrados RGAs ligados ao loco que contém o gene *Ppr1*.

Programas de sequenciamento genômico para espécies florestais ainda são restritos. Estudos de genomas funcionais estão sendo realizados em *Pinus*, *Populus* e, mais recentemente, em *Eucalyptus*, sendo *Populus deltoides* a planta modelo entre as espécies florestais (NSF, 2001; Genolyptus, 2001; Forest, 2001; Taylor, 2002). Como parte do projeto Genolyptus (Rede Nacional de Pesquisa do Genoma de *Eucalyptus*), no qual a UFV participa, bibliotecas de cDNAs foram construídas a partir de mRNAs isolados de diferentes tecidos de *Eucalyptus* spp.. Os cDNAs foram parcialmente sequenciados e uma análise preliminar das sequências obtidas revelou que vários cDNAs dessas bibliotecas possuem similaridades com genes *R* e genes envolvidos em resposta de defesa. Visando o mapeamento físico do genoma desta planta também foi construída uma biblioteca genômica da espécie *E. grandis* em cromossomos artificiais de bactérias (BAC). Esta biblioteca possui 70% dos insertos com tamanho superior a 150 kb (Brommonschenkel, dados não publicados), o que viabiliza o estudo da organização genômica de locos complexos desta espécie, como os locos que contêm genes *R*.

Com base no exposto e na inexistência de um estudo detalhado de genes *R* em *Eucalyptus*, este trabalho teve por objetivos: 1) sequenciar completamente oito cDNAs pré-selecionados que possuem similaridade com genes envolvidos na resistência a doenças em diferentes espécies vegetais, como linho e *Populus* e ao gene *Rar1* envolvido em resposta de defesa (Azevedo et al., 2002; Takahashi et al., 2003); 2) Identificar clones BACs de *E. grandis* correspondentes a cada um desses cDNAs, visando a futura determinação da organização genômica dos genes correspondentes; e 3) sequenciar parcialmente pelo menos um clone BAC visando determinar a estrutura completa de pelo menos um gene similar a gene *R*.

2. MATERIAL E MÉTODOS

2.1. Material vegetal e extração de DNA

O clone G21 da espécie *E. grandis*, cedido pela Cia. Suzano de Papel & Celulose, foi utilizado como fonte do DNA utilizado nas amplificações por PCR. Este clone possui o gene *Ppr1* que confere resistência à ferrugem causada pelo fungo *Puccinia psidii* (Junghans et al., 2003). Folhas jovens de mudas do clone G21 foram coletadas e o DNA genômico extraído de acordo com o protocolo descrito por Ferreira & Grattapaglia (1995).

2.2. Clones de cDNA selecionados para o sequenciamento

Do banco de dados do Genolyptus foram selecionados oito cDNAs que possuíam similaridade com sequências de genes que codificam proteínas R da classe NBS-LRR e à proteína Rar1 envolvida na resposta de defesa em plantas (Tabela 1). Esses cDNAs, pertencentes a diferentes bibliotecas do Genolyptus, já haviam sido parcialmente caracterizados por sequenciamento da região 5'-terminal.

Os clones de cDNA descritos na tabela 1, foram denominados EUGR ou EUPE referindo-se as espécies a partir das quais foram desenvolvidas as bibliotecas (*E. grandis* ou *E. pellita*, respectivamente); PU e XY denomina o tecido da planta do qual foi extraído e purificado o mRNA (folhas infectadas pelo fungo *P. psidii* e xilema, respectivamente); o primeiro código numérico denomina a fração da biblioteca e o segundo o número da placa; o código alfanumérico denomina a posição do clone na placa; e MG denomina o laboratório onde está armazenada a placa.

A estimativa do tamanho dos fragmentos clonados (Tabela 1) já havia sido feita pela realização de PCR de colônia bacteriana, utilizando os

oligonucleotídeos M13 *forward* e *reverse*, que anelam nas regiões que flanqueiam o inserto.

Os seguintes clones foram selecionados do banco de dados do Genolyptus (Tabela 1): EUGR-PU-000-001-G07-MG, EUGR-PU-000-010-D04-MG e EUPE-XY-000-035-C06-MG similares ao gene *L* de linho (*Linum usitatissimum*) que codifica diferentes proteínas que conferem resistência à ferrugem do linho, relacionados com diferentes alelos desse gene: L^9 , L^H e L^1 , respectivamente (Ellis et al., 1999), apresentando, então, códigos de acessos diferenciados; clones EUGR-PU-003-001-C10-MG e EUPE-XY-000-048-H05-MG similares aos genes *R* putativos 60I2G01 e 60I2G11, respectivamente, de *Populus deltoides* que codificam proteínas de resistência da subclasse TIR-NBS-LRR (Lescot et al., 2004); clone EUGR-PU-000-005-H06-MG similar ao gene *Nbi-B* de linho (*L. usitatissimum*), componente do loco N, que codifica a proteína de resistência Nbi-B da subclasse TIR-NBS-LRR (Dodds et al., 2001); clone EUGR-PU-003-020-C11-MG similar ao gene *M* de linho (*Linum usitatissimum*), que codifica a proteína de resistência M da classe NBS-LRR (Anderson et al., 1997); e clone EUGR-PU-003-042-F06-MG similar ao gene *Rar1* de cevada (*Hordeum vulgare*), que codifica a proteína de defesa Rar1 (Shirasu et al., 1999).

A partir das sequências parciais das extremidades dos clones selecionados, foi desenhado um par de oligonucleotídeos (*forward* e *reverse*) utilizando o programa *Primer3* (www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi) e sintetizados pela *Invitrogen* Brasil (Anexo 1). Estes oligonucleotídeos foram utilizados no sequenciamento dos oito cDNAs selecionados visando obter suas sequências completas, além de terem sido utilizados na triagem da biblioteca genômica BAC de *E. grandis*.

Tabela 1 – Clones de cDNAs selecionados de *Eucalyptus grandis* e *E. pellita* cujas sequências possuem similaridade com genes que codificam proteínas R ou proteínas relacionadas com resposta de resistência. EUGR-PU refere-se aos clones de cDNA de *E. grandis* derivados de folhas jovens naturalmente infectadas pelo fungo *P. psidii*, EUPE-XY refere-se aos clones de cDNA derivados de mRNA isolado de xilema de *E. pellita*.

Clone	Tamanho estimado (kb)	Proteína	Acesso	Valor esperado
EUGR-PU-000-001-G07-MG	1,9	Proteína de resistência à ferrugem em linho [<i>Linum usitatissimum</i>]	gi 4588068 gb AAD25975.1	1e ⁻²⁴
EUGR-PU-000-005-H06-MG	2,4	Proteína Nbi-B [<i>L. usitatissimum</i>]	gi 13509232 bem CAC35336.1	4e ⁻²²
EUGR-PU-000-010-D04-MG	1,4	Proteína de resistência à ferrugem em linho [<i>L. usitatissimum</i>]	gi 4588070 gb AAD25976.1	7e ⁻²⁷
EUGR-PU-003-001-C10-MG	1,9	Proteína TIR/NBS/LRR [<i>Populus deltoides</i>]	gi 40644187 bem CAC95120.1	3e ⁻¹⁸
EUGR-PU-003-020-C11-MG	0,8	Proteína M de resistência à ferrugem [<i>L. usitatissimum</i>]	gi 1842251 gb AAB47618.1	2e ⁻²¹
EUGR-PU-003-042-F06-MG	1,45	Proteína Rar1 [<i>Hordeum vulgare</i>]	gi 6581046 gb AAF18432.1	5e ⁻³⁷
EUPE-XY-000-035-C06-MG	1,8	Proteína de resistência à ferrugem em linho [<i>L. usitatissimum</i>]	gi 4588050 gb AAD25966.1	4e ⁻²⁴
EUPE-XY-000-048-H05-MG	1,8	Proteína TIR/NBS/LRR [<i>P. deltoides</i>]	gi 40644189 bem CAC95124.1	1e ⁻¹⁹

2.3. Sequenciamento dos cDNAs de *Eucalyptus*

Para a obtenção da sequência completa do inserto de cada um dos clones, foram utilizadas as estratégias de *primer walking* e subclonagem.

2.3.1. Sequenciamento

Para o sequenciamento dos cDNAs foram utilizados oligonucleotídeos M13 *forward e reverse* e T7 que anelam em regiões adjacentes ao sítio múltiplo de clonagem do vetor pSPORT1 e os oligonucleotídeos RGs (Anexo 1), que anelam nas regiões internas do inserto. As condições das reações foram 100-200 ng de DNA plasmidial, 2 μ L *Dynamic ET terminator Premix (Amersham Biosciences)*, 0,5 μ M do oligonucleotídeo correspondente, num volume final de 5 μ L. As reações de sequenciamento foram realizadas em termociclador *Mastercycler (Eppendorf)* com 35 ciclos a 95 °C por 10 segundos, 52 °C por 5 segundos e 2 minutos a 60 °C.

Os fragmentos amplificados foram precipitados com etanol de acordo com manual de instruções do kit, com modificações. Após a adição do acetato de amônio 7,5 mM e etanol 100%, a placa foi deixada a temperatura ambiente por 10 minutos. Em seguida as placas foram centrifugadas por 45 minutos a 2250 x g a 20 °C. Após o descarte do sobrenadante, o precipitado foi lavado com etanol 70% e centrifugado por 10 minutos a 2250 x g a 20 °C. A seguir realizou-se um *spin* invertido por 30 segundos a 490 x g para retirar o excesso de etanol. O DNA foi deixado para secar por 10 minutos à temperatura ambiente. O precipitado foi ressuscitado com 5 μ L de *MegaBACE loading solution (Amersham Biosciences)*. As reações foram analisadas no sequenciador automático *MegaBACE™ 500 (Amersham Biosciences)*.

2.3.2. Subclonagens

Efetuiu-se a subclonagem de quatro clones de cDNAs. Estes foram inicialmente analisados quanto ao padrão de restrição utilizando-se diferentes enzimas que clivam nas extremidades do inserto. A reação de digestão foi conduzida num volume final de 100 μ L com 1 a 2 μ g de DNA plasmidial dos clones, 2 μ L da respectiva enzima de restrição e 10 μ L do tampão específico. A

reação foi submetida a 37 °C por 2h. O DNA clivado foi aplicado em gel de agarose 1% em TAE 1X e submetido à eletroforese a 6 V/cm. Os fragmentos de tamanho esperado, contendo o vetor pSPORT1 ligado ao fragmento de cDNA remanescente após a digestão foram purificados do gel utilizando *SephaglasTM BandPrep kit*, de acordo com o fabricante.

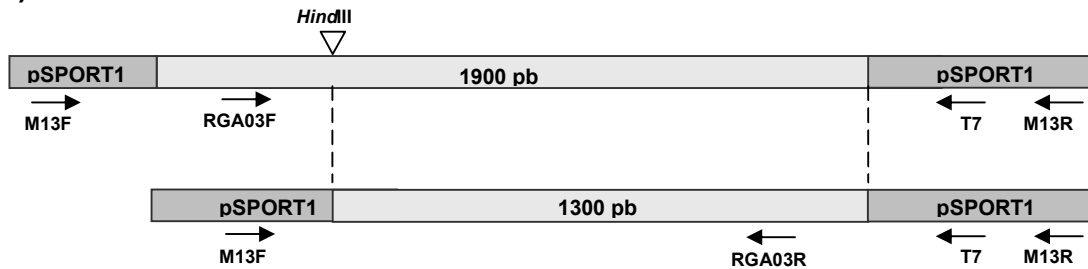
Cerca de 10-15 ng do DNA eluído foi utilizado na reação de ligação, realizada num volume final de 10 µL contendo 2 µL da enzima de ligação T4 DNA ligase (*Promega*) e 5 µL do tampão de ligação. A reação de ligação foi conduzida a 4 °C por 24 h e uma alíquota de 5 µL dessa reação foi utilizada para transformar células competentes de *Escherichia coli* EC100, utilizando o método de choque térmico (Sambrook et al., 1989) e plaqueadas em meio LB sólido contendo 12,5 µg. mL⁻¹ de cloranfenicol.

Colônias transformantes foram coletadas para crescimento em tubos Falcon de 15 mL contendo 3 mL de meio Luria-Bertani-LB enriquecido (2X) contendo cloranfenicol (12,5 µg. mL⁻¹) e submetidos à agitação de 280 rpm a 37 °C por 20 horas. A extração de DNA plasmidial foi realizada por lise alcalina e sua concentração determinada por análise eletroforética em gel de agarose a 1%, comparando-se com quantidades conhecidas de DNA de fago lambda.

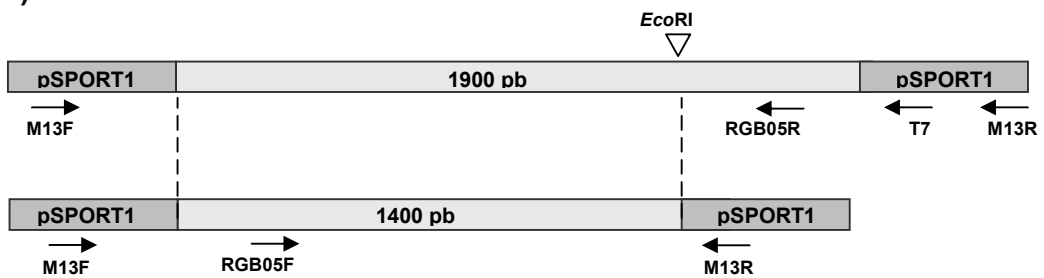
A clonagem foi confirmada por clivagem utilizando-se o DNA plasmidial obtido e duas enzimas com sítio de restrição nas extremidades do inserto. Os seis subclones contendo insertos menores, referentes aos quatro cDNAs, foram estocados em glicerol 50% e armazenados a - 80 °C, com números de identificação 725, 726, 727, 728, 729 e 730, representando os subclones dos clones EUGR-PU-000-001-G07-MG (1300 pb), EUGR-PU-002-010-C10-MG (1400 pb), EUPE-XY-000-035-C06-MG (dois fragmentos de 1200 pb) e EUPE-XY-000-048-H05-MG (fragmentos de 1200 e 800 pb), respectivamente. Uma representação esquemática dos subclones é mostrada na figura 2.

Para o sequenciamento dos subclones, as amostras de DNA foram inicialmente purificadas utilizando extração com fenol-clorofórmio (Sambrook et al., 1989) e ressuspensas em 15 µL de água Milli-Q. A concentração do DNA purificado foi analisada em gel 1%. Na reação de sequenciamento foram utilizados os oligonucleotídeos M13 *forward* e *reverse*, T7 e RGs *forward* e *reverse* (Anexo 1). Estas reações foram feitas sob as mesmas condições citadas para o sequenciamento dos cDNAs, citado no item 2.3.1.

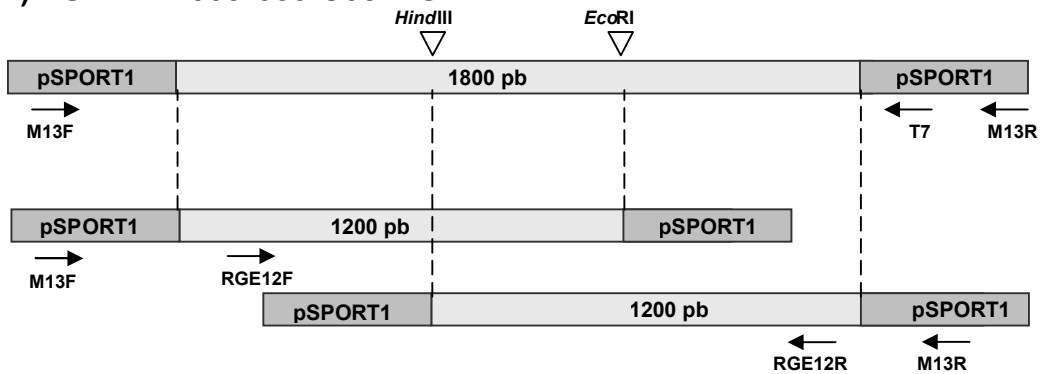
(A) EUGR-PU-000-001-G07-MG



(B) EUGR-PU-003-001-C10-MG



(C) EUPE-XY-000-035-C06-MG



(D) EUPE-XY-000-048-H05-MG

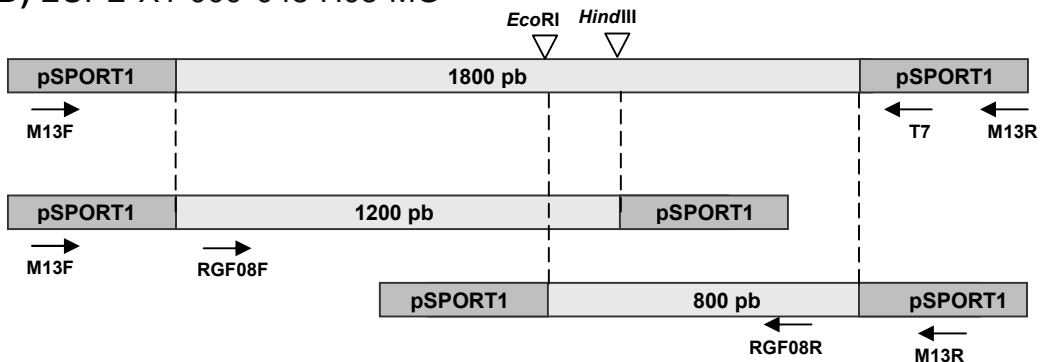


Figura 2 – Mapa de restrição dos clones de cDNA selecionados e dos subclones obtidos. **(A)** EUGR-PU-000-001-G07-MG (1900 pb); **(B)** EUGR-PU-003-001-C10-MG (1900 pb); **(C)** EUPE-XY-000-035-C06-MG (1800 pb); **(D)** EUPE-XY-000-048-H05-MG (1800 pb). As setas representam os oligonucleotídeos utilizados no sequenciamento.

2.3.3. Análise das sequências dos clones cDNAs

As sequências de cada clone foram alinhadas utilizando-se o programa *CAP* (www.infobiogen.fr/services/analyseseq/cgi-bin/cap_in.pl) para obtenção da sequência consenso. Essas sequências consensos foram comparadas com sequências depositadas no banco mundial de dados *GenBank* (NCBI), utilizando o programa *Blastx-nr*.

Cada sequência de DNA consenso foi traduzida utilizando o programa *Transeq* (www.ebi.ac.uk/emboss/transeq/index.html) e, posteriormente, alinhada com duas proteínas com maior similaridade revelada pelo programa *Blastx-nr* para verificar se a sequência consenso codificava uma proteína completa. Os alinhamentos foram feitos no programa de alinhamento múltiplo *Clustal W* (<http://clustalw.genome.jp/>). Para os clones com sequências similares a genes que codificam proteínas R, o alinhamento foi feito utilizando a sequência de aminoácido deduzida com a proteína R com maior similaridade ao clone e uma proteína R de *P. deltoides*.

Somente para o clone EUPE-XY-000-048-H05-MG o programa *Webcutter* (www.firstmarket.com/cutter/cut2.html) foi utilizado para verificação da existência de sítios para *NotI* e o *Expasy Translate Tool* (www.expasy.org/tools/dna.html) para obtenção da localização da metionina de iniciação e o aminoácido de terminação nessa sequência. Isso em função desse clone conter duas regiões de diferentes proteínas, o que normalmente não ocorre.

2.4. Triagem da biblioteca BAC de *Eucalyptus grandis*

Foi utilizada uma biblioteca genômica BAC de *E. grandis*, clonada no vetor pCC1 (*Epicenter*), construída no Lab. de Genômica do BIOAGRO (UFV) como parte do projeto Genolyptus (Brommonschenkel et al., dados não publicados). Esta biblioteca é constituída por 20.160 clones, com tamanho médio de inserto de 150 kb, armazenados em 210 microplacas de 96 cavidades, representando cerca de quatro a cinco vezes o genoma de *E. grandis*, que possui um tamanho estimado em 640 Mb (Grattapaglia, 1994).

A triagem da biblioteca BAC foi realizada por PCR utilizando cada par de oligonucleotídeo RGs (Anexo1). As reações foram feitas em volume final de 25 µL com 1 µL de DNA dos BACs (50 ng/µL), 2,5 µL de tampão 10X (1,5 mM

MgCl₂), 0,2 µM de cada oligonucleotídeo RG, 0,2 mM de cada dNTP e 2,5 U de *Taq* DNA Polimerase (*Phonectria Biotecnologia*). Utilizou-se termociclador *Mastercycler* (*Eppendorf*) com um ciclo a 95 °C por 2 minutos, seguido de 40 ciclos a 95 °C por 30 segundos, 52 °C por 1 minuto e 1,5 minutos a 72 °C, seguidos de uma extensão final de 7 minutos a 72 °C. Os produtos de amplificação foram analisados em gel de agarose 1% em tampão TAE 1X (Sambrook et al., 1989) a 6 V/cm. O gel foi corado em 0,2 µM de brometo de etídio, visualizado sob luz ultravioleta e fotodocumentado.

Para facilitar a triagem das 210 microplacas da biblioteca BAC utilizou-se a estratégia de triagem em *pools* de clones (Figura 3). As células bacterianas de cada microplaca foram plaqueadas, crescidas e coletadas em conjunto (*pool*) e o DNA dos BACs foi purificado por lise alcalina (Sambrook et al., 1989). Cada *pool* de DNA (96 clones) foi acondicionado em um microtubo e armazenado a - 20 °C (210 microtubos). Amostras DNA de cinco *pools* foram combinadas numa nova microplaca gerando um *superpool* de DNA. Os 20.160 clones estavam, então, representados em 42 cavidades de uma microplaca (cada cavidade com amostras de DNA de 480 clones), a qual também foi armazenada a - 20 °C.

Para início da triagem na biblioteca BAC, os oligonucleotídeos RGs foram utilizados na reação de PCR junto às 42 amostras do *superpool*. Após a identificação da cavidade positiva na microplaca de *superpool* pela visualização do produto de PCR em gel de agarose 1%, uma segunda reação de PCR foi realizada, utilizando oligonucleotídeos RGs, nos cinco *pools* de DNA correspondentes àquela cavidade. O *pool* de DNA positivo, detectado pela visualização do produto de PCR em gel de agarose 1%, identificou a microplaca com o(s) possível(is) clone(s) positivo(s). Finalmente, para identificar o clone positivo na microplaca, efetuou-se a terceira reação de PCR utilizando-se como molde as células bacterianas de cada cavidade individualizada (“PCR de colônia”) (Figura 3).

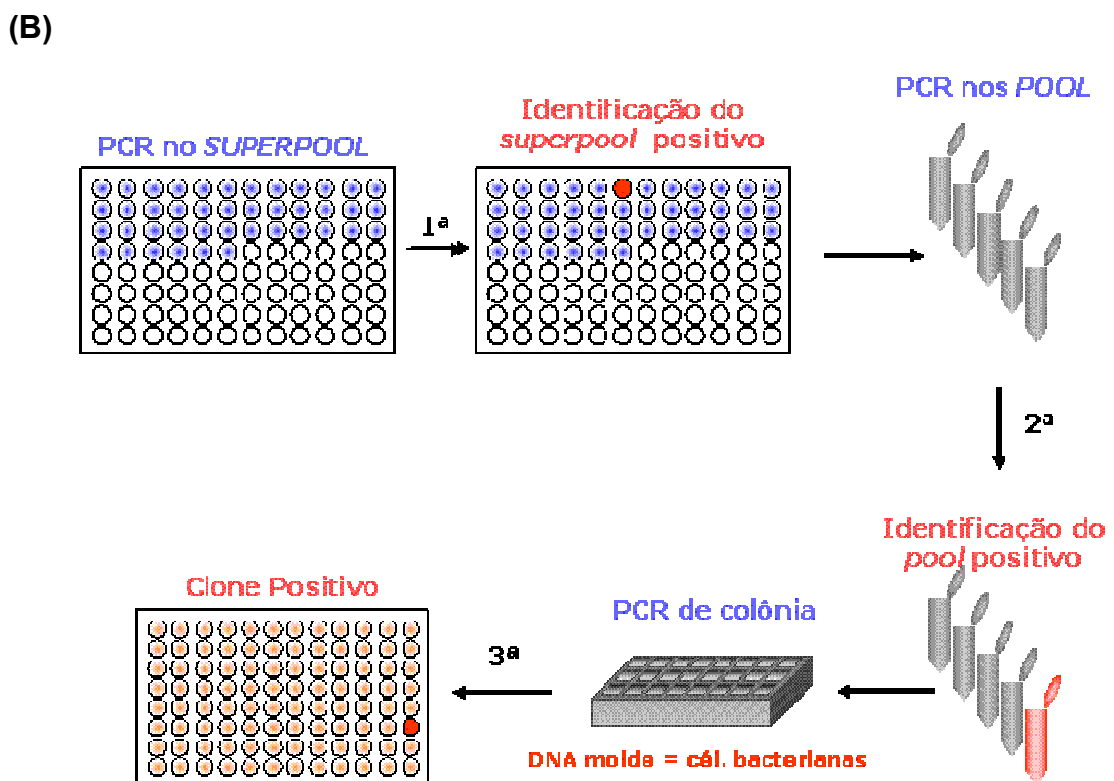
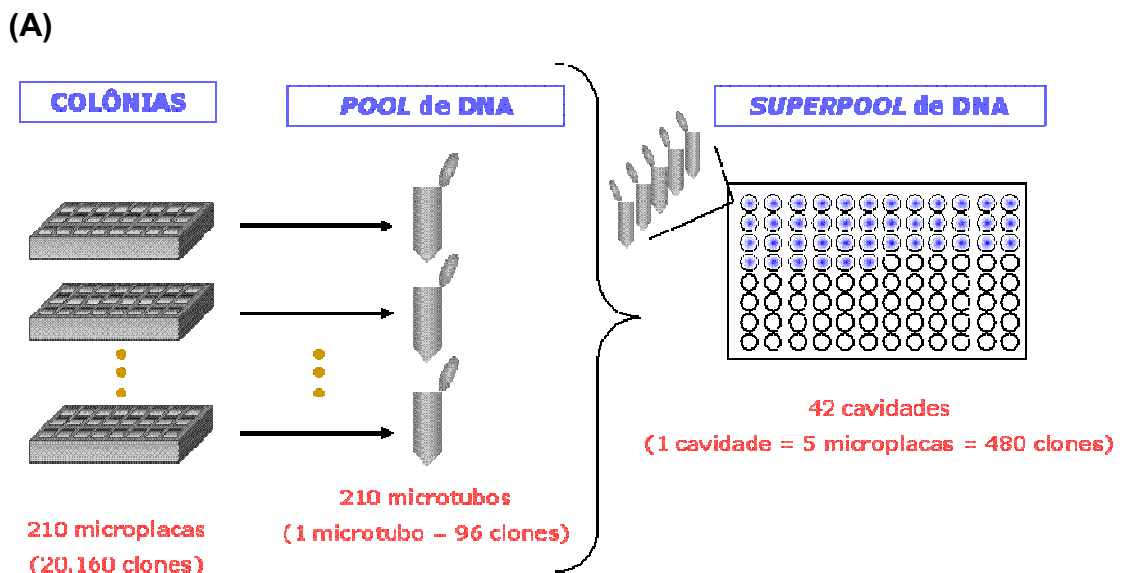


Figura 3– (A) DNA plasmidial de 96 clones de cada microplaca coletado em conjunto. Após extração por lise alcalina o DNA foi acondicionado em um microtubo formando um *pool* de DNA. Amostra de cada cinco *pools* foi utilizada para formação de um *superpool* de DNA. **(B)** Três ampliações por PCR (no *superpools* de DNA, nos *pools* de DNA e na microplaca com as colônias) foram utilizadas para se encontrar o(s) clone(s) positivo(s).

2.5. Caracterização dos clones BACs identificados

Os clones positivos foram plaqueados individualmente em meio LB (Sambrook et al., 1989) contendo cloranfenicol ($12,5 \mu\text{g}\cdot\text{mL}^{-1}$) a partir do estoque de glicerol. Após o crescimento, colônias isoladas foram coletadas para novo crescimento em um volume maior de meio LB enriquecido (3 mL de 2X LB) contendo cloranfenicol ($12,5 \mu\text{g}\cdot\text{mL}^{-1}$) e submetidos à agitação de 280 rpm a 37°C por 20 horas. Após este período foi realizada a extração de DNA dos BACs por lise alcalina e sua concentração analisada em gel de agarose 1 % (Sambrook et al., 1989).

Análises de PCR foram realizadas utilizando DNA de todos os clones positivos e todos os pares de oligonucleotídeos RGs visando verificar a especificidade dos oligonucleotídeos com seus respectivos clones amplificados. Foi utilizado o, mesmo padrão de reação de 25 μL citado na triagem da biblioteca BAC (item 2.4.).

Para análise do tamanho dos insertos dos BACs, 5 μL de DNA foi submetido a digestão com *NotI* (*Invitrogen*). Foram utilizadas 15 unidades da enzima, 2 μL do tampão 10x num volume final de 20 μL . A reação foi conduzida a uma temperatura de 37°C por 3 a 4 horas. O produto da clivagem foi submetido à eletroforese de campo pulsado (PFGE) em gel de agarose a 1 % em tampão TBE 0,5X (Sambrook et al., 1989) a 14°C , usando o sistema *CHEF MAPPER* (*Bio-Rad*) a 6 V/cm, com pulsos de 15 segundos por 16 horas e com ângulo de reorientação de 120° .

2.6. Construção da biblioteca *shotgun* do BAC 143-F12

O clone BAC 143-F12, identificado por amplificação com o par de oligonucleotídeos RGB05 *forward* e *reverse* (F e R) foi selecionado para a construção da biblioteca *shotgun*. O DNA do BAC foi extraído como descrito no item 2.5., mas para um volume de células crescidas em 500 mL de meio 2X LB. A concentração do DNA plasmidial foi determinada por análise em gel de agarose 1 % (Sambrook et al., 1989).

O DNA desse clone foi digerido com *NotI*, como descrito no item 2.5., porém com um volume de reação de 250 μL e digestão por 7 horas a 37°C . Esse DNA digerido foi aplicado em gel de agarose 1% e submetido à

eletroforese de campo pulsado para separação dos fragmentos. O inserto referente ao BAC 143-F12 foi eluído do gel por eletroeluição, utilizando membrana de diálise, por 4 horas com 42 V em TBE 0,5X. O DNA do inserto foi precipitado e ressuspendido em tampão *Shearing Buffer* (Tris-HCl 10 mM, EDTA 1 mM, glicerol 20% (v/v) pH 8,0). A concentração do DNA fragmentado foi determinada por análise em gel de agarose 1 % (Sambrook et al., 1989).

O DNA purificado do clone BAC 143-F12 foi fragmentado por nebulização conforme instruções do *TOPO Shotgun Subcloning kit* (Invitrogen). Os fragmentos com as extremidades reparadas e desfosforiladas foram ligados ao vetor pCR4Blunt-TOPO (Invitrogen) e utilizados na transformação de células competentes. Foram coletadas e isoladas 960 colônias transformantes, que foram acondicionadas em 10 microplacas de 96 cavidades e estocadas a -80 °C.

2.7. Sequenciamento da biblioteca *shotgun*

Uma alíquota de cada colônia estocada em ultra freezer, derivada da biblioteca do BAC 143-F12, foi utilizada para extração do DNA plasmidial. Este foi extraído por lise alcalina após crescimento das colônias em microplacas contendo meio 2X LB. O sequenciamento parcial (*single-pass sequencing*) dos insertos dos clones da biblioteca *shotgun* (BAC 143-F12) foi realizado utilizando o oligonucleotídeo T7 que anela em região adjacente ao sítio múltiplo de clonagem do vetor pCR4Blunt-TOPO. As reações de sequenciamento foram realizadas como descrito anteriormente (item 2.3.).

2.8. Análise de sequências para identificação de RGAs

A análise das sequências dos clones da biblioteca foi realizada com auxílio de diversos programas, sendo iniciada pela identificação e remoção da sequência do vetor pelo programa *Cross_match* (www.phrap.org). As sequências foram depositadas em banco de dados do Laboratório de Genômica (UFV).

A seguir, as sequências foram processadas com uma marca de qualidade *Phred* 20 e agrupadas com marca de qualidade *Phrap* 40 utilizando os *softwares* PHRED/PHRAP/CONSED (www.phrap.org). Bases incorretas ou

ambíguas foram removidas manualmente e a sequência consenso de cada agrupamento contíguo foi comparada com sequências do banco mundial de dados *GenBank*, utilizando o programa *Blastx-nr* (<http://www.ncbi.nlm.nih.gov/>).

Utilizou-se os programas *ORF Finder* (www.ncbi.nlm.nih.gov/gorf/gorf.html) e *GenScan* (http://bioweb.pasteur.fr/seganal/interfaces/genscan_simple.html) para analisar todos os agrupamentos contíguos, visando encontrar a fase aberta de leitura (ORF – *Open Reading Frame*), junções íntron-éxon e possível região promotora. As sequências de aminoácidos das ORFs encontradas foram comparadas com sequências do banco mundial de dados *GenBank* com auxílio do programa *Blastp-nr* (<http://www.ncbi.nlm.nih.gov/>). Motivos conservados foram identificados nos agrupamentos contíguos relacionados com genes *R* (CT12 e CT13) utilizando o programa NCBI *Conserved Domain Search* (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

A sequência traduzida do agrupamento contíguo CT13 foi alinhada com seis sequências de proteínas com maior similaridade utilizando o programa *ClustalX* (Thompson et al., 1997) visando identificar os domínios conservados.

3. RESULTADOS

3.1. Caracterização dos cDNAs de *Eucalyptus* sp.

Oito clones de cDNA pré-selecionados do banco de dados do Genolyptus (Tabela 1) foram caracterizados por sequenciamento. Os agrupamentos contíguos consensos obtidos variaram de 711 a 1.888 nucleotídeos (Tabela 2). Os clones EUGR-PU-000-005-H06-MG e EUGR-PU-003-001-C10-MG não apresentaram contíguo único. Suas sequências formaram dois agrupamentos cada um, mesmo utilizando sequências geradas pelo sequenciamento com oligonucleotídeos que anelavam internamente no inserto e subclonagem, visando obter sequências mais internas do inserto (Anexo 2). Esses clones de cDNA possuem insertos com tamanhos estimados de 2.400 e 1.900 nucleotídeos, respectivamente. Assim, novos oligonucleotídeos necessitam serem sintetizados para completar o sequenciamento desses clones.

O EST relacionado com o clone EUGR-PU-000-001-G07-MG, quando submetido ao *Blastx-nr*, mostrou similaridade com o gene L^9 de linho (*L. usitatissimum*) que codifica a proteína L^9 de resistência à ferrugem do linho (e-value $1e^{-24}$, Tabela 1) (Ellis et al., 1999). O agrupamento contíguo desse clone (1.888 nt) mostrou maior similaridade com o gene *Nbi-B de linho* que codifica a proteína Nbi-B (e-value $2e^{-109}$, Tabela 2), que pertence ao loco complexo N, também relacionado à resistência à ferrugem do linho (Dodds et al., 2001). A similaridade desse agrupamento contíguo com o gene L^9 foi menor (e-value $2e^{-104}$, *ranking* do *Blastx-nr* não mostrado).

O mesmo aconteceu com o EST relacionado ao clone EUGR-PU-000-005-H06-MG que mostrou similaridade com o gene *Nbi-B* de linho, componente do loco N (e-value $4e^{-22}$, Tabela 1) (Dodds et al., 2001). O agrupamento contíguo desse clone (1.135 nt), mesmo não sendo contíguo único (Tabela 2), mostrou similaridade com a sequência do gene *N2-B* (e-value $1e^{-31}$),

relacionado à resistência a ferrugem do linho (Dodds et al., 2001), enquanto que a similaridade a *Nbi-B* foi menor (e-value $3e^{-31}$, *ranking* do *Blastx-nr* não mostrado).

O EST do clone EUPE-XY-000-035-C06-MG mostrou similaridade com o gene L^1 de resistência a ferrugem do linho (e-value $4e^{-24}$, Tabela 1), um dos alelos do gene *L* (Ellis et al., 1999). O agrupamento contíguo obtido (1.641 nt) foi similar à seqüência codificada por um gene do loco *N* (proteína *Nbi-B*) (e-value $8e^{-20}$, Tabela 2), também relacionada à resistência à ferrugem de linho (Dodds et al., 2001), e menor similaridade com o gene L^1 (e-value $9e^{-13}$, *ranking* do *Blastx-nr* não mostrado).

O EST referente ao clone EUPE-XY-000-048-H05-MG apresentou similaridade com a seqüência do gene putativo *60I2G11* de *P. deltoides* que codifica uma proteína TIR-NBS-LRR (e-value $1e^{-19}$, Tabela 1) (Lescot et al., 2004) e foi selecionado para sequenciamento completo. O agrupamento contíguo (1.659 nt) apresenta uma ORF interna na seqüência similar a subunidade beta do proteossoma 20S *PBG1* de *Arabidopsis thaliana* (e-value $1e^{-110}$, Tabela 2) codificada pelo gene *PBG1* (Fu et al., 1998), enquanto que na região 5'-terminal há uma seqüência similar à um gene que codifica uma proteína TIR-NBS-LRR de *P. deltoides* (e-value $4e^{-19}$, *ranking* do *Blastx-nr* não mostrado).

A diferença de resultados de similaridade das proteínas R entre as Tabelas 1 e 2, evidenciada acima., deve-se à diferença entre os tamanhos das seqüências utilizadas na comparação com o banco mundial de dados *GenBank* (NCBI). Os agrupamentos contíguos, com tamanhos de seqüências maiores (Tabela 2) que os ESTs (Tabela 1), apresentam similaridades mais significativas, porém em ambas as comparações não deixaram de apresentar similaridades com proteínas R por possuírem regiões conservadas com os domínios TIR e NBS.

As seqüência de cDNAs similares à genes *R*, foram traduzidas e alinhadas com a respectiva proteína mais similar estatisticamente (menor valor esperado ou E-value) e também com proteínas de *Populus* (planta lenhosa modelo para estudos biológicos em árvores). Nenhum cDNA possui a seqüência completa de um gene R, como verificado para o clone EUGR-PU-003-020-C11-MG (Figura 4).

Os alinhamentos mostraram sobreposição apenas na região TIR-NBS das respectivas proteínas R, mas não na região LRR (figura 4), demonstrando que os clones de cDNA são truncados. Exceção apenas para o clone EUPE-XY-000-048-H05-MG que não é similar à sequência de gene R, mas à sequência do gene que codifica a subunidade beta do proteassoma 20S (*PBG1*). Os supostos motivos codificados pela sequência deste clone foram alinhados separadamente, pois não há no banco mundial de dados do *GenBank* proteassomas contendo um domínio TIR. A sequência que codifica o domínio TIR associado a sequência que codifica *PBG1* contém os códons de iniciação e de terminação sem interrupção (Figura 5A), podendo se tratar de uma nova proteína.

A possibilidade de a sequência obtida corresponder a uma quimera formada durante a construção da biblioteca é pouco provável, pois não há sítios da enzima *NotI*, utilizada na construção, internamente a esta sequência. Além disso, a partir do códon de iniciação os dois motivos encontram-se *in frame* (Figura 5A).

A sequência de aminoácidos deduzida da sequência do clone EUGR-PU-003-042-F06-MG, similar à proteína Rar1, foi alinhado com as proteínas Rar1 de cevada (Liu et al., 2002) e de tabaco (Shirasu et al., 1999). Essa sequência possui ORF completa que codifica a proteína Rar1 (Figura 6).

Tabela 2 – Agrupamentos contíguos de cDNAs cujas sequências possuem similaridade com genes *R* ou ao gene *Rar1*, relacionado com resposta de defesa. Existem dois cDNAs com dois agrupamentos contíguos cada (EUGR-PU-000-005-H06-MG e EUGR-PU-003-001-C10-MG).

cDNA	Proteína	Acesso	Valor esperado	Tamanho do cDNA (nt)
EUGR-PU-000-001-G07-MG	Proteína Nbi-B [<i>Linum usitatissimum</i>]	gi 13509232 emb CAC35336.1	$8e^{-109}$	1888
EUGR-PU-000-005-H06-MG	a. Proteína hipotética [<i>Arabidopsis thaliana</i>]	gi 7268216 emb CAB77743.1	$2e^{-09}$	1002
(dois agrupamentos contíguos)	b. Proteína N2-B [<i>L. usitatissimum</i>]	gi 13509225 emb CAC35332.1	$1e^{-31}$	1135
EUGR-PU-000-010-D04-MG	Proteína de resistência à ferrugem [<i>L. usitatissimum</i>]	gi 4588070 gb AAD25976.1	$3e^{-27}$	746
EUGR-PU-003-001-C10-MG	a. sem similaridade no banco mundial de dados (NCBI)	-	-	689
(dois agrupamentos contíguos)	b. Proteína TIR/NBS/LRR [<i>Populus deltoides</i>]	gi 40644187 emb CAC95120.1	$6e^{-32}$	714
EUGR-PU-003-020-C11-MG	Proteína M de resistência à ferrugem [<i>L. usitatissimum</i>]	gi 1842251 gb AAB47618.1	$1e^{-34}$	711
EUGR-PU-003-042-F06-MG	Proteína Rar1 [<i>Hordeum vulgare</i>]	gi 6581046 gb AAF18432.1	$1e^{-80}$	1171
EUPE-XY-000-035-C06-MG	Proteína Nbi-B [<i>L. usitatissimum</i>]	gi 13509232 emb CAC35336.1	$8e^{-20}$	1641
EUPE-XY-000-048-H05-MG	Subunidade beta do proteasoma 20S PBG1 [<i>A. thaliana</i>]	gi 3421123 gb AAC32074.1]	$1e^{-110}$	1659

(A)

> EUPE-XY-000-048-H05-MG (1.659 nt)

```
5' CTTCCTTGCGCCATCGTCGTTTGTTCAGATCAAAACAGTTTACGCCAGGATGGAATCGCCGATAATCGAGACGAGGGGACAGCTT
TCGGATCTAGGGGGAAATCCTTCTGCGGAGCTGAAAGCGGAGCGCAACCGGAAAGGGAAGCGCCGAGCGGGATGTACGACGTGTTCT
TGAGTTTCAGAGGGCGGGACACTCGCAACAACCTCCACCGACCACCTCTACCGGACCATGGAGCGCCAGTTGACCGCCGCTAACAGAGA
CAGCGAAGAGCTTACGCGCGGTGATCGGATCGACGATCCCTGAGAGCGGTTCGGCGAACCAGGATCCTCATACCGTTCCTCTCCGAA
GATTATCTTTGGAGCGATCGGCGCCTCTCCAAGTCTTCCGGATGACGAAGCTCAGCGAATCGGCCAGGAAGCGGGAGATTCTACCCA
TCTTCTTTAAGGTCACGCCCTCGGACGTCAGGCTCAGGACGGAAGTGTACTGAAGGACTCGGAGAAGCATGAGCAGAGGTACCCAG
GGTGCATGGCAAGGATTTGAAGGATTATGTGGAACAAGTCCCGACTCCCGATCTGGCGACTCTGAAAGAACTCTGTATCCATATGTG
ACGGGTACGCTCTGTCTGGCTCTCAAGTACAGGGATGGGATTCTGATGGCTGCTGACATGGGAGGTTCTTATGGATCCACCTTGAGAT
ACAAAACCTATTGAACGAATAAAAACCTATTGGAAAACATTCTCTCTGGGTGCGAGTGGAGAGATAAGTGACTTTCAGGAGATATTGCG
TTACCTTGTATGAACTGATTCTTTACGACAATATGTGGGATGATGGGAATTCCTTGGGGCCAAAGAGGTGCATAATTATCTGACTCGG
GTTATGTATAATAGGCGGAACAAGTTCAACCCATGTGGAACCTCACTAGTCTAGGTGGAGTTAAGAATGGACAAAATATCTTGGCA
TGTTAGCATGATAGGCGGTGAGTTTGAAGAAAACATGTGGCTACTGGGTTTGGCAATCACCTTGCACGGCCAATTCTTCGCGATGA
GTGGAACGAGAATTGAGTTTGAAGATGGTGTAAAGCTGTTAAGCTGTTGAGAAAGTGCATGCGCGTCTTGTATCGTGACAGATCTGCCA
AATAAGCTTCAGATTGTAAAAATCACAGAAGAAGGAGCAACAAATTTCCGCCCTACTCTTTGAAAACCTTCTGGGGCTTTGAATCGT
ATGAAAACCCACTGTTGGTGAAGAAGGACCTCATTGGAAGATGACCAACTAGAAATAAAGCTGGATGAAGTCTGATTCCGACCCAGCA
CGTTAGCTGGGATATGTCGTTGTGTTTGTCTCGCAGCTCCGATATAGCGGGGATATGTCCTTGTGTTATTTCTTCTCAGCTATGTAT
TTTAGCAGTCTATCCTTCCGTGGCCTTTCATGGGCGTCGGTGGTCTGGCTGTCGGCGTCGGTGGTTTGGTGGCGCAATTAGATGGGAT
GTGGGCATCTCTTCTGATTCCTTTCTGTTAACGTCTGAAACATATGGAGCTTTATGTCCCCCTAATATTTGGAACCTTTATGTCCCAA
ACTTCGAATTTCTGTGTTTCCGCTTGCCCATGTGTTTTCAGCGTCTCTCACATTCAAACAAAAAATAAAAAA 3'
```

(B)

3'5' Frame 3

```
S L R P S S F V R S K Q F T P R Met E S P I Met E T R G Q L S D L G G N P S A E L K A E
R E P E R E A P S G Met Y D V F L S F R G R D T R N N S T D H L Y R T Met E R Q L T A A
N R D S E E L H A G D R I D D P L R A V G E P K I L I P F L S E D Y L W S D R R L S K F
F R Met T K L S E S A R K R E I L P I F F K V T P S D V K L R T E L Y L K D S E K H E Q
R Y P R V H G K D L K D Y V E Q V P D S R S G D S E R T L Y P Y V T G T S V L A L K Y R
D G I L Met A A D Met G G S Y G S T L R Y K T I E R I K P I G K H S L L G A S G E I S D
F Q E I L R Y L D E L I L Y D N Met W D D G N S L G P K E V H N Y L T R V Met Y N R R N
K F N P L W N S L V L G G V K N G Q K Y L G Met V S Met I G V S F E E N H V A T G F G N
H L A R P I L R D E W N E N L S F E D G V K L L E K C Met R V L L Y R D R S A V N K L Q
I V K I T E E G A T N F P P Y S L K T F W G F E S Y E N P T V G E E G P H L K Met T N
Stop N K A G Stop S L I R T Q H V S W D Met S L C L L R S S V L A G I C P C C I S F S A
Met Y F S S L S F R G L S W A S V V W L S A S V V W R N Stop Met G C G H L F Stop F L
F C Stop R L K H Met E L Y V P L I F G T L C P K L R I S C V S A L P Met C F S V L S H S
N K K K K K
```

(C)

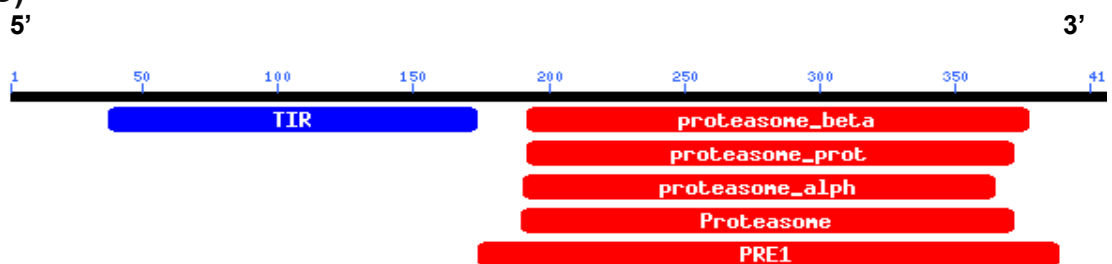


Figura 5 - Sequência de nucleotídeos e de aminoácidos (3'5' Frame 3) do clone EUPE-XY-000-048-H05-MG. Este clone apresenta similaridade à sequência de um gene R que codifica um domínio TIR de *P. deltoides* associada a uma sequência similar à subunidade beta do proteasoma 20S de *A. thaliana*. (A) sequência do contíguo consenso; (B) possível localização da metionina de iniciação (Met) e do códon de terminação dessa subunidade revelado pelo *ExPasy Translate Tool* (www.expasy.org/tools/dna.html); (C) figura representativa dos domínios relativos identificados pelo programa *Conserved Domain Search* (www.ncbi.nlm.nih.gov). ATG = códon de iniciação; TAG = códon de terminação.

```

prt-02          STLPERALKLPRGIFSIILLASRLAQSLTHSLALTRPHLRPSSGAARMMAKVRCQRIGCEA
Rar1-Tabaco    -----MERLRCQRIGCNA
Rar1-Cevada    -----MSAETERSAAAPAPAPMRCQRIGCDA
                                     :*****:*

prt-02          SFSEDDNPEGSCRYHPAGPIFHDGMKEWSCCKKKSHDFSEFLGIPGCKTGKHTTEKPSTK
Rar1-Tabaco    TFTEDDNPENSCTYHESGPLFHDGMKKWSCCKKSSHDFSLFLEIPGCKIGKHTTEKPVIA
Rar1-Cevada    MFTDDNPDGSCHYHPSGPLFHDGMKEWSCCKQRSHDFSLFLAIPGCATGKHTTEKPVTK
                                     *:*****:.* ** :**:*:*:*:*:*:*:*: ***** ** **** *****

prt-02          AVTATKTPSAATSAATSAAPATNASSKESCARCRQGFFCSDHGSRTGKTNIDVAN-ANAE
Rar1-Tabaco    NPAANRN-----RAIPAPTSTTNVSPKEACPRCRQGFFCSDHGSQPREVIPKASNTVTSV
Rar1-Cevada    AVSLNSSKATPPKSAPVQSSQ-GVETEACSRCRQGFFCSDHGSQPKAQKPVAVNGTNTE
                                     : . . : . . : . *:*.******:.* . * ..:

prt-02          PAVESSKDAQISVVPQKRIIDINEPQTCNKKGCGQTFKEKDNHEIACSFHPGPAVFHDRM
Rar1-Tabaco    PSESNTDIQQCHPAPVKKKVDINEPQICKNKGCGKTFTEKENHDTACSYHPGPAIFHDRM
Rar1-Cevada    SVQKSS-----VPEPKKVVDINEPRVCKNKGCGKTYKEKDNHDAACDYHPGPAVFHDRN
                                     . . .: * * :*****: *****:*.**:*:*: **.******:****

prt-02          RGWKCCDVHVKEFDEFMTIPPCTKGWHNADPASGALRCSNKVLYGLYRELILEWLLSFAV
Rar1-Tabaco    RGWKCCDIHVKEFDEFMSISPCTTGWHNASPAS-----
Rar1-Cevada    RGWKCCDVHVKEFDEFMEIPCTKGWNADAV-----
                                     *****:***** *.*.*.*****...

prt-02          KIPLFSLIDMVSTMSRRYPASQLLALYLSLLLLCCETFLDSGFFLLKQFSRSFLVREV
Rar1-Tabaco    -----
Rar1-Cevada    -----

prt-02          PKKKKKKKKKKKKKKKKKKKKKK
Rar1-Tabaco    -----
Rar1-Cevada    -----

```

Figura 6 – Alinhamento da sequência de 382 aminoácidos deduzida da sequência de nucleotídeos do clone EUGR-PU-003-042-F06-MG (prt-02) alinhada com as proteínas Rar1 de tabaco (221 aa) e cevada (232 aa) (gi|21552979| gb|AAM62409.1| e gi|6581046|gb|AAF18432.1|, respectivamente). Este clone contém uma ORF completa que codifica a proteína Rar1 (azul = dois motivos conservados CHORD de ligação a zinco, vermelho = metionina de iniciação).

3.2. Identificação de clones BACs correspondentes aos cDNAs

Oligonucleotídeos RGs (*forward* e *reverse*) foram usados na triagem da biblioteca BAC, mas três deles não identificaram clones positivos para RGAs (Tabela 3). Como esses oligonucleotídeos são capazes de amplificar sequências do clone G21 (*E. grandis*), é provável que as regiões genômicas correspondentes a estes cDNAs não estejam representadas na biblioteca.

Tabela 3 – Clones positivos resultantes da triagem com oligonucleotídeos RGs em biblioteca BAC. O tamanho dos insertos foi estimado por eletroforese de campo pulsado (PFGE) após digestão dos clones com a enzima de restrição *NotI*.

Clones	Oligonucleotídeos (<i>forward</i> e <i>reverse</i>)	Clone BAC	Tamanho aproximado dos insertos (kb)
EUGR-PU-000-001-G07-MG	RGA03	14-A08	170
EUGR-PU-000-005-H06-MG	RGA09	14-H12	155
		138-H11	160
		153-G11	70
EUGR-PU-000-010-D04-MG	RGA11	208-G01	20
EUGR-PU-003-001-C10-MG	RGB05	135-D03	180
		137-E09	110
		140-A07	280
		143-F12*	90
		171-F01	80
EUGR-PU-003-020-C11-MG	RGC05**	-	-
EUGR-PU-003-042-F06-MG	RGD06**	-	-
EUPE-XY-000-035-C06-MG	RGE12	27-C02	170
EUPE-XY-000-048-H05-MG	RGF08**	-	-

* Clone selecionado para construção de biblioteca shotgun;

** Oligonucleotídeos que não identificaram clones positivos.

Foram identificados onze BACs positivos para os diferentes oligonucleotídeos RGs (Tabela 3). Amplificações por PCR utilizando cinco oligonucleotídeos e todos os 11 clones identificados confirmaram a especificidade de cada oligonucleotídeo, como verificado na amplificação da figura 7. Em todas as amplificações, o tamanho do produto de amplificação foi o mesmo para o clone BAC e o DNA genômico de G21. Em um único caso, onde foram utilizados os oligonucleotídeos RGB05 *forward* e *reverse* (F e R), identificaram-se dois produtos de amplificação no DNA genômico G21,

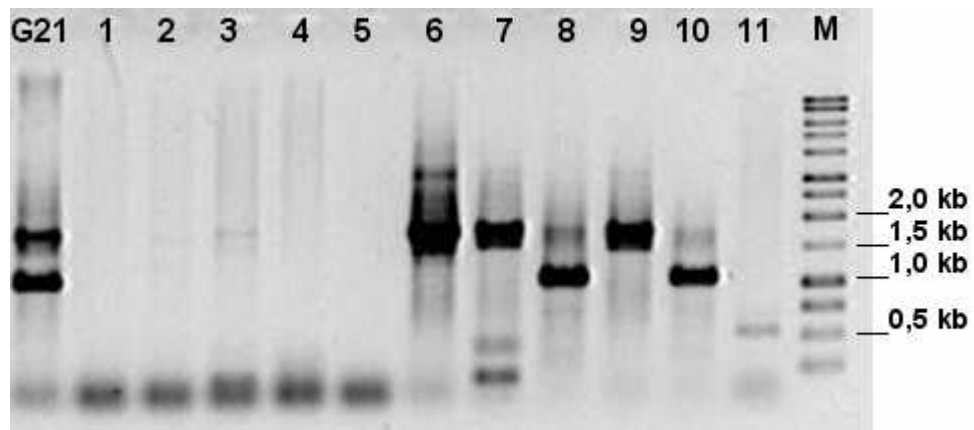


Figura 7 – Análise do produto de amplificação do DNA de todos os clones BACs selecionados utilizando o oligonucleotídeo RGB05 (F e R). As canaletas de 6, 7, 8, 9 e 10 representam o padrão de amplificação dos clones 135-D03, 137-E09, 140-A07, 143-F12 e 171-F01, respectivamente, revelados pela triagem da biblioteca BAC de *Eucalyptus grandis*. A primeira canaleta (G21) representa os produtos de amplificação do clone G21 (dois locos com 1 e 1,5 kb) e a última canaleta (M) representa o padrão de comprimento de DNA 1 kb DNA Ladder (Promega).

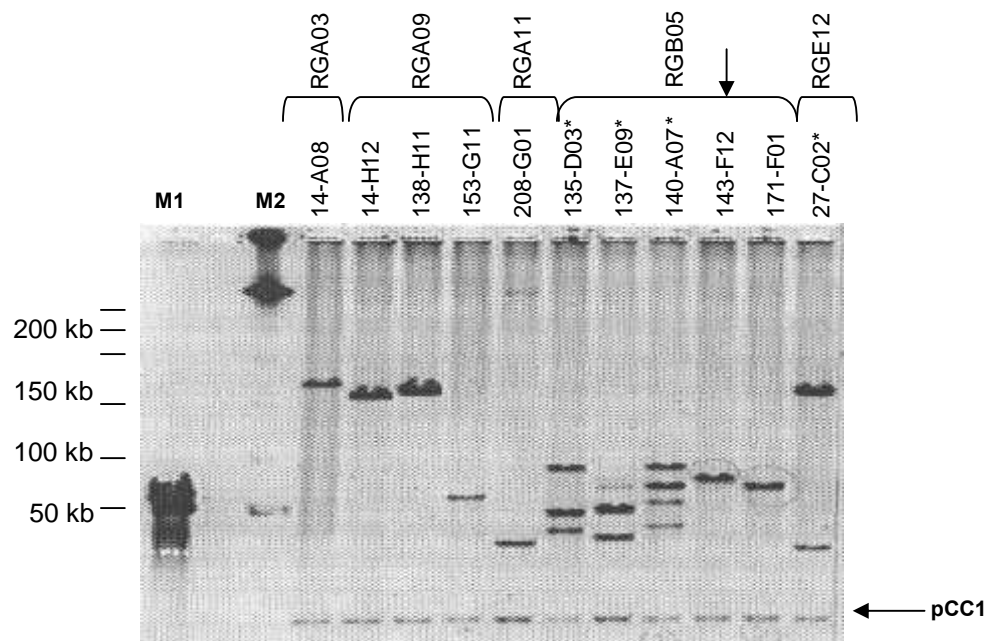


Figura 8- Tamanho do inserto dos clones BAC separado por eletroforese de campo pulsado (PFGE) em gel de agarose 1%. Os insertos foram liberados do vetor pCC1 por digestão com *NotI* (Invitrogen). As bandas inferiores, comuns a todas as canaletas, representam o fragmento do vetor pCC1 linearizado (8,12 kb). M1 e M2 representam os padrões λ *HindIII* e λ *Ladder* (Biorad), respectivamente. (*) Clones com sítio interno para *NotI*; (↓) Clone selecionado para construção da biblioteca *shotgun*.

de 1 e 1,5 kb, indicando que este par de oligonucleotídeo foi capaz de amplificar seqüências de locos distintos. Corroborando para este resultado, clones BAC equivalentes a esses dois locos foram identificados na triagem (Figura 7).

O tamanho aproximado dos insertos, avaliados por PFGE, variou de 20 a 280 kb, sendo que quatro apresentam sítios internos para enzima *NotI* (135-D03/RGB05, 137-E09/RGB05, 140-A07/RGB05 e 27-C02/E12) (Figura 8). Dentre os clones amplificados, o BAC 143-F12 (aproximadamente 90 kb), derivado da amplificação com o oligonucleotídeo RGB05 (F e R), foi o selecionado para construção da biblioteca *shotgun*. É um fragmento que possui chance de conter pelo menos um gene que pode ser um RGA, já que em *E. grandis* estima-se a ocorrência de um gene a cada 20 kb. Fragmentos maiores que 150 kb podem dificultar o estudo em função de possuir muitos genes além do que se propõe e pela necessidade de gerar um grande número de clones na biblioteca *shotgun*. Já fragmentos menores que 50 kb têm maior probabilidade de conter parcialmente um gene.

3.3 Sequenciamento da biblioteca *shotgun* do BAC 143-F12 e análise das seqüências obtidas

Foram sequenciados parcialmente 960 subclones do BAC 143-F12 e foram obtidas 917 seqüências (*reads*) com tamanho médio de 555 bases. Excluindo-se as seqüências do vetor, isto equivale a uma cobertura de aproximadamente quatro a cinco vezes o tamanho do clone BAC 143-F12. Apesar dessa cobertura, as seqüências não se alinharam em um contíguo único. Foram obtidos 14 agrupamentos contíguos sendo nove similares a genes que codificam diferentes proteínas conhecidas de plantas: quatro tiveram similaridade com seqüências que codificam fosfatidilinositol 3-quinase putativa (CT03, CT06, CT08 e CT10); um a poligalacturonase putativa (CT11); dois similares a poliproteínas putativas (CT09 e CT14); e dois a proteínas TIR-NBS-LRR (CT12 e CT13) (Tabela 4 e Anexo 3). Cinco agrupamentos contíguos não apresentaram similaridade significativa com seqüências depositadas *GenBank* (NCBI).

Quatro contíguos tiveram similaridade com seqüências que codificam fosfatidilinositol 3-quinase (PI3-K) putativa de cevada (CT03), de soja (CT06) e

de *Arabidopsis* (CT08 e CT10) (Tabela 4 e Anexo 3). A sequência de aminoácidos deduzida dos agrupamentos CT08 e CT10 tiveram similaridade com a mesma proteína PI3-K putativa de *Arabidopsis*, porém referem-se a regiões diferentes desta proteína (Tabela 4). O CT08 alinha na porção inicial da proteína (2-103) e o CT10 uma porção posterior (103-368). PI3-K fosforila especificamente fosfatidilinositol na posição D-3 do anel inositol para gerar fosfatidilinositol 3-fosfato (PI3-P) (Drobak & Heras, 2002). Evidências sugerem que PI3-K esteja envolvida em processos celulares como cascata de sinalização RAS (que são proteínas G com função ativadora de várias vias moleculares), crescimento celular, tráfego e secreção celular (Odorizzi et al., 2000). Estudos recentes mostraram que PI3-k pode mediar eventos celulares durante o movimento estomático e também participar em funções nucleares (Bunney et al., 2000; Jung et al., 2002; Hernández et al., 2004).

O contíguo CT11 foi similar a sequência que codifica poligalacturonase (PG) putativa de *Arabidopsis* (Tabela 4 e Anexo 3). PG é uma das enzimas envolvidas na solubilização e redução do tamanho das cadeias de pectina, um dos principais componentes da parede celular (De Veau et al., 1993).

Tabela 4 – Agrupamentos contíguos da biblioteca de *shotgun* do BAC 143-F12. O resultado do *Blastx-nr* possibilitou verificar que os agrupamentos possuem similaridade com quatro tipos de genes, sendo dois deles com similaridade a gene R. (* sequência inversa).

Agrupamento Contíguo	Proteína	Acesso	Valor esperado	Coordenada Query (nt) / Sbjct (aa)
CT03 (3 reads-1.066 nt)	Fosfatidilinositol 3-quinase [<i>Hevea brasiliensis</i>]	gi 50404258 gb AAT76869.1	$1e^{-18}$	+1 = 676-903 / 112-188 +1 = 43-138 / 94-125*
CT06 (16 reads-904 nt)	Fosfatidilinositol 3-quinase putativa [<i>Glycine max</i>]	gi 7434344 pir T07745	$1e^{-14}$	-2 = 133-2 / 573-616
CT08 (24 reads-2.565 nt)	Fosfatidilinositol 3-quinase putativa [<i>Arabidopsis thaliana</i>]	gi 24030195 gb AAN41278.1	$3e^{-08}$	+2 = 2099-2200 / 70-103 +2 = 665-748 / 43-70 +1 = 289-399 / 2-38
CT09 (28 reads-3.175 nt)	Poliproteína [<i>Oryza australiensis</i>]	gi 2443320 dbj BAA22288.1	e^{-134}	+3 = 126-1829 / 334-866 +2 = 1802-2296 / 858-1022* +3 = 2295-2660 / 1060-1182
CT10 (42 reads-4.899 nt)	Fosfatidilinositol 3-quinase putativa [<i>A. thaliana</i>]	gi 24030195 gb AAN41278.1	$4e^{-49}$	+1 = 445-654 / 156-225 +2 = 194-355 / 103-156* +3 = 3771-3950 / 285-344 +3 = 2328-2504 / 226-285 +2 = 4313-4420 / 333-368
CT11 (61 reads-5.931 nt)	Poligalacturonase putativa [<i>A. thaliana</i>]	gi 42562835 ref NP_564758.2	$1e^{-77}$	-1 = 3446-2937 / 277-409 -3 = 2814-2416 / 409-539 -1 = 4673-4530 / 139-186 -3 = 4377-4228 / 181-232
CT12 (175 reads-16.712 nt)	Proteína Ngc-D [<i>Linum usitatissimum</i>]	gi 13509207 emb CAC35321.1	$3e^{-71}$	+2 = 2432-3595 / 307-638 +2 = 500-946 / 31-184 +3 = 4401-5510 / 648-1068 +3 = 2052-2333 / 181-276 +3 = 4449-5399 / 789-1076 +3 = 3705-3818 / 148-685*
CT13 (216 reads-20.788 nt)	Proteína N1-D [<i>L. usitatissimum</i>]	gi 13509221 emb CAC35330.1	$7e^{-74}$	+2 = 1712-2848 / 185-557 +2 = 470-913 / 31-183 +1 = 2980-4464 / 555-1063
CT14 (240 reads-18.332 nt)	Poliproteína putativa gag-pol [<i>O. sativa</i> cv. japónica]	gi 31249704 gb AAP46197.1	$1e^{-96}$	+3 = 2757-3223 / 408-596 +1 = 3325-3996 / 635-858* +1 = 1825-1983 / 228-289 +2 = 2633-2770 / 367-412* +3 = 3996-4079 / 859-886*

Os contíguos CT09 e CT14 apresentaram similaridade com sequências que codificam poliproteínas putativas em diferentes espécies de arroz (Tabela 4 e Anexo 3). O contíguo CT09 apresenta similaridade com a sequência do retrotransposon *RIRE1* (*Rice RetroElement*) de *Oryza australiensis* (espécie selvagem) e o contíguo CT14 possui similaridade com a sequência da poliproteína gag-pol putativa (parte de um retrotransposon) de *O. sativa* cv. japonica (espécie cultivada). Retrotransposons são elementos móveis que se movimentam no genoma por meio de uma molécula intermediária de RNA que sofre transcrição reversa. As sequências do CT09 e CT14 são similares a retrotransposons do grupo Ty1-*copia* que possuem uma região LTR (*long terminal repeat*) que se localiza nas extremidades da sequência, o qual não codifica nenhuma proteína conhecida, mas que contém promotores e terminadores associados à transcrição dos retrotransposons LTRs. Na região intermediária, flaqueada por dois LTRs, existem dois dos três maiores genes dos retrotransposons LTRs: *gag* e *pol*. O gene *gag* codifica uma proteína envolvida na maturação e empacotamento do RNA, além de codificar proteínas que auxiliam na integração de sua sequência no genoma. Proteínas codificadas pelo gene *pol* são poliproteínas que são clivadas em peptídeos funcionais como protease, transcriptase reversa, RNase H e integrase requeridas pelo retrotransposon. Estas são auxiliam na transcrição do elemento de transposição, dando origem a um RNA, no seu empacotamento em partículas, na conversão do RNA em DNA fita dupla e na inserção desse cDNA em um novo sítio genômico (Noma et al., 1997; Kumar et al., 1999).

Dois contíguos apresentaram similaridade com genes do loco N de linho que codificam proteínas da classe TIR-NBS-LRR. O contíguo CT12 e CT13 foram similares aos genes que codificam as proteínas Ngc-D e N1-D, respectivamente (Tabela 4 e Anexo 3). Em linho existem cinco locos relacionados à resistência à ferrugem: *K*, *L*, *M*, *N* e *P*. Três genes, (*N*, *N1* e *N2*), com diferente especificidade de resistência a ferrugem, estão localizados no loco N, e foram identificados em diferentes genótipos de linho (Islam & Mayo, 1990) Uma baixa frequência de recombinação foi encontrada entre os genes *N* e *N1*, sugerindo que seja um loco complexo (Flor, 1965). No isolamento de genes TIR-NBS-LRR derivados de uma sequência genômica de 15 kb da linhagem *Forge* de linho, Dodds et al. (2001) encontraram quatro candidatos a genes *N* (*Ngc-A*, *Ngc-B*, *Ngc-C* e *Ngc-D*). Introduzindo o gene

Ngc-D em linhagem suscetível a raças de ferrugem reconhecidas pelo gene *N* de linho (*Ward*), esses autores demonstraram que a presença desse gene foi suficiente para conferir resistência a várias raças de ferrugem do linho, o que não foi comprovado para os genes *Ngc-A*, *Ngc-B* e *Ngc-C*.

A ORF contida no contíguo CT12 é dividida em oito éxons e oito íntrons, incluindo um íntron dentro da sequência que codifica o domínio NBS (Figura 9). A sequência dos íntrons obedece à regra de GT/AG, que são conservadas em suas extremidades. Para o contíguo CT13, representado na figura 10, a ORF apresenta quatro éxons e três íntrons, que também possuem as sequências conservadas GT/AG em suas extremidades. Essa ORF apresenta tanto o códon de iniciação (ATG) quanto o de terminação (TGA).

A sequência dos contíguos CT12 e CT13 contém sequências que codificam para os domínios conservados de proteína R. O agrupamento CT12 possui sequência que codifica uma ORF que contém domínios TIR e NBS (NB-ARC), e um domínio LRR truncado (Figura 11A). Já o agrupamento CT13 possui uma ORF típica de uma proteína com os domínios TIR-NBS-LRR (Figura 11B).

Uma comparação, por meio de alinhamento, foi realizada entre a sequência de aminoácidos do CT13 e seis proteínas com maior similaridade com a ORF encontrada nesse agrupamento contíguo (Tabela 5) para verificar a conservação dos domínios TIR-NBS-LRR. Foram observados aminoácidos conservados no agrupamento contíguo C13 relacionados com estes domínios (Tabela 6 e Anexo 4).

A subclasse de proteínas com domínios TIR-NBS-LRR é a mais frequentemente encontrada em dicotiledôneas (Lescot et al., 2004). Acredita-se que o domínio TIR, que normalmente possui 200 resíduos de aminoácidos, esteja relacionado com processos de sinalização celular devido a sua similaridade com domínios citoplasmáticos das proteínas *Toll* de *Drosophila* e *IL-1R* de mamíferos (Hammond-Kosack & Jones, 1997; Cannon et al., 2002; Martin et al., 2003). A região N-terminal do CT13 possui similaridade com sequências que codificam o domínio TIR, contendo 139 aminoácidos. Neste domínio foram encontrados os motivos TIR-1, TIR-2, TIR-3 e TIR-4, que são típicos desta região (Meyers et al., 1999; Young, 2000). A presença de domínios NBS sugere uma possível ativação de uma quinase ou atuação como uma proteína G (Hammond-Kosack & Jones, 1997). O domínio NBS do CT13,

com 305 resíduos de aminoácidos, incluem regiões altamente conservadas, como estrutura *P-loop* (GxxGxGK(T/S)T), quinase-2a (LxxxDDVW), quinase-3a (GxxxxTxR), e o domínio hidrofóbico (HD) (GLPL(A/T)L. (van der Biezen & Jones, 1998; Meyers et al., 1999; Young, 2000; Dinesh-Kumar et al., 2000; Dilbirlig et al., 2004). O domínio LRR, também identificado no CT13, possui de 9 a 41 repetições imperfeitas, com cerca de 25 aminoácidos consensos. É uma região com um menor grau de conservação quando comparado com a região N-terminal de uma proteína R, contendo motivos ricos em leucina (xxLxLxx) (van der Biezen & Jones, 1998; Young, 2000; Ellis et al., 2000; Cooley et al., 2000) (Tabela 6 e Anexo 4).

Tabela 5 - Proteínas com maior similaridade com sequências de aminoácidos deduzidas do agrupamento contíguo CT13 que foram utilizados no alinhamento com o programa *ClustalX* (1.83).

Proteínas	Valor esperado
.gi 13509221 emb CAC35330.1 Proteína N1-D [<i>L. usitatissimum</i>]	e^{-134}
.gi 22329811 ref NP_174037.2 Proteína putativa de resistência a doenças (classeTIR-NBS-LRR) [<i>A. thaliana</i>]	e^{-134}
.gi 10121908 gb AAG13418.1 AC000348_15 T7N9.23 [<i>A. thaliana</i>]	e^{-131}
.gi 37574597 gb AAQ93075.1 Proteína TIR-NBS-LRR [<i>Malus baccata</i>]	e^{-130}
.gi 40644189 emb CAC95124.1 Proteína TIR/NBS/LRR [<i>Populus deltoides</i>]	e^{-130}
.gi 51477390 gb AAU04763.1 MRGH8 [<i>Cucumis melo</i>]	e^{-128}

Tabela 6 – Aminoácidos conservados no agrupamento contíguo CT13.

Domínio	Motivo	Sequência
TIR	TIR-1	176 YEVFLNFRGADTR 188
	TIR-2	225 IGAELLQAIENSKIYIPIFSANYASSCWCLRELAYMAEC 263
	TIR-3	276 ILPIFLDVEPVVDVK 289
	TIR-4	325 WQKALIAVG 333
NBS	P-loop/Quinase1a	408 FLVIHGMGGVGGKTTTLTKVVFN 428
	Quinase-2 ^a	491 KVLIVLDDVD 500
	RNBS-B/Quinase-3	518 SSRIITTRY 527
	HD	591 GKLPLALEVTGSSLSGR 607
LRR	Motivo 1	930 LMRTPDLSMLVSLERLILEGCR 950
	Motivo 2	1010 LDALTEILMPHSSYAFKL 1027
	Motivo 3	1037 LLTLDVSHSWIGMLPDSIGNLKSLTL 1064

Classificação dos aminoácidos:

- Aromáticos	F, Y, W, H
- Alifáticos	I, V, L
- Hidrofóbicos	A, G, M, C, K, R, T
- Positivos	H, K, R
- Negativos	D, E

(A)

>CT12 (5.470 nt)

```
5' ATGGCCTTTTCCGGCGAACGCCGACCTCTCGACTGCTCTCTCCCTCCCGAGCTCACCTGCGAGACTCCCTGCCCCAGCTCAAGCT
CTACTTCTCATCGGAATCCTCAGCCTCTCTGCACTCTCCATCGCTCTCCTTATATACCTCTGCTCCTTCTCAGCCGAGCTCGAAG
AAGCGGAAGTCCACGAAAAGGGACGAATTCCTGAGAATCAGATGCACCTTTCTA CGAGGTTTTCTAAGTTTTCTATGGACAGAC
ACCGTCATGGATTCACCGATTTCTTTTACCGTGGCATGGTTGAGACCGGGATCTCGTCTTTAGGGATAATGAATCCCTCATGTCGG
TCAAAGAATCGGCGATGAACTTCTACAAGCCATCCAAACTCCAAGATCTACGTTCCCATATTTCTTGAGAAATATGCGTCGAGTCA
TGGTGCCTCCAAGAGCTTGCATATATGGTTGAGTGCACCTCCAATCTAATGGAAAAAAGAGATTTTGCCCATTTCTTTGGATGTGG
AGCCTGATGATGCAAGCTGAAAACAGACTTATACCCGAAAGACTTTTCAGCGCACCGAAAAAGTTCCGCTACTGAAGTTGAATCATG
GAAAAAGGCTCTCATTGAAGTGGACGAGATAAAGGGATGGAACTGGAGGAAAGATGAAGGgtaaatcttaaaaacttttcaattatcat
ttgtagatctcattttctctctatttagtttaattgacagtgtgaggatgaaaaattattatttcattctttcattttctctctat
tttctaatttttatttcgattttcaccgtgtttaagtatgaaagagagttttcttaaaaattgagataaaaaatataatcaatcgcataaa
aatcatgtgcaactacactgtaacgacaagctaataagagagataacagtttaacagaataaaagtgcaggaaatattgtataaacag
ttgaatgtttcagaataaatggttactaactgtacttttaaggggattttgtgtagtgcctcttgcatttcaactcctcacttctatg
gcagtgcgatttttactttcgatttcataagcattgaaataaaatacagaagataagggccagtggtgtaaggaatttagaat
gaaatgattacttaaaaaggtgaaaagctctcggcactattgtggcgatcgctcctagagCTCCACCAAGCCCGCCGGCCAGCCATCA
GCTCTCCGTCAGTCCCTGCCACTTCTTTGTCACCAGCCCGACACCAAGCCACAGAAAGCCGCGCGAGCCTTCTGGCCCCACCTT
TCTTCTCTCCTCAATGGCGAGTCCGATTCGGCTCGGATGTCGGgcttaagatcaccaagccatattccagcgtcgcgagcttg
ggctaccctcaaatctagTGTTAGCGAAGGCCAGCAAAGCCATCCCAGTGGTCGCAAGGCTGCAACCCACTTGACTTGAAGGAGAGGC
AAGGCCGAGATGGCGAGTGGATGGGGTGCATCAGGCCGGCCACTCGAATGTGACAGTGTGGCCCGGGAGGgtagcaacgattatc
ggcggttggaatggcaggggcccagcgggtgatgactaaattggggtagtggagatagactgggaaattttgattgttttagcg
tttagatgttttgcgtaattctcgtcaaatggaaggtataattgagcttagggagtgtttgtatcttaagctacttcttaga
acggtgtgtacgtttttgcagGCAAGCAGATCTGACACAATCCATAATTAACACTGTGTGGATAAGCTGAAAGTCAGATATAAAGG
ATTGTGCCTGAAGATCTAGTTGGAATTGACGATTGTGTGGAAAGCTATAATCAAAATGTTGGATTGAGATCTAGTAGCGTACAATTTTC
TCGGAATCCACGGAATGGGCGGAATGGCAAAACAACACTTGCCATATGTCATTTTCAACTGCCTGTCTCTCATTAAAGACTTTA
TTTCTTCCAGATGGTCCGAGAGTTGTCAACAAGCAATGGACTGCCATATTTGCAAAAATTTGGTATCTCAAGgttccttgggtctagt
tctcatcttgacaatattcaacacacggatgaggggataaacatgataaaagagattcttggggatagAAAAGTTCTCATCGTCTTGGGA
TGACGTGGACGGGAAAGAGCAACTCAAAGTCTAGCAAAGAAAGGTGATTGGTTTCGGTTCTGGTAGTAGGATTATCATAACTACTAGG
GACCAAAGGGTCTTAAGGATTGTGGGAGAAGCAACAGGTGAAGGCCATTTAGAAAAGTCTACAAAAGTTTTGACTTACGAAGTACGTG
AAGCTAGAATTTGAAGATGCTCTTAAGCTTTTTCGTAAGCATGCTTTAGAAGAGACTCTCCACAGATCATATAGGCTCCCTTTCAA
TGAGATTGTCTCCACTTTGGGAATGCTTCTTTAGCCCTTGAGGTTATAGGTTTCACTCCCTTAACAATGAACCCAAAGAATTCTGGGAA
GCCACGTTGAAGAGCTAAAAGATGCTCCTCCGATGAAGTCCAAGTAAATTGATGATATCTTTTGATAAGTTAAATAATGAACAAA
AGCAAGTATTTTGGATATAGCTTGTTTTTTGTAAATGAGGATAAAACATACCCTCTCATATGTGGGATGCCTATGGATA CCATCC
AGCAATTGCCATTAAGGTCTCTTGTCTCATGTCCTTGATCAAAATTAAGATGGCAATACTTTTCGATGCATGATCAAGTGCAGAC
CTTGAAGGAAATTTGACGTCAAGAGAACAAAGAACTCTGTAAGCGTAGTAGAGTGTGGAATCATGAGGATGGCTTGAACATTGTGA
AGCAAAAAGGgatagaaaggtatagcaatttaatttcatgtgaattttgttagattattcttgcaaaaggaatttaaatgcta
ttggcgcttgcaagtgtctgtatttgaccatcacacgtggactcttgggtgtttccattctcagGGAAGTAAGAAAATAGAAGCACTG
TCAGTTGGACTTCGGATTGAGCGGGAAGTCAATTATTATGAAGGATGATCAATTTGCTGATTTACGAAACCTGCGTCTTTCTGGGG
GAAGGTGTCTTCTGTTGGAGATTTCAATAATCTTCTCTCAAATTTAAGATGGTTTTCTAACTGTTATTGTTCTTCTATGCTTGGGC
AAAAACTTTTGTCCAACATAATTTAGTCTTCTTGAACCTTCGATCCAGCGACATTTTCAGAAAAAATAATGATTGGATCAAGgatga
tacgagctcactccttgtattcataaactcgtttcttttaattttttttcttttttaggtcatttctcaataaattgtatatacaaat
gtctcgtcaagtagcaaggaatttaaacgtcctagatcttaaggtgcttaaaccttaagagaaacactagatttctatctatgctggg
ccttggagattttgagcttatatagttgttgaattttggaaagtattgatttgattaaagtatgatacaagctccacttcttatacac
gtaaatcgcttcttttctattttttgtagGCCGTTTCTCAAAAATAATTTTATCGAATGTTTTGTCAGGCAGCAAGGAACTAAAAG
TCTTAGATTTTAGCTTCTGCTATAACTTAAGGAGATTGCCGATTTATATACGTTGGTGTCTTTGGAGATTTTGACATTAGCAAATG
TGCACTTTTGGAAAGTATCGATTGGATCGAAgtaagataggagtctcactccttgtattcataaactcatttcttttataattttttg
taggtcatttctcaataaataattattgttttgcagccaccaaggaactaaaagtcctagatctcagctactgccaacacttgaggag
attgcccgatattatctacgttggtgctcttggagattttgaaataaatagattgtgaacttttgaaagatcagattggatcgaagta
cgatacagagctcactccttgtattcataaataattttatatttttttaggtcatttctcaataaataattattgttttgcg
ggggcgaaggaataaaagtcttagatctcagcttatgctttaaacttaagagagtgctgatttaactacgttgggtctcttgggag
attttgagattatgattattgttcaaagATAAAGAACTTCCAGACTCGATTGGGAAATTACAATCATTAATCGAGTTGGATTTGTTAG
AAACGAATATTGATCACCTACCGGATTCATTCAACAATCTAAAGCAACTAACAACACTTAGGATGATGGACATTAAGGGAGGGTTGAC
AAAATTACCTAGTGCATCTGGTTTGTTCAGTTAATGGCTTATCTTTTTCGTTCCACGGTAATAAGCTGTACTGTTTGAATAATGCCA
GACTTGTATGACCCAAATATTGGATTACCCACCGTGGCAGTCTGCTCTAATATCCGTCACTCCGATGACCGGATGGGATGATA
GCAGACGGCCAACTTTTATTGGAGTACCTCGAGCGTAGCAATCTCAGAGGGAAAAACAAATATTGGCTCCCAACATGCGGTT
CCTCCTCCAGCTCCCGTCTAGTTTAAAGAGACTCAATATTTCGAGAAGTGTGGAGTCTGGGTAACGTGCATAATCGGGTCCGATATTC
GTTGGTATATGTGATTCCGAGTTGGAGAACTCCATCTCGAGGATTCACGCTTTCAAGACGGATTGAACAGTGTGGCTGAGAAACA
GGAAATCCCGTCCAGATACCGTCCGAGGATTCGTGACAGCCGATCAGCCACGAAAGCATTGAGCCGTCGCATATGAAGAATCT
GCAAGAGGTAGATCTGTTTGAATGCGAGCTACTAGTTCGAGATTTCGCGGCTCGAAGAATTGGGATCCCTTTTCACTCCCTTGGCTGCGA
TATGTAGTTCGATCGAAAGGATGCCAGACTATCGAAATCGAGGAAGCTAAGGAGCTTCGAGTGGGAAAATGTCGAAGCTAAGAA
GCATGGAAGGTTTGGACCGCTTGAATCTCTAAAGTGTTTGCGGATTCACGATTGCAGGCTTTTGGAGAGTCTGGTGGACACTCGAA
TTTTGATTTAGAATGTTGCTGATTGTCCGTTGCAAGCGGTTGCCGACCGTCCGCAATTTATTGTAACAGTGGCCGGGTTGCTATTGT
GAGATGCTTTTTCGAGGGGTTATGGAAATCTGACCCGATGGAAGTCCAGCAAGTATTACAGGGACACCCGGATGTTGCGCCAgtaagga
atttagaacatggcttggtttttgtgcttgggtgtagccattttgtttgtttcttctgttctgtttttttgttctatgaaaaagct
ttctcccggacaaata 3'
```



Figura 9 – (A) Sequência de nucleotídeos representando parte da região codificadora do CT12. Em minúsculo estão representados os oito íntrons e em maiúsculo estão representam os oito éxons. (B) Diagrama esquemático do CT12, onde os retângulos representam os éxons e as linhas os íntrons. (vermelho = sequência similar ao domínio TIR; azul = sequência similar ao domínio NBS).

>CT13 (4.226 nt)

5' **ATG**GAGAAAGGCACAAATTCACGGGCATCATCCGGATCTTCATA**TGAGGTTTTTTTGAATTTTCGAGGAGCAGATACTCGTAATGG**
ATTACCCGATTTTCTTTACCATGGGATGGTGGTGTGGAATCCTTGATTTAGGGATAGCGAATCCCTCCATGTFGGTGAAGAAAT
GGTGTGAACCTTCTACAAGCAATTGAGAACTCCAAGATTACATTTCCCATATTTCTCGCAAATTTCTGCAAAATTTGTTGGTGTCTTC
GAGAGCTTGCCTACATGGCCGAGTGCACCTTGAATCAAATGGGAACAAGGAGATTCTGCCTATTTTCTGGACGTGGAACCTGTTGA
TGTCAAGCTAAAAACAACTTATATAAACAAACTCTTTCAAAGCGCCAGAAGACGTTTTGCAGTGAAGCAGAGTCATGGCAAAAGGCT
CTCATTGCAGTGGGCAAGATAAAAGGATGGAACCTTGAAGAAAGATGAAAG**gt**gaatgtcttaaacttttcagttgttacgtagatttc
ttcttccccctaaattagttcgtttgcaagtattgttgggaaaattatagtttattcatctcgtcaagcatgggatttctatgggttag
tagccaatagcggccgacaccaatacgggtgggacacacctaccatttctaattctttttctcctctcactttctatctctgacata
gtagccacgaatagtaacatcttgagcagcttcaccaagccgcccagccagccactgctacaactcatccatcaatcatcgccatttct
ttgtcgccgcccgtgacaccaagcggccagaacctgcccgtgaccatccaaactcccacctttctatccttcaactcactggtaagtca
attcaggtgggtttgacgtcaaccttgccctatatcttgcaattgcccaccttggggaccccaaatctgggtctcctaaagccggcgat
tgcaaggccacgacctatgtgacgtggaggattgagataaggctgtgatggcgggctcgattgggtgtgatgggtcaaggactctggaat
gcaatggtgatagtgcaggaggcaagtgatgtcaacagtgccatgatgggtggcagaggctgacgggtgatgggtgatgaaagtgggga
gtggagagagattgaaaaaatcagaaattcttagcatctagatttttttcttgtaatttctcacaactgtgagggcaattgaaatca
cttgagaagtagctttttacaacataaagtgttgccttaacatggttttacatcatggc**ag**CCAAAGTAGATTTGATCCAGTAA
TTGAAGCTGTTTTGGTTAAGTTGAATGTCGGA TATAAAATGTGACT**TGAGCACCTAGTTGGAGTTGATGATCGTGTAGAAGCTATAAT**
CAAGATGTTGGACATGGAACTGATGATGTAATTTCTCGTGATCCATGGAATGGGAGGTGTTGGCAAACAACACTAACCAAAGTC
GTCTTCAACCGATTTTCTTCTCGGTTCCAAGGCTGCAATTTCTTTCCGGATGTTTCGAGAGTCATCAGAACGACATGGCCTGGTACATT
TTCACAAACAATTTGATTAATCAAGTTCCTTAATTGTTGTTGATGTTGACCAAAATCAGTGAAGTGGATCAATGGAATCAACATGATTAAGAG
GGTACTTTCATAATAAAAAAGTTCTCATTGTTCTTGACGATGTGGATGAAAAACAAGCAGCTGAAAAATTTGGCAGAAAAAGGAAATTTGG
TTCCGTTCTAGTAGTCGGATTATCATTACTACTAGATACAAAAGTGTCTAATGATTGATGGAGAACGACAAGTGAATGTCCTATAA
AGTGTGCAAAATGTTTTGATTTATGAAGTACAAGAAATGGAATTTCCGCCAGCTCTTGAGCTTTTCAGTAGGAATGCCTTCAGGAGAGA
CITTCACATCCGATGTTGATGATCAAGTGCAGGACCTCGGAAGGCAATCTGTTCCGTAAGAGAAATTTCAAAGAACCTCGCAACCTGAGCC
TCCTTTTCTGGTAGATCCAAGAATTTTGGGTAGACACTTTGAAGAAGCTAGAAAAAGCTCCTTCCATAGAAGTCCAAAAAACATTAA
GGATAACTTATGAAAGGTTAGATGATGAACAAAGGCCAAGTATTTTGGATATAGCTGTTTTTTTGTTAATGGGATAAAACCTATCC
TTGCTACATGTTGGGAGGACTGTGAATACCACCCGTACAATGCCATTGAAATCCTCCGTCTCATGTCTTGATAAAAAATCAAAGATGAC
AATACCATTTGGATGCAATGATCAAGTGCAGGACCTCGGAAGGCAATCTGTTCCGTAAGAGAAATTTCAAAGAACCTCGCAACCTGAGCC
GAGTGTGGAATCATCAGGAAGCCCTGAGCATTCTGGAGCAGAAAAAG**gt**gtcgaaagggaaatttcaacttattttctctgtaattgtt
gtggcctattcttgaaaaaaggaaaaaatcatccttattagcgttgaaattgtttgcttttgaccttcacatgtgacttcttggtgtt
tccatcatt**ag**GGAAAGCGAGGAGATAGAAGCACTGTCACTAGGATCTCGTGGAGTCATCGTAACGCCTGGTGAAGTTGCTAATTTACG
ACACTTGAGGTTCTTGGATGGGAATGAGATGTTCTTGTGGAGACTCAATAATCTTCTCTCAATTTAAGATGGCTTTCTTGGCGA
TGGTGCCTTTTCGAGTTTGTGGCAACAAACTTTTCTG**gt**caatttaactgcttcttgacctttcacatagctatgtatctgagaat
ggattggctggaaccaaatcaaagtatgatacagagtctcaactccttgatataagtatattttctgcttcttctcatttttttccct
ataaatcgtttcttaatatatatatattttt**gt****ag**GTGGCAAGTAAACTAAAAGTACTAGATCTCAGTAATTGCAA**CAACTTAATGAGA**
ACGCCTGATTTATCTATGTTGGTGTCTTGGAGAGATTGATTTCTTGAAGGTTGTCGCAATTTAATTAATAATCGACCCTTCCATCGGAA
AATTAAGCTCCTCAATGCTTTAAACTTGAATGGATGTCACGCTCTTGAAGGATTGCCTTACGAAATAGGATGTTTAGATGCTTTGAC
GGAGATTCTCATGCCTCATTCGTATATGCATTCAAACCTCCGGAGACGATTGTAATTCGAAGTCATTGTTGACCTTGATGATCA
CATAGTTGGATGGCATGCTACCAGACTCGATTGGTAATTTGAAGTCAATTTGACCTTGATGATACATAGCTGGATCGGCATCG
TGCCAGACTCAATTTGGAGGCTGGGAAACTTAGCCGTTGAAATTTAATTAAGTGTTTAAAGATAAAGAAACTTCGGACTCGGTTGG
GCAATTACAATCATTAGTCGAGTTGGATTGTCATCAACAAGTTTGGGACACTTACCTGATTCATCGGCAATCTAAAGCAATTGACA
GTACTTAGGATGAGCCACATAATTTGGGATAACAAAATTGCCAAGAGCGATTGGGTTGCTAGAGAAGCTTGAAGAGTTAGAGCTAGTG
GATGTCACCACCTGACCGCGAGATCCCTGAAGAAATTTGGGAGGTTGTCTCATTAAAGGATCTTAGACTTGTCACTCACACGTATATC
CAGATTACCCACTGTGCATTACCTCTTAATCTCAAACACTTAATCTAGAATCATGTTCTAAGCTTAATGAGTTGCCACAGCTT
CCCCAAGTTTGAAGTTGCTAAGGTGGACCTTTAATACTCTCAAGTTTTGTGGAAATAGGACATATATGCTAGAGCGGGACTCTGCTT
GTCACTTAGCATCGTACCCTATATGAGGTGGAAGAAAGACAGCGATTTTGGCCCCAGCTCCCGTCTTCTAGTTTAAAGAGATTGGA
AGTTGGAAATCTGTCAATTAGACAACGGTGGGATTAATCCCGTTTGGAGCAACTGTCAACCTTTCGATTACAAAGTTGTTTCGATGTC
GAATTCCTTTGTGTGAGCCCTGAGGTGGTTCGAGCGCTTGACTTTGTTGGAACAAAAAGGCAATTTGGGTGTGGTTCCGTTTAAATGT
GCCCAAGAGCACTTCTTCTATTAGCACATGAAGATTCTGGAAGAGGTAATCTGTTTGGATGCTCCTTGTTAGTCGAGATCTGGGCCC
TGA 3'

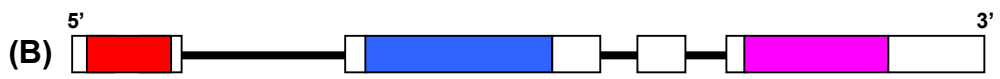
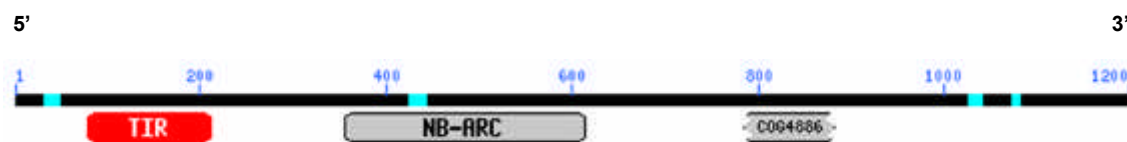


Figura 10 – (A) Sequência de nucleotídeos representando a região codificadora do CT13, que possui similaridade com genes *R* da classe TIR-NBS-LRR. Em minúsculo estão representados os três íntrons; em maiúsculo estão representados os quatro exóns. **(B)** Diagrama esquemático do CT13, onde os retângulos representam os quatro éxons e as linhas os três íntrons. (vermelho = sequência similar ao domínio TIR; azul = sequência similar ao domínio NBS; rosa+ sequência similar ao domínio LRR).

(A)



Domínio	Acesso	e-value
smart00255, TIR, Toll - interleukin 1 - resistance	gnl CDD 22729	1e-21
pfam00931, NB-ARC, NB-ARC domain	gnl CDD 24492	2e-27
COG4886, Leucina-rich repeat (LRR) proeín	gnl CDD 14020	0.001

(B)



Domínio	Acesso	e-value
smart00255, TIR, Toll - interleukin 1 - resistance	gnl CDD 22729	1e-20
pfam00931, NB-ARC, NB-ARC domain	gnl CDD 24492	1e-33
COG4886, COG4886, Leucine-rich repeat (LRR) protein	gnl CDD 14020	8e-10

Figura 11 – Representação dos domínios TIR-NBS-LRR dos contígus (A) CT12 e (B) CT13. As respectivas tabelas contêm os acessos e e-value, enfatizando similaridade dos domínios. Os domínios foram identificados pelo programa NCBI *Conserved Domain Search* ([http://www.ncbi.nlm.nih.gov/ Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)).

4. DISCUSSÃO

Neste trabalho foram caracterizados oito cDNAs de diferentes bibliotecas do projeto Genolyptus, com similaridade a genes *R* e ao gene *Rar1*, envolvido em resposta de defesa. Baseados nestes cDNAs uma triagem foi realizada em uma biblioteca genômica, que possibilitou a construção de uma biblioteca *shotgun* de um BAC visando determinar a estrutura completa de pelo menos um gene similar a gene *R*.

Os oito clones cDNA caracterizados apresentaram similaridade com proteína RAR1 ou proteínas *R*, mesmo aqueles sem um agrupamento contíguo único (EUGR-PU-000-005-H06-MG e EUGR-PU-003-001-C10-MG) (Tabela 2).

Os dois clones citados, mesmo com estratégia de *primer walking* e subclonagem, formaram dois agrupamentos contíguos cada (EUGR-PU-000-005-H06-MG (sequências de 1.002 e 1.135 nt) e EUGR-PU-003-001-C10-MG (sequências de 689 e 714 nt)). Isso ocorreu em função dos insertos desses clones serem grandes (estimados em 2,4 e 1,9 kb) e dos oligonucleotídeos e enzimas de restrição utilizadas não terem gerado fragmentos menores que abrangessem a região mais interna dos clones. Subclonagens complementares, utilizando outras enzimas de restrição e/ou a utilização de novos oligonucleotídeos no sequenciamento podem levar a obtenção de sequências que abranjam regiões mais internas desses dois clones, possibilitando a obtenção de agrupamento consenso único. Apesar de se ter obtido dois agrupamentos, foram feitas comparações dessas sequências com o bando mundial de dados *GenBank* (NCBI), assim como para os outros agrupamentos obtidos neste estudo, onde se confirmou a similaridade com genes *R* e com a proteína *Rar1* (Tabela 2).

Os cDNAs similares a genes *R* não possuem a sequência completa de genes *R*. Todos os cDNAs possuem sequência codificadora da região N-terminal, onde se encontram os domínios conservados TIR-NBS, mas não possuem a sequência correspondente à região C-terminal (Figura 4). Os

insertos, com tamanhos entre 711 e 1.888 nucleotídeos (Tabela 2), não possuem sequência suficientes para codificar uma proteína R completa, o qual, geralmente, possui genes com cerca de 3.000 nucleotídeos (cerca de 1.000 aa).

Curiosamente, o clone EUPE-XY-000-048-H05-MG (1.659 nucleotídeos) contém uma região que codifica um domínio TIR na porção N-terminal da proteína associada a uma ORF similar a subunidade beta do proteassoma 20S (Figura 5). Não foram encontrados relatos de um proteassoma associado a um domínio TIR. O proteassoma 20S é a subunidade catalítica do proteassoma 26S, que possui estrutura de cilindro oco formado por quatro anéis, cada um formado por 7 subunidades alfa ou beta num arranjo alfa-beta-beta-alfa. Embora as sequências dos polipeptídeos alfa e beta estejam relacionadas, somente a subunidade beta contém os sítios catalíticos. Essa estrutura 20S deste complexo está associada a uma porção 19S, que tem função regulatória, necessitando de ATPs para se associarem, formando o proteassoma 26s. Esse complexo proteolítico tem como função a degradação de proteínas intracelulares de vida curta e proteínas anormais, principalmente aquelas marcadas com ubiquitina. (Fu et al., 1998; Pickart & Cohen, 2004).

Mudanças em regiões codificantes do genoma podem levar à evolução de novas proteínas. Íntrons separam essas regiões do DNA (éxons) que codificam diferentes domínios de polipeptídeos. Os éxons comportam-se como módulos na construção de genes, os quais experimentam diversas combinações ao longo da evolução. Em evolução, a separação dos éxons permite sua recombinação ou embaralhamento para criar novos genes funcionais, criando novas proteínas compondo novos domínios estáveis. Porém, podem ser criados genes não funcionais, definidos pela incapacidade de codificar proteínas (Lehninger et al., 1997; Lewin, 2000).

No caso da sequência encontrada neste estudo, que inclui um domínio TIR associado à ORF similar a subunidade beta do proteassoma 20S (Figura 5), pode ser que se trate de uma nova estrutura de gene. Entretanto, maiores investigações deverão ser realizadas para se confirmar a possibilidade de ser um gene funcional ou não-funcional, compondo a família dos proteassomas.

O clone EUGR-PU-003-042-F06-MG (1.171 nucleotídeos), similar ao gene que codifica a proteína Rar1, relacionada com resposta de defesa, contém uma ORF completa dessa proteína (Figura 6). As proteínas Rar1

(*Required for Mla12 Resistance*) são citoplasmáticas e possuem dois domínios que ligam-se ao zinco denominados CHORDs (domínios ricos em cisteína e histidina), os quais são organizados em repetições diretas (*tandem*) e possuem sequência conservada em todos os eucariotos analisados (Martin et al., 2003). Proteínas Rar1 são componentes essenciais na ativação de genes de defesa em resposta a diferentes patógenos. Essa proteína é um ponto de convergência de várias vias de sinalização de vários genes de resistência (Azevedo et al., 2002; Takahashi et al., 2003) A função bioquímica de Rar1 ainda não está clara, mas mutação do homólogo do gene *Rar1* em *Arabidopsis* interfere no acúmulo de espécies reativas de oxigênio e na ocorrência da resposta de hipersensibilidade (Muskett et al., 2002).

Em diferentes espécies vegetais como soja, cevada, alface e eucalipto, diferentes estratégias foram utilizadas para se identificar análogos de genes *R* (RGAs), como a construção de oligonucleotídeos degenerados, que são baseados em regiões conservadas de proteínas R, como os domínios NBS e LRR (Kanazin et al., 1996; Shen et al., 1998; Yu et al., 2000; Brueggeman et al., 2001; Laia, 2001). Entretanto, domínios como NBS e LRR podem ocorrer em diferentes tipos de proteínas com diferentes funções, favorecendo, então, o encontro de sequências que não sejam RGAs.

Na tentativa de ser mais preciso na identificação de RGAs, utilizou-se oligonucleotídeos baseados em ESTs similares a genes *R* ou a genes relacionados com resposta de defesa em plantas. Na triagem em biblioteca genômica de *E. grandis* (Genolyptus), encontrou-se onze BACs positivos para RGA, dentre eles o BAC 143-F12, com cerca de 90 kb, que foi selecionado para construção da biblioteca *shotgun*.

Na caracterização da biblioteca *shotgun* do BAC 143-F12, dois agrupamentos contíguos apresentaram similaridade com genes *R* (CT12 de 16,7 kb e CT13 de 20,8 kb). Análises desses contíguos revelaram que CT13 contém a ORF completa de um gene da classe TIR-NBS-LRR e o CT12 contém um gene truncado com domínios TIR-NBS e parte de um domínio LRR (Figura 11). Pelo fato de se ter uma estimativa, em *E. grandis*, da ocorrência de um gene a cada 20 kb, baseado no tamanho do seu genoma, estimado em 640 Mb (Grattapaglia, 1994) e com cerca de 30.000 genes, o encontro desses dois agrupamentos está dentro do esperado. ,

A existência de dois RGAs no BAC 143-F12 sugere que este clone seja derivado de uma região genômica de *E. grandis* que contém um *cluster* de genes da classe TIR-NBS-LRR. Estudos de mapeamento de RGAs no genoma de plantas comprovam sua ligação com locos que contêm genes *R* funcionalmente conhecidos (Yu et al., 1996; Collins et al., 1998; Seah et al., 1998; Shen et al., 1998; Chen et al., 1998; Mago et al., 1999; Gowda et al., 2000; Collins et al., 2001). A identificação e mapeamento de RGAs em cevada, demonstrou a sua importância para clonagem de genes *R* (Brueggeman et al., 2001). O mapeamento genético, então, poderá dizer se há ligação entre este BAC e genes de resistência em *E. grandis*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALGHISI, P. & FAVARON, F. (1995). Pectin degrading enzymes and plant parasite interactions. *Europe Journal Plant Pathology* 101: 365–375.
- ANDERSON, P.A., LAWRENCE, G.J., MORRISH, B.C., AYLIFFE, M.A., FINNEGAN, E.J. & ELLIS, J.G. (1997) Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell* 9, 641–651.
- AZEVEDO, C., SADANANDOM, A, KITAGAWA, K. FREIALDENHOVEN, A. SHIRASU, K. & SCHULZE-LEFERT, P. (2002) The *Rar1* interactor SGT1, an essential component of R gene-triggered disease resistance. *Science*, 15: 295(5562):2073-2076.
- BELKHADIR, Y., SUBRAMANIAM, R. AND DANGL, J.L. (2004) Plant disease resistance protein signaling: NBS–LRR proteins and their partners. *Current Opinion Plant Biology*, 7:391–399.
- BRUEGGEMAN, R., DRUKA, A., KUDRNA, D., & KLEINHOF, A. (2001) Efficient cloning of resistance gene analogs from barley. *Barley Genetics Newsletter*, 31.
- BRUEGGEMAN, R., ROSTOKS, N., KUDRNA, D., KILIAN, A., HAN, F., CHEN, J. DRUKA, A., STEFFENSON, B. & KLEINHOF, A. (2002) The barley stem rust resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proceedings of the National Academy of Sciences*, 99:9328-9333.
- BUNNEY, T.D., WATKINS, P.A.C., BEVEN, A.F., SHAW, P.J., HERNANDEZ, L.E., LOMONOSSOFF, G.P., SHANKS, M., PEART, J. & DRUBAK, B.K.. (2000) Association of phosphatidylinositol 3-kinase with nuclear transcription sites in higher plants. *The Plant Cell* 12: 1679-1687.
- CANNON, S.B., ZHU, H., BAUMGARTEN, A.M., SPANGLER, R., MAY, G., COOK, D.R., YOUNG, N.D. (2002) Diversity, Distribution, and Ancient Taxonomic Relationships Within the TIR and Non-TIR NBS-LRR Resistance Gene Subfamilies. *Journal of Molecular Evolution*, 54:548–562.

- CHANG, J.H., TAI, Y.S., BERNAL, A.J., LAVELLE, D.T., STASKAWICZ, B.J. & MICHELMORE, R.W. (2002) Functional Analyses of the *Pto* Resistance Gene Family in Tomato and the Identification of a Minor Resistance Determinant in a Susceptible Haplotype. *Molecular Plant - Microbe Interactions*, 15:281-291.
- CHEN, X.M., LINE, R.F. & LEUNG, H., (1998) Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. *Theoretical and Applied Genetics*, 97: 345–355.
- COLLINS, N., PARK, R., SPIELMEYER, W., ELLIS, J. & PRYOR, A.J. (2001) Resistance gene analogs in barley and their relationship to rust resistance genes. *Genome*, 44:375-381.
- COLLINS, N.C., WEBB, C.A., SEAH, S., ELLIS, J.G., HULBERT, S.H. & PRYOR, A. (1998) The isolation and mapping of disease resistance gene analogs in maize. *Molecular Plant - Microbe Interactions*, 11:242–52.
- COOLEY, M.B., PATHIRANA, S., WU, H., KACHROO, P. & KLESSIG (2000) Members of Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycetes pathogens. *The Plant Cell*, 12: 663-676.
- DANGL, J.L. & JONES, J.D.G. (2001) Plant pathogens and integrated defense responses to infection. *Nature*, 411:826-833.
- DESLANDES, L., OLIVIER, J., THEULIERES, F., HIRSCH, J., FENG, D.X., BITTNER-EDDY, P., BEYNON, J., MARCO, Y. (2002) Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive *RRS1-R* gene, a member of a novel family of resistance genes. *Proceedings of the National Academy of Sciences*, 99:2404-2409.
- DE VEAU, E.J.I., GROSS, K.C., HUBER, D.J. & WATADA, A.E. (1993) Degradation and solubilization of pectin by β -galactosidases purified from avocado mesocarp. *Physiologia Plantarum*, 87: 279-28.
- DINESH-KUMAR, S.P., THAM, W.H. & BAKER, B.J. (2000) Structure-function analysis of the tobacco mosaic virus resistance gene *N*. *Proceedings of the National Academy of Sciences*, 97:14789– 94.
- DODDS, P.N., LAWRENCE, J. & ELLIS, J.G. (2001) Contrasting modes of evolution acting on the complex *N* locus for rust resistance in flax. *The plant Journal*. 27:439-453.

- ELLIS, J., DODDS, P. & PRYOR, T. (2000) Structure function and evolution of plant disease resistance genes. *Current Opinion Plant Biology*, 3:278-284.
- ELLIS, J.G., LAWRENCE, G.J., LUCK, J.E. & DODDS, P.N. (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11(3):495-506.
- FAVARON, F. (2001) Gel detection of *Allium porrum* polygalacturonase inhibiting protein reveals a high number of isoforms. *Physiology Molecular Plant Pathology* 58: 239–245.
- FLOR, H.H. (1965) Tests for allelism of rust resistance genes in flax. *Crop Science*, 5: 415-418.
- FERREIRA, M.E. e GRATTAPAGLIA, D. (1995) Introdução ao uso de marcadores RAPD e RFLP em análise genética. EMBRAPA-CENARGEM. pp. 220.
- FOREST. FORESTs: Eucalyptus Genome Sequencing Project Consortium. <https://forests.esalq.usp.br/>
- FU, H., DOELLING, J.H., ARENDT, C.S., HOCHSTRASSER, M. & VIERSTRA, R.D. (1998) Molecular organization of the 20S proteasome gene family from *Arabidopsis thaliana*. *Genetics* 149 (2); 677-692.
- GENOLYPTUS. Projeto Genolyptus. <http://genolyptus.ucb.br>.
- GRATTAPAGLIA, D. & BRADSHAW, H. J. (1994) Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research*, 24(5):1074-1078.
- GOWDA, B.S., MILLER, J.L., RUBIN, S.S., SHARMA, D.R., & TIMKO, M.P. (2000) Isolation, sequencing, and mapping of resistance gene analogs from cowpea (*Vigna unguiculata* L.). IN: Resistance gene analogs from cowpea p. 1-167.
- HAHN, M.G., BUCHELL, P., CERVONE, F., DOARES, S.H., O'NEILL, R.A., DARVILL, A. & ALBERSHEIM, P. (1989) Roles of cell wall constituents in plant-pathogen interactions. In: Nester, E. & Kosuge, T. (Eds.), *Plant-Microbe Interactions*. McGraw-Hill, New York, pp. 131–181.
- HAMMOND-KOSACK, K.E. & PARKER, J.E. (2003) Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology*, 14:177–193.
- HAMMOND-KOSACK, K. & JONES, J.D.G. (2000) Response to plant pathogens. IN: *Biochemistry and Molecular Biology of Plants*, ed. Buchanan, B.; Gruissem, D. and Jones, R. pp. 1102–56. Rockville, MD: American Society of Plant Physiologists.

- HAMMOND-KOSACK, K.E. & JONES, J.D.G. (1997) Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48:575-607.
- HERNÁNDEZ, L.E., ESCOBAR, DRÛBAK, B.K., BISSELING, T. & BREWIN, N.J. (2004) Novel expression patterns of phosphatidylinositol 3-hydroxy kinase in nodulated *Medicago* spp. *Plants. Journal of Experimental Botany* 398 (55): 957-959.
- HULBERT, S.H., WEBB, C.A., SMITH, S.M. & SUN Q. (2001) Resistance gene complexes: Evolution and utilization. *Annual Review of Phytopathology*, 39:285-312.
- HWANG, C.F., BHAKTA, A.V., RUESDELL, G.M., PUDLO, W.M. & WILLIAMSON, V.M. (2000) Evidence for a role of the N terminus and leucine-rich repeat region of the *Mi* gene product in regulation of localized cell death. *Plant Cell*, 12: 1319-1329.
- ISLAM, M.R. & MAYO, G.M.E. (1990) A compendium on host genes in flax conferring resistance to flax rust. *Plant Breeding*, 104: 89-100.
- ISLAM, M.R., & SHEPHERD, K.W. (1991) Present status of genetics of rust resistance in flax. *Euphytica* 55:255–267.
- JIA, Y., MCADAMS, S.A., BRYAN, G.T., HERSHEY, H.P. & VALENT, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO Journal*. 19(15):4004-4014.
- JOHAL, G.S. & BRIGGS, S.P. (1992) Reductase activity encoded by the *Hm1* disease resistance gene in maize. *Science*, 258:985–87.
- JONES, D.A., THOMAS, C.M., HAMMOND-KOSACK, K.E., BALINT-KURTI, P.J., JONES, J.D.G. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*, 266:789–93.
- JUNG, J.Y., KIM, Y.W., KWAK, J.M., HWANG, J.U., YOUNG, J., SCHROEDER, J.I., HWANG, I. & LEE, Y. (2002). Phosphatidylinositol 3- and 4-phosphate are required for normal stomatal movements. *The Plant Cell* 14: 2399-2412.
- JUNGHANS, D. T., ALFENAS, A.C., BROMMONSCHENKEL, S.H., ODA, S., MELLO, E.J & GRATAPAGLIA, D. (2003). Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. *Theoretical and Applied Genetics*, 108: 175-180.
- KANAZIN, V., MAREK, L.F., & SHOEMAKER, R.C. (1996) Resistance gene analogs are conserved and clustered in soybean. *Proceedings of the*

- National Academy of Science of the United States of America. 93:11746-11750.
- KAWCHUK, L.M., HACHEY, J., LYNCH, D.R., KULCSAR, F., VAN ROOIJEN, G., et al. (2001) Tomato *Ve* disease resistance genes encode cell surface-like receptors. *Proceedings of the National Academy of Sciences*, 98:6511-6515.
- KOBE, B. & KAJAVA, A.V. (2001) The leucine rich repeat as a protein recognition motif. *Current Opinion in Structural Biology*, 11:725-732.
- KUMAR, A. & BENNETZEN, J.L. (1999) Plant retrotransposons. *Annual Review in Genetic* 33:479-532.
- LAIA, M.L. (2001) Clonagem e caracterização de análogos de genes de resistência e desenvolvimento de um marcador SCAR ligado ao gene *Ppr1* em *Eucalyptus grandis*. Viçosa, MG: UFV. 54p. (Tese MS).
- LAWRENCE, G.J., FINNEGAN, E.J., AYLIFFE, M.A., ELLIS, J.G. (1995) The *L6* gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *RPS2* and the tobacco viral resistance gene N. *Plant Cell* 7:1195-1206.
- LESCOT, M., ROMBAUSTS, S., ZHANG, J., AUBOURG, S., MATHÉ, C., JANSSON, S., ROUZÉ, P. & BOERJAN, W. (2004) Annotation of a 95-kb *Populus deltoids* genomic sequence reveals a disease resistance gene cluster and novel class I and II transposable elements. *Theoretical and Applied Genetics* 109:10-22.
- LEISTER, D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics*, 20:116-122.
- LEHNINGER, A.L., NELSON, D.L. & COX, M.M (1997) *Principles of Biochemistry*, 2nd ed., W. H. Freeman.
- LEWIN, B. (2000) *Genes VII*. Oxford University Press and Cell Press.
- LIU, Y., SCHIFF, M., MARATHE, R. & DINESH-KUMAR, S.P. (2002) Tobacco *Rar1*, *EDS1* and *NPR1/NIM1* like genes are required for N-mediated resistance to tobacco mosaic virus. *Plant Journal*, 30 (4), 415-429.
- LUCK, J.E., LAWRENCE, G.J., DODDS, P.N., SHEPHERD, K.W. & ELLIS, J.G. (2000) Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination. *Plant Cell*, 12:1367-77
- MAGO, R., NAIR, S. & MOHAN, M. (1999) Resistance gene analogues from rice: cloning, sequencing and mapping. *Theoretical and Applied Genetics*, 99:50-57.
- MARTIN G.B., BOGDANOVA, A.J. & SESSA, G. (2003) Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology*, 54:23-61.

- MARTIN, G.B., FRARY, A., WU, T., BROMMONSCHENKEL, S., CHUNWONGSE, J., EARLE, E.D. & TANKSLEY, S.D. (1994) A member of the tomato pto gene family confers sensitivity to fenthion resulting in rapid cell death. *The Plant Cell*, 6: 1543-1552.
- MARTIN, G.B., BROMMONSCHENKEL, S.H., CHUNWONGSE, S., FRARY, J., GANAL, A., et al. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science*, 262:1432–36.
- MEELEY, R.B., GURMUKH, G.S., BRIGGS, S.P. & WALTON, J.D. (1992) A biochemical phenotype for a disease resistance gene of maize. *The Plant Cell*, 4: 71-77.
- MEYERS, B.C., DICKERMAN, A.W., MICHELMORE, R.W., SIVARAMAKRISHNAN, S., SOBRAL, B.W. & YOUNG, N.D. (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant Journal* 20:317-332.
- MEYERS, B. C.; KAUSHIK, S.; NANDETY, R.S. (2005) Evolving disease resistance genes. *Current Opinion in Plant Biology*, 8:128–134.
- MICHELMORE, R.W. & MEYERS, B.C. (1998) Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. *Genome Research*, 8:1113–1130.
- MILLIGAN, S.B., BODEAU, J., YAGHOUBI, J., KALOSHIAN, I., ZABEL, P. & WILLIAMSON, V.M.(1998) The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell*, 10:1307–1319.
- MUSKETT, P.R., KAHN, K., AUSTIN, M.J., MOISAN, L.J., SANDANANDOM, A., SHIRASU, K., JONES, J.D.G. & PARKER, J.E. (2002) *Arabidopsis Rar1* Exerts Rate-Limiting Control of R Gene–Mediated Defenses against Multiple Pathogens. *The Plant Cell* 14:979-992.
- NIMCHUK, Z., EULGEM, T., HOLT III, B.F. & DANGL, J.L. (2003) Recognition and response in the plant immune system. *Annual Review in Genetic*, 37: 579-609.
- NOMA, K., NAKAJIMA, R., OHTSUBO, H. & OHTISUBO, E. (1997) *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. *Genes and Genetic Systems* 72: 131-140.
- NSF (National Science Fundation): Genomics of Wood Formation in Loblolly Pine Program. <http://pine.ccgb.umn.edu/>.
- ODORIZZI, G., BABST, M. & EMR, S.D. (2000) Phosphoinositide signaling and the regulation of membrane trafficking in yeast. *Trends in Biochemical Sciences* 25: 229±235.

- PAN, Q., WENDEL, J., FLUHR, R. (2000) Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *Journal of Molecular Evolution*, 50:203–213.
- PARNISKE, M. & JONES, J.D.G. (1999) Recombination between diverged clusters of the tomato *Cf-9* plant disease resistance gene family. *Proceedings of the National Academy of Sciences*, 96:5850–5855.
- PICKART, C.M. & COHEN, R.E. (2004) Proteasomes and their kin: proteases in the machine age. *Nature Reviews Molecular Cell Biology*. 6: 177-187.
- ROMEIS, T. (2001) Protein kinases in the plant defence response. *Current Opinion Plant Biology*, 4:407-414.
- ROSSI, M., GOGGIN, F.L., MILLIGAN, S.B., KALOSHIAN, I., ULLMAN, D.E. & WILLIAMSON, V.M. (1998) The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proceedings of the National Academy of Sciences*, 95: 9750–9754.
- SAMBROOK, J., FRITSCH, E. F., & MANIATIS, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SALMERON, J.M., OLDROYD, G.E.D., ROMMENS, C.M.T., SCOFIELD, S.R., KIM, H.S., LAVELLE, D.T., DAHLBECK, D., STASKAWICZ, B.J. (1996) Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell*, 86: 123–133.
- SEAH, S., SIVASITHAMPARAM, K., KARAKOUSIS, A. & LAGUDAH, E.S., (1998) Cloning and characterization of a family of disease resistance gene analogs from wheat and barley. *Theoretical and Applied Genetics*, 97:937–945.
- SHANMUGAM, V. (2005) Role of extracytoplasmic leucine rich repeat proteins in plant defense mechanisms. *Microbiological Research* 160: 83-94.
- SHEN, K.A., MEYERS, B.C., ISLAM-FARIDI, M.N., CHIN, D.B., STELLY, D.M., & MICHELMORE, R.W. (1998) Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. *Molecular Plant-Microbe Interaction*, 11:815–823.
- SHIRASU, K., LAHAYE, T., TAN, M.W., ZHOU, F., AZEVEDO, C. & SCHULZE-LEFERT, P. (1999) A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. *Cell* 99 (4), 355-366.
- SONG, W.Y., PI L-Y, WANG, G-L, GARDNER, J., HOLSTEN, T., RONALD, P.C. (1997) Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell* 9:1279–1287.

- SONG, W.Y., WANG, G.L., CHEN, L.L., KIM, H.S., PI, L.Y., HOLSTEN, T., GARDNER, J., WANG, B., ZHAI, W.X., ZHU, L.H., FAUQUET, C., RONALD P. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*, 270:1804–1806.
- TAI, T.H., DAHLBECK, D., CLARK, E.T., GAJIWALA, P., PASION, R., WHALEN MC, STALL RE, STASKAWICZ BJ (1999) Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proceedings of the National Academy of Sciences*, 96: 14153–14158.
- TAKAHASHI, A., CASAIS, C., ICHIMURA, K. & SHIRASU, K. (2003) HSP90 interacts with *Rar1* and SGT1 and is essential for RPS2-mediated disease resistance in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 100:11777-11782.
- TAKKEN, F.L.W. & JOOSTEN, M.H.Q.J. (2000) Plant resistance genes: their structure, function and evolution. *European Journal of Plant Pathology*, 106: 699–713.
- TAYLOR, G. (2002) *Populus: Arabidopsis* for Forestry. Do we need a model tree? *Annals of Botany*, 90:681-689.
- THOMPSON, J.D., GIBSON, T.J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25:4876-4882 www-igbmc.u-strasbg.fr/BioInfo/.
- VAN DER BIEZEN E.A., F. C. T., KAHN K., PARKER J.E. & JONES J.D. (2002). *Arabidopsis* RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant Journal*, 29: 439-451.
- VAN DER BIEZEN, E.A. & JONES, J.D.G. (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology*, 8(7):226-227.
- VOS, P., SIMONS, G., JESSE, T., WIJBRANDI, J., HEINEN, L., HOGERS, R., FRIJTERS, A., GROENENDIJK, J., DIERGAARDE, P., REIJANS, M., FIERENS-ONSTENK, J., DE BOTH, M., PELEMAN, J., LIHARSKA, T., HONTELEZ, J., ZABEAU, M. (1998) The tomato *Mi-1* gene confers resistance to both root-knot nematodes and potato aphids. *Nature Biotech*, 16:1365–1369.
- ZHU, H., CANNON, S.B., YOUNG, N.D., & COOK, D.R. (2002) Phylogeny and Genomic Organization of the TIR and Non-TIR NBS-LRR Resistance Gene Family in *Medicago truncatula*. *Molecular Plant - Microbe Interactions*, 6(15): 529–539.
- YOUNG, N.D. (2000) The genetic architecture of resistance. *Current Opinion in Plant Biology*, 3:285–290.

YU, Y., TOMKINS, J.P., WAUGH, R., FRISCH, D.A., KUDRNA, D., KLEINHOF, A., BRUEGGEMAN, R.S., MUEHLBAUER, G.J., WISE, R.P. & WING, R.A. (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. *Theoretical and Applied Genetics*, 101:1093–1099.

YU, Y.G.; BUSS, GR. & SAGHAI-MROOF, M.A. (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Genetics*, 93: 11751-11756.

ANEXOS

Anexo 1 - Oligonucleotídeos utilizados para amplificar os análogos de genes de resistência.

Clone	Nome	Oligonucleotídeo (5'- 3')	T_m (°C)
EUGR-PU-000-001-G07-MG.A03	RGA03F	CACTTGCCAAATTGCTTTCA	59,85
	RGA03R	CTGCCCATTTTCTTGGATGT	59,93
EUGR-PU-000-005-H06-MG.A09	RGA09F	GGCAAATCAAGCATGGAAAT	59,91
	RGA09R	AACAGATTGAGGCCCTTCT	60,07
EUGR-PU-000-010-D04-MG.A11	RGA11F	ACTCCTTCTCATGCCTCGAA	59,95
	RGA11R	CCACACCACAACCTCCTCTT	60,00
EUGR-PU-003-001-C10-MG.B05	RGB05F	GCTTGCCCTGCAACAATGTAA	59,88
	RGB05R	CGACCTCTCGACTGCTCTCT	59,88
EUGR-PU-003-020-C11-MG.C05	RGC05F	GCCATCGGTTTTTGTGAAT	59,77
	RGC05R	TCATGGAGGTTGATGCGATA	60,03
EUGR-PU-003-042-F06-MG.D06	RGD06F	GGGATTTCACTGCAAAGGA	60,05
	RGD06R	CGATCCTTCTTCTTGCTTCG	60,09
EUPE-XY-000-035-C06-MG.E12	RGE12F	GCCATCTGCAAAAAGGATGT	60,08
	RGE12R	GTAGCCGGGGGTCATCTACAA	59,96
EUPE-XY-000-048-H05-MG.F08	RGF08F	CAAGGACATATCCCCGCTAA	59,92
	RGF08R	ACCGTTCCTCTCCGAAGATT	60,07

Anexo 2 – Sequência de agrupamentos contíguos de oito cDNAs. Dois deles apresentaram dois agrupamentos contíguos (EUGR-PU-000-005-H06-MG.A09 e EUGR-PU-003-001-C10-MG.B05).

>EUGR-PU-000-001-G07-MG.A03 (1.888 nt)

3' TTTTTTTTTTTTTTTAGAATAAATGAAATGCTTATGCAAGGAGTGGGACTCATATCATACTCTAAGTTGTTTTCCAACCAATCCATT
CCTCTGAAATAAGGCTCCATGAAAGGTCAAGAACGACTAAGTTAGCCGGATGAAAGTTTGTGCTTCAAATGGTTGAGTAAAATTT
CCAAGAAAGCCATCTTAAACTAGGGAGAAGGTTGTTGAAGTCTCCGATGAGCGCTTCCCCTCCCTTGACAGAACCTTAGGTTTCGA
AACGGAGCAACTTCATCACTTGCCAAATGCTTTCAAGAAATCTAGTGACATGCTTTTATTTTCTATTTCCCTCTTTCTCCTTCA
GAATGCTCCAAGCTTCATGATAGTTCCACACTCTACTACGCTCATAGGGATCGGAGAAGTTCTCTTGACGGACAATTTCCCTTCCGAG
GTCCCGCACTTGGTCATGCATCCAGAAAGTTTGTGCTCTTTGATTTTTTCAAGGACATGAGAAGGAGGACATCAATGGCAACATGT
GGTGGTATCCATAGTCATCCACATATAGAAAGGGCATGTTGATCTTCAATTAACAAAAAACAGGCTATATCCAAAAATACTTCCC
TTTGTGTACGATCTAACCTGTTATAAGTTATCATCAATTTCTCTGGACTTCTTAGGAGGAGCTTCTGTAGCTTCTTCCATGTGTC
AACCCATAATTTCTCGGTTTACCATTAAGGGATGAACCTGTAATCTCGAGAGCTAAAGGAAGCTTCCCAAAGTATGGACAATTTTC
TGAGAAAGGAACATACGATCCAGCTATGGTGGACAGTCTCTTCTAAAGGCACGCTACGTAAAAAGCTTAAAGAGCATGATCAAATTC
CATTTCTTGATACATCGTCAAAGCTTAAAGCTGTGTGAACCTTTGACATCTCAGTTCTACACATCCAAAGAAATGGCTTCTCCCTTAACTATAGGATACT
TTGGTCCCTAGTAGTTATGATAATCTACTACCAGGACCGAACCAATTACTTCTTCTGCTAGACTTTCAAGTTGTTCTTTCTCATCC
ACGTGCTTAAACAATGAGAATTTTTTATTTTAAAGTACTCTGTTTCACTGTTGATCCCATCATTACTTGATAAAATTTGGTTCGA
TAGAATGAGAACCAAGGAGCTTGGATAACAATTTGGTTTTGCAAAATTTACCATACCATATTGCTTTTGTGATTTCCCAACGTTTGC
AAGGAACTGCAATCGTCAAAGCTTAAAGCTTAAAGATAGTATGTTGGAACCAACCGGAGGAGTTGTTTTGCAACACCGCCCTTCCGTGG
ATTACAAGAAATCGTACACTATCAGATCCCACTTTCAACAATTTGATTATATCTTCTATGCGATCATCAACTCCGACTAGATGTTTCCAG
TTACATTTTTCATGTGCAACCTTTCAGCTTAACTGAAATCGTTTCAATTAACCGAACATATAAGAGCTCCTTGGCCTTGATCTTCTTCAA
GTTCCATCCCTTATCTCGTCCACTTCATTGAGAGCTTTTTTCCATGACTCCACTTCTGCACTAAAATTTTCTGGTGTCTTGCAGG
CATTTGGCTATATAAGCTGTTTTGAGCTGTGATCCCAAGCTTCCAGTTCCCAAGAAATGGCGAGAATCTTTTGTCTTCTTCAATTT
TCGAGGTGCACTCGACCATGTGTGTGAGCTCTCGGAGGCACCACTGACTCGAAGCATAATTTTTGGAGAATATGGGAATGTAGATTTT
GGAGTTCTCAATTTGCGTGTAGAAGTTCCACCACAAATTTTTGACCTACGCGAAGGGATTTCATCGTCCCTAAAGACAAGGATCCCGACC
CTAACCATGGCATGGTAAAGGCAATCAGTGGGAAGCGTGGG 5'

>EUGR-PU-000-005-H06-MG.A09a(1.002 nt)

5' TCTCCTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGGCTAACTAGGTCATTCGCCCTGCTGGTTTTCCCATGC
CCCGAAGAAATCGGAAGCCAAACCTTTTTGACCCGCCATCCACCGTAGGCAGGCGAGCAGGTCGGTTCGGCGCGGGGAGATGACAAAG
GCAACGGTGCATAGGCGACGCCACGGCTGGCGACGCCCTTCCATCCGTTACAGGTAACGCACGCATGCGTCTCTCGAGCTTTCTCTG
CGTCTCTCAGCTTCAGCTGCATTTTGAACCCCGTTCGAACCCAGCAGGCTCATCGATTCTGTCAAGATTGTTTTGCAACACCGCCCTTCCGTGG
AATTGAGTAGGCGGTTGGAGCTGGAAGCCGCTTGGCTGCCATTGTACTATCGAGATTTGATAGTTGGGATGGCGGTGTTTCAAGTTTCC
TAGCAATGGCCTTCTACGTCTTCCGGGGACTGTGCTTGGGAGCGGGGCTGCTGAGATCACCTGACGACAATATCTCTTTTATGGT
TTTCTCGCGATGTTTCTGTTCTGTTGGTGTACTGCGATAGATCCAATGGATAAGACTAGTTATAGAAAATGAAAGAGAGGCTTCGA
AGCTATGGCACTCGCGGACTTTTGTGCTGATTCGCCATCGCCTTCCAATTTCCATGCTTGAATTTGCTACCTGGCTGGGATGTG
ACATGCTATGTATGAGCTATCTGAAAAATGAATGCTTCCGGCATATCTATATGAAAATGGCTCTAGGCAATCCATCAAGATTCAAGA
TCTTAGTCTTAGGCTTGTGCTTTATTAAGTGAATCTCTAGAATCTAGATGATGGAATCCACTAGAATCTAGGATTGTGGAATC
CTCTTCTATTTCTCACCTTGTGTTTCTTGATTCTGCAAGATTGGCCAGATTTTTTACTGATTCTTGTCTTCTACCTTCGTATGAATC
ACATACATCTTTTTCGGGTCTAAAAAAGAAAAA 3'

>EUGR-PU-000-005-H06-MG.A09b (1.135 nt)

3' TCTTAATTGCTAAACCATATATAGGTAATTGAGTTAACGAAACCAAGGAGCCCATGGATTTAGGAAGTTGAGAAATTTTACAATCA
CGAAGAGATAGACGCTCCAACCTCATCAAAACAATCACTCGAAATAGGAATATCTTGTATATCAGTTTCCGCTATGTTAAGGTCCTCA
AGTTCTATCATTCTACCTATACCTTTTGGTAGCTCCTTGAGTTCTCTACAAAAGATAGACATCCAATGAAAACAAGATCTTGTGTC
CAATAGAAAGGATGAAGCTCTTCCAACCTGATGACATTTTTTAAAGAGTCAAAAATCTCCAACCTTACGAAAAGTGGACAAGTCTGGAGTTCT
TTTCAAAGATTGACAAGACGTGAGGTTGAGAATTTGAGCTCAGTCGCAATCTTGTAGTGGAAACCATCCTCCCAAGGCTCCGAAAT
GCACTGTTTGACAAATCAAGTCAAACTAATTTTTTCCGGATGGAAGTTGACTGCAATGAAATCCAAGGGACAGCCTTGGCACTGAAGCC
ATCTTAATTTAGGAAGTAAAGCTTTGAAAATCTCCGGTAAAATTTGCAATAATCACTTGAAGGAATCTCAAGTTGGTCAATCTGTTAAA
ATGTTTCACTGTCATATCTTTTTGATAACCCCATCCGCGGAGAGGAGCGTATCTTTTTGATAACCCCATTTGCGGAGAGGAAGGCGC
TCAATCTTCTAGTACCCTCATGTTTATCAAGCACATCTAGGGCTTCCCTCATGGATGCACAATCTGCTAGCTCTTGTAGGCTCCTTTG
GATTTTCTAGACGAAACAATTTCCCTCCGAGATCTCTCAATGATCATGCATCAATAGCCTTCCGTTTTCATCAATTTGTAATTTAGGG
ACATAAAACATAAATCACTTCAATCCCTTCTCGGGGAAAAATTTGTCCTCAATATAAGGTTGATTTTGGTTTACATGATCCAAAT
GAAGAGGCATGCGATATCAAAAAATTTTGTGGTCTCGTACTCTAATGCTTCATAACTTATTTTCAACTTCTCTCGCACTTTCTCA
TGGGCACTTTCTTAAATTTTTTCAACGTATCTTCCATACGTTTTCTCTCTTTTCTACATAAAGAAATGAACCTGTAACCTTCT 5'

>EUGR-PU-000-010-D04-MG.A11 (746 nt)

5' CATTTCATCGGGCCACACCACAACCTCCTCTTTTTTCTCTTTAGGGGTGAGAGATATCCGGAAGGAATAAGCTTCTAGCTATG
GCGAACTCAGAGGCCGGAACAAAAAGTATGTTGCACGGGAATTAGAAGTTGAGTATCAAGTGTCTTCTGAGTTTAGAGGACCTGACA
CTCGAAATGAATTCACCGACTTCTCTACTATGGCTTGGAAAGATGTTGGGGTCCATGATTTAGGGACCAAGATGAACCTCCGCTCGG
CGAAGTGTGGTAAAAACATTTACGTGCTATTGACAACTCCATAATCTACATACCCATCTTCTCTCGGACTTATGCTTCCAGTAAA
TGGTGCCTCCGTGAGCTTGTGACATAAGACAAAAATTTGTCCTCAATAGAGGTCAAAAATGCACTTCTCCCATTTTTTTTTTTCATG
TGAACCTGATGATATACAAGCTTAAAGCTCAAAGGTACATTGAAGACTTTTTGAGGCATGAGAAGGAGTCTCCCAAGGGAAGTCAAG
GCATGGAGAGAGGCTCTTGCAGAGTTGACGAAATTAAGGGATGGAACGGTGAAGGATCAAGGTAATTTTTTAAAGATACCTAC
CTATTTTTTCTTTTTTTTCTGTTTTAAGGGATAAAATGTTGCCCTGTCCAATGATTTCTTTTATACTTTTTCTTGACGGAATAATGTG
AAGCTTTTTTAAAGATCCCTACCAAAAATAAAAAAAGAAAAA 3'

>EUGR-PU-003-001-C10-MG.B05a (689 nt)

3' ATCGGCACTATTTGTGGCGATCGCTCCTAGCAGCTCCACCAAGCCGCTGGCCACCATCGCAGCAGCTCTCCGTCAGTGCCTGCCAC
TTCTTTGTCAACAGCTCGACACCAAGCCACCA GAAGCCGCGCCGAGCCATCTGGTCCCCACCTTTCTTTCTCTCACTCAATGGCAA
TCCGATTCGGGCTCGGATGTCTGGGCTAATGTCAACCAAGCCGATCATAGGCATCGCCGAGCTCCCAAAATCTAGGTTA
GCCAAGGCGCAAGCCATCCTCCGGTGTGCGAAGCCGCGACCCACATGACTTGAAGGATTGAGGCAAGGCTGAGATGGGCAACG
GCACTCGAATGTGACAGTGTGAGGCGGAGAGGTAAGTACGTTTATCGGCGGCTTGGTAGTGGCAAGGCGGACGTTGATGATGACTA
AATTTGGGCAACCGGAGATGATTTGGGGAATTTCCGATTTGTTTTAGCGTCTAGATTATTTTTCGCTAATTTCTGCAAAATTTAACGACAA
TTGAGTTACTTAGGGCGTAGTTTTGCAATATAAAATATATTCTTAAAATGGTTGTACATTTGTCAGGCAAGCAGATCGAAAAAT
CCATAATTAAGACTGTGTTGGATAAGCTGAATATTTGGATAAAAAAGATTGTTTTCTAAAGAAAAA 5'

>EUGR-PU-003-001-C10-MG.B05b(714 nt)

5' CTCCAAGTTCCATCCCCTTATCTGTCCACATCAACAGAGGCTTTTTCCAGGATTCAACTTCAGTGCAAAACTTTTTCTGATGCT
TTGAAAGGGCTTTGCTGTATAAGTTTGTTCAGTTCAGTATCATCAGGCTCCACATCGAAGAAAATGGGCAAAATCTCTTTATTCCC
ATTGATTTGGAGAGCGCACTCAACCATTGTGCAAGCTCTCGAAGGCACTAGTACTGATGCATAATCTTGGAGAATATGGGAATG
TAGATTTCCGGAGTTTCGGATAGCTTGTAGAAGTTCATCAACAAATCTTTGACCGACATGGAGGGATTCTTTATCCCTAGAGACGAGGA
TCCCGGTCTCAACCATGCCATGGTAAAGAAAATCGGTAATCTATGACGGGTGTCTCGTCCATGAAAACCTTAGAAAAACCTCGTAAAG
AGGTCCAATTCGCTCGAACAATCTCAAGGAAACAACAATCTTCCAAAATGAAAAGAGAGTGGTGAAGAACGACCTTTGCTAGAG
AGGAGAGCGATGGAGAGCGTGGAGAGGCTGAGGATTCCGATGAGGAAGTAGAGCTTGGGCTGGGGCAGGGAGTCTCGCAGGTGAGCT
CGTGAGGAGAGAGCAGTCCGAGAGGTCGGCGTTACCCGAAAACGCCATGTTTTCCAGTAGAGCGAAGAGCGAGAGTGTAAAAGGC
TAAGGGAGTGCG 3'

>EUGR-PU-003-020-C11-MG.C05 (711 nt)

3' TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAAAACATCTTATTCTGGTGTAAATCAATATCTAGATCTGAAAGCAAAGCACTGT
TTCCAACTAATTTCTCCACTTGTCTTTGTTATCAACGTCATCAAGTACAATGAGAAGTCTCTTATTGCAAGTACTTGTGAATGCT
TTTCATTCCGGAATCCATTTGTTAATGCTCCTTGGATGGCCAAATTCAGATAGTAATCTTTTTGCAAGTCAACTAAGCCATCGGTT
TTTTGATGAATTCGCTCGAACAATCTTCAAGGAAACAACAATCTTCCAAAATGAAAAGAGAGTGGTGAAGAACGACCTTTGCTAGAG
TCGTTTTACCAATACCCCCATGCCATAAATTTAATGAGTCTCGTACCAACGGAGTCCGATGTCTAATAAGTTGCTTACTGCCAGTAC
TTGCTCATCAATTCCAACTAAATGCTCAGTACCAATCTATGTTTTGCTTCAATTTTTCAAGACCTTCAACTATCAATTTGATT
AGCTCGCCGTGGCTGTACTTCTCACTTCCACCCCTTTATCGCATCAACCTCCATGAGAGCTCTCTCCAAAACATTGACTTGGCT
CATTTGCTCAACTTCTTCACTCTCCAAATTTGATATGGCATCGTGATATAGTGGAGTTTTCAACTTGACATCATCAGTTCCACATC
AAAGAAAAT 5'

>EUGR-PU-003-042-F06-MG.D06 (1.171 nt)

3' TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGGACTTCCCGAACCGAA
ATGATCGGGAAAACCTGCTTAAGAGAAGAAAATACCCGAATCAAGAATAGTGTTCACAACACAGCAATAATTAAGGGAAAG
GTACAGTTAAGCTAATCACAGTTGAGAAGCAGGATATCTTGTGACATTTGTCGATACCATGTCAATCAAGGAAAACAGTGGGATTTTC
ACTGCAAAAGGATAACAACCTACCCTCCAAATCAACTCCGATACAATCCATACAAAACCTTATGCTACACCTTAGTGTCCCTATG
ATGCCGGATCTGCATTATGCCATCCCTTGGTACATGGTGAATGGTCACTCAAAATCCCTTAACATGAACGTACAGCATTT
CCACCTCTCTCGTGTGGAAAACAGCAGGCCAGGCTCCAGTGGAAAAGAGCAGGCTATCTCATGATGTTCTCTTTAAAAGTTTGA
CCGCATCCTTTATTTTTGCAAGTCTGCGGCTCATTTATGTCGATTATCTCTTTTGGGAACCCAGAGATCTGGGCATCCTTACTGC
TTTCAACAGCAGTTTCACTTTGCATTTGCCACATCTATGTTTGTTCAGTTCTTGAACCATGATCAGAACAGAAAGAACCCCTG
CCGACATCTCGCGCAAGACTCTTTGATGATGCATTAGTTGCAGGTGCAGCACTGGTGGCAGCACTGGTGGCAGCACTGGGAGTCTTA
TAGCTGTACTCCCTTGGTGTGATGGCTTTTCAGTAGTGTGTTGCGAGTCTTACATCCAGGAATCCCAAGAACTCACTAAAATCAT
GACTTTTCTTCTGCAAACTCCATTCTTCCATCGTGAATAATAGGACCCCGGGGTGGTAGCGGCAGGAACTTCCGGATT
GTGCTCCTCGGAAAAGGAGGCTCGCAGCCGATCCTCTGGCACCGACTTTCGCCATCTCGCCGCGCCGAGCTGGGTGGAAGTCAA
TGCGGGCGGGTCAATGCGAGGGAGTGAGTCAAGCTCTGCGCAATCTCGAAGCAAGAAGAGATCGAGAAATCAGATTCCACGAGGGA
GCTTTAAAGCTCGTTCAAGTAGGGTAGAG 5'

>EUPE-XY-000-035-C06-MG.E12 (1.641 nt)

3' TTTTTTTTTTTTTTTTTTTCAAGTCAAGCAAGTACTTTAACTTTCCAATAGAGTCGTCGATTTCTTCAAATTCCTACAACCTTTAA
AAGTCAACCTCTCTAAATCTAGGCATTTAGACAAGTCTGGAGTTCTAATAATGCCACGACATTTACAAGAGTAAGATTTTAAGTTT
GGTTGCCATCTGCAAAAAGGATGTATAAATGCTACAAAAGAGTCAAATAATGTATGAGACAAGTGTATAAATGAGATGAAACCA
TACCTTAATTAAGTCCCATACTTTCGAATCATTTGTTAAGTCAATGTTAGAAAAGTTCGAAGACAATAAATCTTTGAATGCATACTA
AGCCCAAAACATCTTGACGGGCAAAACGCCATGAAATCCATCTCAACTTTGGTAGAGAAATGGTGAATCCCAACATATGGCCCAT
GACTTAATCTCAGGAACCTCAAACCTTTAAACCTTCTAATTTCTTATGGCTAATTTGAGATATTAGACCTCCACTGCAGATCAAGTGC
TTCAACATCTTCCGACTCTCTTCCGTTCTTACGAAATCAAGAACTCTTTGCTAGTCCACACCTACTCTGTTGAATTTTAA
GGATTTTTTGACGAACAATTTCCCGTCCCAAGTCTCTCAATTTGGTCAATGCATCCAAAATTTGTTATCTTTATCTTTAACAAGGACT
TCCTAACAGGACATCAATTTCCAATCCAGGATAATTCACAAGCTCCACATGATTTGGCATTAGTAATATCTCATCGATGAAAA
AAGCATGCGACATCTAGAAAAATTTGCTGTTGACGATAGTTTAGGTCCTCATAGCTAATCATCAACCTTTCTTGTACATGCTCATCT
GGTATTTTTCTTTAAATTTGTTTTAAATGTATTCTCCATTTGTTTTGCTGTCCACAAGTGAAGAACTATAAATCTATAGAG
CAAAAGGAAGGCACCCAGTAGTAGAGACAATATTCTCTTGAAGATCGTCAAAAATCTGTTGCTAGAAAAGTCTTCAAAAACGCATGC
CTCTGAAAAGTTGGAGTGCATCGTCAAAATTTATCGCCACCATTTTATAAGGAGAAATTCGGACCGGTACCTGCAGGGCCAGTAA
CTCAGCTTTCCCTCATAGTCAGGAGCCATTTATAGCTTAGTAAACTTCTAGTACGCTTTGTGTTCTTAGTGGTGATAAATTTCTT
GTTCCGGATGCAATCGATGTTGTTCAATTTAGATTTCTCGATGTTGTTTTCTCATCAACATCATCAAGTATGAGAACTTTCTTAC
TGCCAAGTGTGACTCTATCTCTCTTCCGCGATCAACATCATTAATAGTTTCTGCTGCTCCAGAATCATCAATATCTTTTAGTAA
CTTCTTCTGACGAGACTAAACCCACTTCTTTGAAGTCTCTCGAACAGCATCAAGGAAGGCGAATGCTTTCCAAAGCGAGAAGA
AAGTTCAATTTGATAGATGACCCCGGTACCGTGGTTTTACCGATGCCCCCATTCATGAATTCAGTAGGCTCACTTCACTACTACAG
TTAATCTAATAATTTAGTACTTTTGGCACTCGATCATCAAGTCCGACTAAATCTTC 5'

>EUPE-XY-000-048-H05-MG.F08 (1.659 nt)

3' TTTTTTTTTTTTTTTTTTTGATGAGAGAACGCTGAAACACATGGGCAAGGCGGAAACACAAGAAATTCGAAGTTTGGGACAT
AAAGTCCAAATATAGGGGACATAAAGCTCCATATGTTTCAGACGTTAAACAGAAAAGGAATCAGAAGAGATGCCACATCCCATCT
AATGCGCCACCAAAACCCGACGCGCAGCCAGACCAGCCAGCCATGAAAGGCCACGGAAGGATAGACTGTAAAATACATAGC
TGAGAAGGAAATACAACAAGGACATATCCCCGTAATACGGAGCTGCGAAGCAAAACACAACGACATATCCAGCTAACGTGCTGGGTC
CGAATCAGACTTCAATCCAGCTTTATTCTAGTTGGTCACTTCAAAATGAGGTCCCTTCCACCAACAGTGGGTTTTTCATACGATTCAA
AGCCCCAGAAGGTTTTCAAGAGTAGGGCGGAAAATTTGTGCTCTTCTCTGTGATTTTACAATCTGAAGCTTATTGACTGCAGA
TCTGTACAGTACAGAAGCAGCGCATGCACTTCTCAACAGCTTAAACCACTTCAAACTCAAGTTCTCGTCCACTCATCGCGA
AGAAATGGCCGTGCAAGGTGATTGCCAAACCCAGTAGCCATCAGAAATCCCTTTTCCAAAACCTACGCTATCATGCTAAACCATGGCAAGAT
ATTTTTGTCCATTTTAACTCCACCTAGAAGTGTGAGTCCACAATGGGTTGAAGTGTTCGGCTATTATACATAACCCGAGTCCAG
ATAAATATGCACTCTTTGGGCCCCAAGGAATCCCATCATCCACATATTTGTCGTAAGAATCAGTTTCAAGGTAACGCAATATC
TGATGAAAGTCACTTATCTCTCACTCGCACCCAGGAGAGAATGTTTTCCAATAGGTTTTATTCGTTCAATAGTTTGTATCTCAAGG
TGATCCATAAAGAACCTCCATGTCAGACCCATCAGAAATCCCATCCCTGTACTTTGAGAGCCATGAGCAGTACCCGTCAATATGG
ATACAGAGTTCTTTAGAGTCGCCAGATCGGGAGTCCGGACTTGTCCACATAATCTTCAAATCCTTGGCATGCACCTGGGGTAC
CTCTGCTCATGCTTCTCCGAGTCTTCAAGTACAGTTCCGCTCTGAGCTTTCAGCTCCGAGGGCTGACCTTAAAGAAAGATGGGTAGAA
TCTCCGCTTCCGCGGATTCGCTGAGCTTCTGTCATCCGGAAGAATTTGGAGAGGCGCCGATCCCAAGATAATCTCGGAGAG
GAACGGTATGAGGATCTTTGTTGCGCCGACCGCTCTCAGGGGATCGTCGATCCGATCACCGCGCTGAAGCTTCTCGCTGTCTGTFTA
GCGCGGTCAACTGGCGCTCCATGGTCCGGTAGAGGTGGTCCGGTGGAGTTGTTGCGAGTGTCCGCCCTTGAACCTAAGAACAGCT
CGTACATCCCGCTCGGCGCTTCCCTTTCCGTTCCGCTTCCGCTTTCAAGTCCGCAAGGATTTCCCCCTAGATCCGAAAGCTGTCC
CCTGCTCCATTTACGGCGATTCATCTGGGCGTAAACTGTTTTGATCTGACAAAACGACGATGGGCGCAAGAAG 5'

Anexo 3 – Agrupamentos contíguos (CT) da biblioteca *shotgun* do BAC 143-F12. As ORFs obtidas com os programas ORFfinder ou GenScan. São representados com letras maiúsculas os éxons e letras minúsculas os íntrons.

>CT03

5' aatTTTTgtgacagcattgCGtaaatattgagctggcgagcttctccgggtggtatggtgctgtggaacttcatgattgtgctTTTTgc
taaacgatactactgtacttacgagatcttggaaagacatattgatgaaggtttgagttcttcaagggtctggcttTgtgttataatct
ttcctcttatgggtggttctgtaccAAAatgttaaacctgtcttggatggtttgggttgagatccttttagttaagttctttaaagcAAA
gtatggggcatttAAattaggagcaattgtgggggttccactccgactgaattgggctattgttttcttcttcttcttcttcttTg
gttcatctatttatctagggaaagaccaatgggcttctctgttTgtatggtactggaggtgctattcgtgatTgagcatgtatgcttct
caaatgtgtctacacgtggatgatcagccaactgtagaaagaatgagaaatgccaatccgattTgtgtgacagttTTTTTggacttt
tgataaagaaatcatcgaatttCattgtaatttcttTgagttccttattcttgggattggctctgttttagtggaaatcaccatcttTcag
tatagTgctacatgctgtgtctatttctgaaagactaaagctagagatggtattaaatagttctgcaacagcagttattagattgtaa
tttTgtacagTTGGCATTCTGTTGTCAATGGAAATGAGGATGGATTAAACTCTGGCAAAGTTTGGTGCCTGACAGAGAGTTGACTGCC
AGTTGTGTTCCATAATGAGGGATGTTGAAAATGTACGTGGCAATACCCAGAAAAGATCGAGAAGCTTAGACAGCTTCTTTCTGGTCT
CTAAAgTgagctcacatattttgag 3'

>CT06

3' cCAAAGTCCGAGAAGGTATAAATTC AACATGCCCTTCATCTTGTCCCGTGGCTAGTACCTTATACGGAGTCAAGCACAAATCCAGG
TTCTCCGACTTAAGTAATCGATCCATAAGAGAGACCATctggataaacctacagcaaacagaatattttgggtgtcccaaacattttt
ccatcgatccaactacttctcttactcaaaagagcaaaaacagactgagaaacaacggcataaagaaacaagaaataatccacgtaa
gccagaaagttaaagctatggaacatatttttaatgcaataatagaggcttaattaccaattgatcttTgTcgaatatcatctctctt
ttaaataattatcttggcacttccacac 5'

>CT08 (03 éxons e 03 íntrons)

5' acatttTgagaattcgatagccagtgtaattttccgtaataaaaataaaaactcatgtgCaggaacgatccaacgaaattaacctgctg
atctagcgggtgggtccccgcccgggtgtctctccggcgccaccgttgcaccagcgcgggcttccccttctcaaccgaaaccgaaactcag
cgtccggtgctcgggatcggcgaacc caccgcttTgtccgacgcggccctctcgggctaggagattcgcggtgctctctgtgtccag
ccccggcgaagtgcagcagcATGAGCGGCGGCAACGAGTTCCGGTCTTCTCTGTGCGCGCAGATCAACCTCCCCGTACCTTCCGGGT
CAAGAACTTGGAAAGCGCTCTGCCTCCTAAGCGCTCTCCAAACCAGGTAACCTTTTTCGAGTTTGGCTAGTAGAAGAACGAGAGAA
GGAAAAAATTCGAGAAATGGAGTGAATTTGAGGTGGGTGAACCTGCTTGTATGGTGAAGGAGACTTCTGTGGTCTGTGTTTTGAGGAA
AAGGCTGGAGACGTGTGCTTCTTGTGTGgttaatggttaagcggaccgagatcttttagagctgcccgtggctttTgtgctttcag
ctgttatataccgcttccactgtattccttctgCagAGATCGATCATGCCGCGGAGGAGAGAAAGGCGGAGCTCTACGTCGAGTGTCT
TCTGTACATTGACGGTGGCCATTGGTCTTCCCACTAGAACGAGGTGCCACCTAAATTAATCCGCTCATGTCGGCAGCTTTTTT
CCCATGCATGATCATTGTTATCTGCCGTGCCAAGTATTACTTCCAGTGACATTTTTCTTCTATTATATCTTGGTTCACTATTTgt
gagacaattgaaaaaggtttttTgctttaaaggaacatgatggacgacgatgagatattcgggtaaaccttaaaaaatctattcactt
caCaagagacttatattatgtgacatagtctctggggaaggacgttTgctgatttTgtaacacaatggatctctttTgtggtttcatgcc
aaagtggctgttcagaggagtgcagagatgagcgtgagtgatgctggtaccatcagcatggcctatgtTgattTgggaatgtggg
atgcttcatcatcgatcgttttagggcgttagtgacgtTgcatgttttctggcacaactcgtctgtTgcatgaaaggt agcct
gtaaagagacagactccgatcttcatgctgagtggaatgaaagtttaaacctggaaataattTgattcatatgctggcctc
aatactcataattgtaaacctacagttgggtggtgctaaatacatatgacagttctctgctttttttttttttttttctgtTgga
agctctgtgtTcaaatagctcattggaataatatttcatattccttaactgaatgatacatattatctggagcttaattcggcac
ttctggctcatgtttcttagaaagatagggatctgtTgcttccagaggtgcttaaatTaaagccatagactgtaggaattTcaaa
aatagttcgtttTgggtatattcttttagatgagcttatcaatttcttagattttatcttaggagatggaagatgaccaaataaac
ggctcctatgTcaatggccatcaatggcaaaaatattgaaacaaaaaattctcaaggtctttattgacagctagttgctatgagaagat
ttTgtctttTgtcttcattttccagaagacaaagtcaogcaattgggtctgcttctgcttctgcttatctgatttatatactgactgactg
catgggtgggcaataaacatgctgcaaaaataaactgTgggagcttcaatttagttttacttatattatggaatgaaacaattTgct
gctgttacaggtagatgtTgggatccagcatgggtgggtccccctaacagtgatcgattgaaatacaactctttttattctgtcaataag
agTATATGACATGGGTAGTTGGAGATTAGATTCTGTTCAACATCAATCCAAATCTTAGCTATTTCATTTCAATGCCAGGTTGGAACT
GCGGGCCACATATTTTTGGAATGAACCTCCTCAGTACTAGTACTAAATATCGAGACTTAACTGCTCACTCACAACTTGGCTTAAACAG
TGAGTCTGGAATGTTTTCTTCACTTGTGATTTgtagatatacttagctgactttgTatctataccaaggtgctccgcttcttaaagt
acacctctcattacatacatttTgtatctctc 3'

>CT09 (05 éxons e 04 íntrons)

5' ctgaattattggctgaaatccaactcctttTgtggaactatagattcaggaataacagaccatgtagcgaaggtatgagaaatccttctg
ggaaatccgacgataaccgactggaactaagtggatctatgtgggatataactccagaaatgaagtTaaagggattggcactgccaat
TgaaacATGCGTGGTGGACGCATCTTATTCTACATGACGTTCCATATGCTCTAGAGATTCTGTCGAAATTTAACGCTGTGTTTGGTGT
GGTTAAACTTGTTTTTTATATGAACCTCCATAATATAGTGTAGATTGTGTTTGTAGATACAAACTACTATGTTTGTGGTCACTTTCTA
GATGGTTTTATTGACTGAATGTTGACTGTGGTGTGATGCAATATGTTAATCCCTGTTTACGCTTTTTACTTTCATGTGATAATGATG
TGAATGTGTGGCATGCTAGACTTGTATCTATTGGCAACACATATGAATAGACTTGGCAAAGAGGGCTTgttcaacaatattgaaata
agtcgattTgtccatattgtaagcatTgcttagaagggaaacgacaaggaaaccattTggaaaaggtTaaagagctgacgttctctg
catttaactccattatgacgtttTgtggttcaatgaaatgagaggaagcagggactggttacttcatcacttttaatatagatgattat
actcgattcggatggtgtctattTgatttctcataaactgagCATAAGTTGTTTCAAAATATTTATGAACCTTAGTAGAAAATCAATT
AGACAAGAAAAATAAAGTATTGAGATCCGATCCGAGGTCGAGAAATATTTATCTGATGAGTTGAGAAATTTAGTGAATGAAAGGAA
GAAAGACAGCTATCTATTATATATACTCCTCAACAAAATGGTGTGCGAGAGAGAAGAAACAGAACCTTATGGAATGGTTAGGTCCA
TGATGGCCATGCTAACTTACATATCACTTTTTGGAGTGTATGCTTTAACTACTGCCTGTATATTTAACCGGTTGCTTCCAAATC
AGTTACTCCACTTCATATGAGTTATTGACAGGTAGAAAACCGGACTTAAGTTTTTTTGGTCACTGGGTTGTGCTTACTATCTAT
GATGCTCTCAAAAGTTTTGGAAATATGTCCTCAAAAGGAGAAATAGAGTATTTAATGAGATACCTCGAACCTTAAATGGTATGTG
TCGTAGGTGAGCAGGAAAGTGGGAGTATAACTGAATATGAATCACGGGATGTCACATTTCTTAGTAGATTTTTTCTTAAGAAGGGCGA
AATAGGGAAAGATTACTTCATATTTGAAACCTTGGATCTAGATGATGATTTGgttagttcatccaagTggggataaatgagaag
gatgaattgaaatcaactcattctcaattacaaccacatgatgagatgatattcgtcattatctaactcagagTgggagcataatgggga
atgattgtgctaagtgtacaatctccatataagcgaacaagtcgcaaaagttattcccgtTgagctttTgaaattgaaagggaaacttt
tataattgctccgcaagATGAGGATGAGCCGAGAAATATTAATGAGGCTCTTAATGCCCTAGTAAGGAAAATGGATTAATGCAATG
AAAGAGGAGATGGAGTCAATGAAGTCAACCAAGTTTGGAAATTTGGTTGATCTTCCAAAAGGGGACGCAAAAGCTATTGGGAACAGT
GGTCTCAAAAATAAAGCAAAGGCTGATGGCTCGATAGAAAGCATAAGGCTCGCTTGTGGCAAGGGGTATAATCAACAGGAAGGA
ATTGATTATGAAGATACATTTTTCTTgttagagggatttaccctgattcgcattatctggcaaatagTggctagatggatctgagT

tacatcaaatggatgtaaaagactgctttcctcaatggagaattagaggaataaataaataataaacaacctgttggttcatagtgaa
aggccaagaagaaaaggatgtcaactttgggggctgatatatggccttaagcagtcgcaagacaatgatataacgtttcataat
gccaataatagcatatgatcttgcaatgatagatgaggatcattgtatataatacaaaatccaaggatcaatttggatcatatca
tatatgtaaatgatataactaatgcccgaagtaatatggagtttggctcactgtcaaaagttgggtgtcttccaatttggagatgaa
tगतatggggtgttaacccatagacactcctatgtcaaaagttgaagatgagccataaaactgtgtcca**agGACTCCACAAAAGAA**
GGAACAAATGAAACATGTTCCCTTATGCTAGTGTGGTGAAGCTTTATGATGCTATGATATGTAAGACCTAACATATGTTTGGCA
GTTGGAATGATGAGTAGTTACCAATCCAATCCAGgtcaagcacatgataagccgttaagagaatactgaggtatataagggaaactg
ttgattatagctgagtcattgaggaaaggatctgcgactctaaagctattatgatgctgatt**gagAAGGAGATTTAGATGAAAGGAA**
GTCTACCTCTGGGTTTGTCTTC 3'

>CT10 (05 éxons e 04 íntrons)

5' atatacatttggctcctcaagtgatgaggaatgggttcatctgcagatgcaaatgggttcataaaactttatccctcaaatgtggg
tcagagagaatttgatgcttgacccttgggtgtaactctatggatctcaattgaaatttatacctcaattgagagcaaaaaattctgt
tttgacataaaaaattttaggtctggggtgttcttgggaaagaggaggggtgattgggtggggtacaattcttctatttagcagc
aaaaagcagctcaaaactgggaagcagaagcttagactttggcaaggaaggaagctgatggctcattccaacaataactcctggaa
aggttctgcttggttacatgcttgctgggtgattacaactttcagattgggacagaaagcaatgatgatttgggtacttgcacaatgt
ttt**agGTCCCTAAGCATGAGCGGGGTGAGCTGGAGCGCTTGGAAAAGCTTATGAATAAATATGAAAAGGACAGATTCACGAATTG**
ATTGGCTAGACCACCTTGCAATTTAAAGCTATGGACAAAGTCAAGGAAACGGGAGAGCAATAAAATTTGGAAATTTCCATCTATACCTGGT
AGTTGACTTTTGTAGCTTTGAACATCAAGTTGTTTCCAGgtaaacattcaggaagaacgctataacaatggttatctctaaagg
gtttggttttttttggctcgagttgtgcaacattgctcagcaaatctgtcatttagaccagacagcaagcttgaagcacaatgta
tcaccgctcatgactttctctgtccatctattttcttctcctctcctcattcagcgtcttactttgcataaagtataaggcctctca
ttctgtgtgatgggtgctgtgtgaaaaaacatttagacgctagatattgggttatcagacattatggtggatgtagatcttcca
tगतataagcaaatgaaaaataagatgtgtgctgggtgactttctctcctgggttaacacatgctttgttaattgaaagcgaat
tcattttcattttccacagtttggttatctgttctgcacagaaaactttcatttagcctgtaactgtcagttagtgctgggtttctca
tctgtctaaagctccttctcctggcacaacagtaggaaatccaagagacatgccatagcttaatatgacaactatctgttgtatt
ggtaatatcatttagtaagctgatgcaatataagcattgggttctgtaagagtggtggaaggaaagggagagagatatacttccc
aaagatattggcctattgggttaattgaaaggaagaaatgtcgggaagttagaggaagcaaggaagcaatccttataggacaac
atgatagggaaatgcaaccggcactttccattcactacctaacttttggttgaaattggtgcaatcaaaaggtcatttaggtcc
attatctagtttctatttggttttgtattttcccttaaatatgcttttagtgaaggaactaagatattttttatggtttctttaa
tttcttctctactttgttaattctcactccttaactcctaattgtgaaacatgatgtaagaaaagaggttgtattcactcact
gatagatctgcttatttggatgtgcatgcagagtgactggagtttagtgctgtaactctctgtgactcaatctccgtattttctt
agatgagggatttcttggttccggagaaaactctcttggatgctacttttaaggactatatttttctcctaagaaaggtttctctct
actacgtgtgatagatattcaatgtacttctgtaacccttagtctcacttactacaaaacttctgagctgtagcatttaatttagcct
cttgcaactattagcttacaccatggcaaaatataatgctgagatgagacgttagaagtcattctggtttgatcaacattctcctaca
tagattggcgtatgatggctgttgccaacttaatgcataatgaaagctcttaatgctcaatccttactttgggtgttttatctat
ctgtcgaatttgcagagttgatgccccttttgaatctataaaaggtgcttgatgtgcttagtattgctttatcagttttccgg
gcttgggtggctcatgtggctatttcttggcaagacactatgttctgtgcatcagtagcaagttgacctccacaaaaacatttaa
ttctccttttggtaaaaaggtgattatgggtcgattt**agGAATCTGGGGCGAATTTCTTATCTCCACCAATAGTTTTAACAAATGA**
ACTTGTATAGTATGGGACCTGAAGTGGGAAGGATAAATCCATCTGAGCACAAGCACTAAAGCTAGCAAGGAGTTTAAACCGTGGT
ATCATTGCAAGGATCTTAAACCAAACTCCACTCAAAGGAAgtaaagctgacttttctttagtgatagaactgtgatacaattac
ggattttagtagtataagggcaaatctctgctcttttctctcaagcagctgtcgtcatcatatcatccagaagaaagatacatg
cgatataaagttgcatgaaatgtagagctattatgataaaacctcctgtagataaaggctctatagtttccagctattctgtgtt
gaccgctgcttctaggcgtatgtgtggctatgcatggtgataatgtagatagagtggtcattgattttaaattggccagaacgcaag
tgatttctcaatgggataaattccgcattgtagtggctcagttatgcatcagttatcttaaatgaaattgaaaggttttgggtttctt
ctcagtttgaccacacaagaaaaaacctcttaaatcaattatcatgtaccgttggtgactttgaaatgggactctgatttacgtaaa
gatcattatttcttattccactcttttcttccattaaagaagttgggtgaaagccgaagtaggtgaaagggagtaaacactttat
gttaacatagatttggctatatacttgtcccaccatttggcttgggtgaaagttgttccctgctgctgcaagagcagactgccacc
aactgaaacatcatagtaacctactcagtttggcgtatctatgagttcagtaaatattatcctcctcagcaagcaatcagtgctta
gtcacttgatgggtgaggtctctcatggttagttattttgaaagccattttctcatgataccttaacatgcttgaagaaacttcta
cttaaaatattaacctgctgccaatataatgcatgaaccaaagagaagaaatattgagggcgtttgagatgtgttggccttgttaaat
atttaattgaggtttatgggatacttcaatgtccagatccagctctcttggcaaacagttagattgaaatgagttgcaacctcaat
tgctacttcaactgacttaaatcaagtcaacttaacttaagtcagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa
ttgaggttaaaatagctataaaatcttgcacaaatcatttttttatttggtttagttagtcatcatatgcttctgataactcgtt
ccactgctaatatgtacttctcttcaactttcaactttcagctgataaaaacctgctcaatcttttgattgta**agGTCAATACAA**
AAACTACTTAAATATCTCTCTACAAGGACTTTAAGTGGAGACGAGGACAGCTCCTTTGGAAATTTGCTTTCTTTGATGTCTGAGA
AGAGAGCTTCTCACAAAGTTCCCTTCGCTGTGTTGAATGGAGTATGTTCCAGgtcagcactatttggatgctgcttataattg
tattcccactaatgctgtgactggcacaatcatttctgctcctcatgctctgagatcgctctagtttggctcctacttgtctgta
taaaagaatattaggttctcttgcataatgcaatgtgcaatcagctgggtcaaatcatgcaagacataaacagagcatttgacag
gctttaggattgcaagatatttaattgcaagcaggttaattgcaaaagttttctgggtcctagttaaagcactcttgcaggaacta
tgcaagtaaaaaccttaaatagagatccaatgatcattgatgtagcaaccgaatcataatccttggttcttaatttttttccctt
tcttgc**agGAAGCTAAGCAGGCATTAGAATTAATGGACCGTGGGAAACAATTGATGTATGCGATGCCCTAGAATATTATCTCTGT**
TTTCAAAAAGCGAGGAGgtaaaggttttgaagaggtttgttcttgggttctcaggaataatccttattctcacaatgctagcagaa
aattttgggttgggtgctttatcacataggatacaggatacaacctgataattggaatctctctttagtactctgcttttagattat
cagtgatttccattaaatactttcagttacaagtaactaaaaggaataaaaaaaatgaaagggattataagtttgaagactcagata
gttggcaaatatgagcattccatgctgtgacccccccccccagccagatagaggttgggtgtgtgattgtagttcattagca3'

>CT11 (05 éxons e 04 íntrons)

3' cgagtcataatagataacccgatctgctgactggcagcctgagatcctatctcgggggtggccatgaatcgaatcggatagaa
ttcttttgggtcaaaaatcggatgagacgcttacttaactcagagttcttggcaaatatacaaaaagcatabaagcacaatacaaaatcta
aaatgcataggggggttagtaacccaaccactaagaagacactttatgctgtgggcttttgcagcaaatgtgctgttttatctgctc
tcggctgcaatgggctattgtaagggccgaaatgtctgcaactttgtctggccgctccaccaatgggtgtgcatcccaactgaa
tctctcaaaagtgaaatgctatgcaagagcctgctgagtcagattgcttgcagcttcagtttagttctgttcaggggaaacttgg
tggagataaaatcaaggtttgcacaagcaccagcctctgcttggctcggcagctgattgcaagcatttggcacaaccatttag
gggtgctttggatgggcatttgaaagccattcagcccccaagttctctcagccaaaatattgggtgtttggcacaaccattcaaga
agctttcaaatgaaagccagccttggaaatgctaaaaacatttttctcttcaaacatctcatttaattctcacaatctcctggt
attacccttgtattttattttttaaagaaaaaggaagcaacctctcggcggcagcagatctggctcgtgggtgctcagatccct
cgccggtccggcgaaggttttagctgaaattttcaaaaaaaggaaaaagaaaaagaaaaagaaaaaagaaaaatatttatactcat

caaaagcccttgcgaaagtttttttaccacaaatgcacattcaactcaaagtccatttgcaaatttttttaccaaacacaattgcatt
 ttccccaaagtcttggtctttcagcatttgcatggactcaaagccattgcaaatgctaaactaaacgcagccttactaatgtgc
 caagggagttggcagagtagtagatataaaatatacaactagtatttcaaagtaagtttttaagtgattcaataaataca
 aatgtggcatttagatgataaagtcaaatttctcttggatttattttacagtagatttctccgtagagattcgacatcatataa
 tatgaatttttagtgacattacaaatgtaattactttttatattttcatttgaaatattttaatgtgcaatttttactttcaatgatata
 tatgaattagataatcatctttggcgtttgggtttacgtttctattttgtttcccttttccatggtttatatataatttctgt
 caagaactaaaatgggtcttgatttatacaaaataatgataatataaaataaaccataactaaactaaacaaagagcggcagtgttaa
 tatgtggtttaataaaaagattttccacaaaataaaccatatacaatgaagatatatacatttagttagactaaact
 tttgttgacgtaagtagaagtaagcttgacttaacaaatatataaactttttataaactagtaattaagatatatttcaactatgagtaa
 tataagattgccttatgcaaacatttagataatcacaacttaagtagatggaactgatatatttactatttcttgaattatttta
 gaggcttttctcctatgaaaaactcaagatttgaaaaaatattttttgcaaaataatttggttatatcgcttagaagaatagttaa
 taaaactttttttcatatcaacacagtttagactaacatttttgggtgagtaaaatatttttattcatataatttttaag
 cgatacaagtagctattttttcgaaaagtgttttccaaactctggacttttgtaaacacacgcctcttaacataaatatgattgtg
 atacaacacatacttttagcacaatatttttagtaaaacatttatatgctataatataaactatgcaaaaatattgtaattgggt
 ttagaattacataaaaataatacaaaataatcacaataatgtagtaacttagctataaattataatattcaccgcttagcctggc
 gggtaaaatgagagtggaacggccaccccttgcaacacaccatttgcaaacctcttaaagctcttgcttgaacaaaggaagcaaac
 tatagaataatcctaagtagaagcgttagatgtccatacaaaaactcgcagcagacgtaaacggggcagcaagcaagggagactccgt
 aagaaaaatataaaaaactatataatgataataatctTTAACATGAGTCTTGACTCGACCAATCCCCTGCTGAGCCGCTCCC
 TCTGCACACAGTCTGGCGGGGAGATCGGCGTGGTGCAGCTGCCGAAGGTCTGCCAGCAGAAGGGGGCTGTGCTGAGAGGGGG
 CTCTGCATGATGGCCCAAGTCTATCGCCGACAGGGTGTATTTCTCGCAAGGCACGCTGTGCTGCAAGCCAAAGTCCACTGGCC
 TCTGTGTAGTCTCGTGTAGCCTGCTcgttaagcagcgtctcgcaacgcggcagctgatgactggctcgtgcaagctcgctgagct
 agtactggtcagtgatgagtagggccatcggactcggagactcgaatggttgaaaacttcaaccataaacggactcctgaccacccgtg
 catcgcaacaattctcgttaaaaagacataaaaagaacttaaaatgtagggttatataaattgcaatttcaaggtcgtctatggt
 tgttcaggagcgaatttcttttttccctttt**ac**CTGCCACGATTTTATCTAACGCCGTTTCGTGTCCCTTGACATCGACGCTCTC
 AACGGTGTATTGTGCACACATGCTTTTAGAGTGGTCTTTTCCTAGGCTTCAGGATGCTGCCGTGCCCCGGACCGCAGTCACATCG
 TGTATGTGAATGTTGGAACATCCGCTGACGATCGAATCGATCGTCC**C**ctgtatgaaacatggttaacagaatctaaactcgaacattc
 tctaataatgaaagaacgtaaacattcaatctatgactttccaaaagaactctcgcaagaatttgggtt**ac**CGCAGGCAGAGTAGC
 ATCGTGATCTGCAGCTCTGGGAGTTCCEAAGTGATCCCATCAGTGTGGGGCTGTGCCCCGGGACGAAATGCTCATGTCCGCGC
 ACCGAAACCGCCGCTGTGGTCGAACGCTGAGGTGGAAGTTGGGCTGTCTGATCTTGTATGCCGCTGAACTGACATTGAGACAGT
 GATCGAATTTTATTGCCATCAACCAAGCCATGAAAAAGAGAGGAGATCCACATCAATCTT**C**ctgcatgatcacagctttggaaa
 aagtcaggcaatttatagcagaagtagatcggtctgctgctattatcagctgggttggctcggggcattgaaactcaactcctcttcg
 ctgcccacaagtctgacaagagtagacatggaacacagatttagtgaattctcaagcaatgatctgcaagaaaggcatatggttac
 ctgcccacaacccacgcacagcagatagtagtagttagttagttcccaattgtaatacaaaaatgtagcagaagaaatcaaaatcatgtg
 caattcaagtaaaaacttttgaattaaatcataaggttaaataaaaacacataaaactcgaaaaataatgctaaaaataactttttaattg
 tcaaaatcaagaataaaaaggattagagaacgacaacaacacaaaactatggaactataggaatctcagatttttctcatcagaat
 atggtctgttcataatcggaatattttttgttggttatataacatggcatttttaagacaataatgacttatttagcgggttaaag
 aatctggtggagtcacaacaaataaaaaatagtggtcgtcaaaatgaaacattttaaattagcagagcgggaagccaccgctggaagggtt
 gt**ca**CTGTGTATCGGCTCCTCGAAGTCCATTCATCTCGCATCATTATCGACATCGAAGATCATCATTGAGATCCCTCAGAC
 CGTGAGCGCCACCAACCGGGCCTCGGCCGTCGATGCTTCCTCGCATTCCTTGAATGTGAAGCAGTCCGCTGTGCAAGCC
 AATGCTCGCACCCAAACCTGGGATCCGTCGAGCAACTACCGTCCCGTCTAG**C**tgaaaaaataaaacaaaccgtgtagtg
 tgatcaggagcagattacagcaaaaggttttttttttttttttgatagccagtgcaattggttatctgaacaaatataaaaaga
 atggtcaattctaaatctccccctcttatttggtgctatatctg**ac**CTGGAGTATAATGTTTCTTGCAGTGTTTACCACCA
 CGAGATGGCCCAACAAGGAAGACATGGTTGCAGGAAT**C**ACTcgttgatgctccacggttgatggaagctccatctgg
 aatgctctgtgagcagcaacagaaggaaataaaaactcaattgcttatctagcataatcttaacgccaacaagagatcagggagca
 ttaggttttttataatggttagacttaactgtagtagtgcctttccagttaagcaagctaccgattagattttgattactgaactc
 ggctactcaagtgatttggaagattatcaaaaagcttttaaggtttagaaccccttaggcatttaggtttagtttaaagcagcgg
 catgggtactctgggacttttttgcatcataaatttaggttttacttctcttctttcttttctttctttcttttctttctttc
 attatatttaaaggacatttctagaacagcataaagagtataaggtatagaaatgaggttagagcacaagtgtagacctctgatca
 taaggttcccattgcccattggtgccccaaatctagcagcttgaaggtggttagagttagctggcactgatggtggtggtggtgccc
 ggtggtggcaccggaagtgtggtagaagcaaggggtgtgatactgcccagtggtgctggtgtagcgggtgagccggagggtgtggtc
 tactggaggtgtggttagcactaccgggtgcatcagtgaggtggtggtgagatctgggtgaaactcggaggtggttgcacagcggggtg
 gatgcccagggtgtagtgaagtggtgaggtggtggttaggttagctcgggtgcatggggtggtgaggtggtggtggtggtggt
 ctgataactggtgcccagcagctcgggtccaccattggtggtggtggttatttgccttggggggaccatgg 5'

>CT12 (08 éxons e 08 íntrons)

5'caatgaggcgggtctacttctgaaacaattgttttctttgggtggtgtagcaattctgaaacccatcactccacttttgc
 aatttatacactaatccatgacaatgacactttcgtaatagtagtgtaaactgtagctctctctctctcctgactgctccttttct
 ctctctctctgagcagagctctcattcttgaaattctcaacgatccacgatttcccttagccttttcaactctctcgtctctgctctg
 ctggaaaac**ATG**CGCTTTTCGGCGAACGCCGACCTCTCGACTGCTCTCTCCCTCCCGAGCTCACCTGCGAGACTCCCTGCCAGCC
 TCAAGCTCACTTCTCAACGGAACTCTCAGCCTCTCGACTCTCCATCGTCTCCCTTATATAGACTCTGCGTCTCTCAGCCAGC
 CTGGAAGAAGCGCAAGTGCCAGCAAAGGGACGAATTCTCGAGAATCAGATGGACCTTTCTACGAGGTTTTCTAAGTTTTCATGGA
 CGAGACAACCGTCATGGATTACCCGATTTTCTTACCGTGGCATGGTGTAGACCAGGGATCCTCGTCTTTAGGGATAATGAATCCCTCC
 ATGTCGGTCAAAGAATCGGCGATGAACTTCTACAAGCCATCCAAACTCCAAGATCTACGTTCCCATATTCTTCTGAGAAATTAAGCTC
 GAGTCACTGGTGCCTCAAGAGCTTGCAATATAGTGGTGGAGTCACCTCAAATCTAATGGGAAAAAGAGATTTTGCCCAATTTTCTTG
 GATGTGGAGCCTGATGATGTCAAGCTGAAAAAGACTTTATACCAGCAAAGATCTTTACAGCCAGCACAAAAGTTTCCGTAAGTTG
 AATCATGAAAAAGGCTCTCATTGAAGTGGACAGATAAAGGGATGGAACCTGAGGAAAGATGAAG**G**taaatctcaaaactttcaa
 ttactcatttgtagactctcattttctctctatttagtttaattgcaaggtgtagcggatgaaaaatattatttcatctctttctttt
 ctctctcttaattttatttccgattttacagtatttcaagttaaaatgagagaggttttctttaaataatgagataatcaaatcaat
 gcataaaaaatcagtgactacactgctaacgcaaaaagtaaatagagaataacagttttaaacaagaataaaaagtagcaggaataatgt
 ataacagtgtaagtgttcagaaataatgtagactacctgactttaaggggatttgtgttagtgcctctgtatcttccatccctcac
 ttagatggcagtgagatttttactttcgatttcataagcattgaataaaaatacaagaagataaagggccagtggtttaaaggaat
 ttagaataaataatgatactttaaagaagctcggcagctcggcagctagctcctgag**cg**CTCCACCAAGCCGCGCCGCA
 GCCATCAGCTCTCCGTCAGTCCCTGCCACTTCTTTGTACCAGCCGACACCAAGCCACCAGAAGCCGCGCCGAGCCTCTGGCCCC
 CACCCTTTCTTCTCTACTCAATGGCAGTCCGATTCCGGCTCCGGATGTG**G**cttaagatcacaagcccaatcagcagctgccc
 gacttgggtcaccacaataat**ag**TGTTAGCGAAGGCCAGCAAAGCCATCCCCAGTGGTCGCAAGGCTGCAACCCACTTGTGTAAG
 GAGAGCCAGCGAGATGGCCAGGTGGATGGGTGCATCAGGGCCGACCTGCAATGTGACAGCTGACCCCGGGGATgtagcaa
 cgttatcggcggttggtaagtgccaggggcccagggtgtagctaaattggggtagtgagtagactgggaaattttgtagtg

tttagcgtctagatggttttgcgtaattctgcaaatggaagataatgagtcacttagggagtagtttgcaatataaggatata
tcttagaacggttggttacggttttgcagGCAAGCAGATCTGACACAATCCATAATAAAACGTGTTGGATAAGCTGAAAGTCAGATA
TAAAGGATGTGCCTGAAGATCTAGTTGGAATTGACGATGTGTGGAAGCTATAATCAAAATGTTGGATTGAGATCTAGTAGCGTA
CAATTTCTCGGAATCCACGGAATGGCGGAATTGGCAAAACAACACTTGCCAATGTCATTTTCAACTGCCTGTCTTCTCACTTTAAG
ACTGTTATTTCTCCAGATGTCCGAGAGTTGTGCAACGCAATGGCATGCCCATATTTGCAAAAATTTGGTCTATCCAAGttccttggg
tcttagttctctatcttgcaaatatcaacacacaggtgagggataacatgataaagagagttcttgggtatagAAAAGTTCTCATCTC
TCTTGGATGACGTGGACGGGAAAGAGCAACTCAAAGTCTAGCAAAAAGAAGGTGATTGGTTCGGTCTGGTAGTAGGATTATCATAAC
TACTAGGGACCAAAGGGTCTTAAGGATTGTGGGAGAAGCAACAGGTGAAGGCCATTTAGAAAAGTCTACAAAAGTTTGTACTTACGAA
GTACGTGAAGTGAATTTGAAGATGCTCTTAAGCTTTTTTCGTAAGCATGCTTTTGAAGAGACTCTCCACCAGATCATTATGGCTCCC
TTTCAAATGAGATGCTCCACTTTGGGAATGCTTCTTTAGCCCTTAGGTTATAGGTTACATCCCTTAAACAATGAACCCAAAGAAAT
CTGGGAAGCCACTTGAAGAAGCTTAAAGATGCTCTCCCGAATGAAGTCAAAGTAAAATGATGATATCTTTTGATAAGTTAAATAAT
GAACAAAAGCAAGTATTTTTGGATATAGCTTGTTTTTTTGTTAATGAGGATAAAAACATACCCTTTCATATGTGGGATGCTATGGAT
ACCATCCAGACGTTGCCATTAAGTCTCTTGTCTCATGCTTTGATCAAATAAAGATGGCAATACTTTTTTCGATGATGATCAAGT
CGGAGACTTGGAAAGGAAATTTGACGTCAAGAGAACAAGAATCTCTGTAAGCTAGTAGAGTGGAAATCATGAGGATGGCTTGAAC
ATTTGTGAAGCAAAAAGGttatagaaaagggatagcaatttaatttctatgtgaattttttagatattctcttgcaaaaggaatttata
aatgctatggcgcttgcaagttgcttggatgtttgaccatcacacgtggacttcttgggtttccattctcagGGAAGTAAGAAAATAGA
AGCACTGTCAGTTGGATCTGGATTTGAGCGGGAAGTCATTTATTAAGAGATGATCAATTTGCTGATTTACGAAACCTGCGGTTCTTT
CGTGGGGAAAGGGTGTCTTCTGTTGGAGATTTCAATAATCTTCTCAAATTTAAGATGGTTTTCTAACGTTATTGTTCTTCTATGC
TGTAGGCAAAAACCTTTTGTCCAACTAATTTAGTCTTGAATTTGATCCAGCGACATTTCCAGCGACATTTGTTGATTTGTTGATTCAA
Ggtatgatagcaggtctcactccttgtatctcataaatcgtttctttaaattttttttcttttttaggtcatttctcaataaatgttata
tcaaatgttctgcaagtagcaaggaatataacgctcctagatctaaaggctgctataacttaaggagaacactagatttatctatg
tcggtgctttggagattttgagcttatatagttggtgaattttggaagatgattgattgataaagatgatacaagtcctcctct
tatacagataaactcgtcttcttcaatttttttggtagCGGTTTTCTCAAATAATTTTTATCGAATTTTTCAGGCAAGCAAGGAAA
CTAAAAGTCTTAGATCTTAGCTTCTGCTATAACTTAAGGAGATTGCCGATTTATATACGTTGGTGTCTTTGGAGATTTTGGACATTAG
CAAATTTGTCATTTTGGAAAAGTATCGATTGGATCGAAGtagcagtagaggctcactccttgtatctcataaatcatttcttttataat
tttttttaggtcatttctcaataaatatttattgttttgcagggcacaaggaactaaaagtcctagatctcagctactgccaacact
tgaggagatgcccagatttatctacggttgggtctttggagattttgaaatttaataagattgtagattttggaagattcagattgggat
cgaagtagcagatagcaggtctcactccttgtatctcataaataatttatttttataatttttttaggtcatttctcaataaatattatg
ttttgcaaggggccaaggaataaaagtcctagatctcagcttatgctttaaacttaaggagagtgctgatttaactacggttgggtgct
tttggagatttttagattatgatttattgttcaaagATAAAGAACTTCCAGACTCGATTGGGAAATTACAATCATTAAATCGATTTGGAT
TTGTTAGAAACGAATATTGATCACCTACCGGATTCATTAACAATCTAAGCAACTAACCAACTTAGGATGAGGAAATTAAGGGAG
GGTTGACAAAATTAACCTAGTGCAGTCTGGTTGTTTCAAGTAAATGGCTTATCTTTTTCTGTTCCACGGTAATAAGCTGTACTGTTTGA
AATGCCAGACTTGTCTAGGACCAAATATTTGGATTACCCACCGTGGCGAGCTCTGCTCTAATATCCGTCACCTCGGATGGACGCG
TGGTATAGCAGACGGCCAACTTTTATTTGGAGTACCTCGAGCTAGCAAACTCAGAGGGAAAAACAATTTGGCTCCCAAAAACA
TGGGTTCTCTCCCTCAGCTCCCGTCTAGTTTAAAGAGACTCAATATTTCCAGAAAGTGTGGAGTCTGGGATGTCATAATCGGTTGCC
GATATTCGTTGGTATATGTGATTCGGAGTTGGAGAACTCCATCTGCAGGATTCACGCTTTCAGACGGATTGAACAGTGTGGCTG
AGAAACAGGAAATTTCCCGTCCAGATACCGTCCGAGGATTCGTCGACAGCGGCATCAGCCACGAAAGCATTGAGCCGTCGCATATGA
AGAATCTGCAAGAGGTAGATCTGTTGCAATGCGAGCTACTAGTCAGATTCGCGGCCTCGAAGAATTTGGATCTCCCTTCACTCTTGC
TGCGCATATTGTTAGTTCGATCGAAGAGGATGCCAGACTTCAAGAACTCAGAAAGCTAAGGAACTTCCAGTGGGAAATTTGTCGAAG
CTAAGAAGCATGGAAGGTTTGGACCGCTTGAATCTCTAAGAGTGTTCGCGATTTCAGATTCAGGCTTTTGGAGAGTCTGGTGGACA
CCTCGAATTTGGATTTAGAATGTTGCTGATTGCCGTTGCAAGCGGTTGCCCGACCGTCGCAATTTATTGTAACGTTGGCCGGGGTTG
CTATTGTGAGATGCTTTTTCGAGGGGTTATGGAATCTGACCCGATGGAAGTCCAGCAAGTATTACAGGGACACCCGGATGTTGCGCCA
Gtaaggaatttagaacaatggccttgggttttgggtgtagcattttggttttcttctggttctggttttttctgatttttctgatttttctgata
aaagcctttctccgggacaata 3'

>CT13 (04 éxons e 04 introns)

5' ctatcttttagtttcaactttataaactttatttttatttttattttttttggatgaagtagtgttttcttcatccataaaaggaac
tctacaagagctagacaagtgagttctgaaagcctgaaatctatacaataatttgggtgctgacttctggttgggtgctaggtttcagat
tttaatttttttcaacaaatgggaccagaaacttcaaccattgattgattcatggaacttttaggaaacttctcagaaacttct
cgtatgactcaagggcattgactatatttactcaagggatattcggagatttctccccctccatgatgccagagactgagagctcatca
tatgacagcctcacttggatttaagcacaccaagcaaaacttctctgagatattcgggtggttagagatccttccATGGAGAAAAGG
CACAAATTCACCGGCATCATCCGATCTTCATATGAGGTTTTTTTGAATTTTCGAGGAGCAGATCTCGTAATGAGTTTCCCGATTTT
CTTTACCATGGGATGGTGGTGTGGAATCCTTGTATTTAGGGATAGCGAATCCCTCCATGTTGGTGAAGAAATTTGGTGTGAACCTTC
TACAAGCAATTGAGAATCCAAGATTTACATTCACATATCTCTGCAAAATATGCTTCAAGTGTGTTGGTGTCTTCGAGAGCTTGGCTA
CATGGCCGATGTCACCTTGAATCAAATGGGAACAAGGAGATCTGCTCATTTTCTTGGACGTGGAACTTGTGATGTCAAGCTAAAA
ACAAACTTATATAAACAACACTCTTTCAAAGCGCCAGAAGCGTTTTCGAGTGAAGCAGAGTCTGGAAGAAAGGTTCTATTGACGTGG
GCAAGATAAAGGATGGAACCTTGAAGAAAGATGAAAGGttgaatgtcttaaaacttctcagttggttacgtagatttctctctcccccta
attagttcgtttgcaagtttgggtgaaattatagtttatctcctcgtcaagcattgggattctcagggtagtagccaatagcgg
ccgacaccaatacgggtgggtggcacaccttaccattcttaattcttttctcctcctcacttctctcagatagtagccaagcaata
gtaacatcttgagcagcttcaaccaagccagccagcagcactgctacaactcatccatcaatcctcggcatttcttggcggccct
gacaccaagggccagaacctgcccgtgaccatccaaactcccaccttctatccttcaactcactggtaagtcagattcaggctgggt
ttgacgtcacttggcctatatttgcattgcccaccttggggaccccaaatctgggtgctcctaaagccggcgattgcaagggccaaga
cccatgtgacgtggaggattgagataaaggtgtgattggcgggtcgattgggtgtgattgggtcaaggacttggaaatgcaatgggtgatag
tgcagggagggcaagtgatgtcaacagtgccatgattggcagagctgacggtgattggtgattggaagttggggagggagagagatt
gaaaaaatcagaaattcttagcatctagatttttttcttctgtaatttctcaacatgtgagggcaattgaaatcacttgagaagtagc
tttttacaacattaaagttgctccttaacatggttggtaacatcatggcagCCAAGTAGATTTGATCCAATCGGTAATTGAAGCTGTTTT
GGTTAAGTTGAATGTCGGATATAAAAAATGTGACTGAGCACCTAGTTGGAGTTGATGATCGTGTAGAAGCTATAATCAAGATGTTGGAC
ATCGAATCTGATGATGAAATTTCTCGTGATCCATGGAATGGAGGTTTGGCAAAAACAACACTTCCAAAGTCTGTCAGCGGAT
TTTCTTCTCGGTTCCAAGGCTGCAATTTCTTTCGATGTTTCGAGAGTCTCAGAACGACATGGCCTGGTACATTTGCAAAAACAAT
GTTATCCAAGTTCTTAATTTGTTGTTATGTTGACAAAATCAGTGAAGTGGATCATGGAATCAACATGATTAAAGAGGGTACTTTCATAAT
AAAAAGTTCTCATTTGTTCTTACGATGTGGATGAAAACAAGCAGCTGAAAAATTTGGCAGAAAAGGAAATTTGGTTCGGTTCTAGTA
AAGGTTAGATGATGAACAAGGCAAGTATTTTGGATATAGCTGTTTTTTTGAATTTGAAATGGGATAAAACCTATCTTGTGTCATGTGG
GAGGACTGTGAATACCACCGTACAATGCCATTGAAATCCTCCGTCTCATGTCTTGTATAAAAATCAAAGATGACAATACCTTTTGGTA
TGATGATCAAGTGCAGGACCTCGAAGGGCAATCGTTCTGTAAGAGAAATTTCAAAGAACTCGCAAGCGTAGCCGAGTGTGGAATCA

Anexo 3 – Continuação

TCAGGAAGCCCTGAGCATTCTGGAGCAGAAAAAGgtgtcgaaagggaaatttcaacttatttctgtgaattggtgtggcctattcct
gaaaagggaaaaaatcatccctattagcgcttgaaattgtttgttggacctcacatgagcttctggtgttccatcattagGG
AAGCGAGGAGATAGAAGCACTGTCACTAGGATCTCGTGGAGTCACTGTAACGCCTGGTGAAGTTGTCTAATTTACGACACTTGAGGTTT
TTGGATGGGAATGAGATGTTCCCTTGGTGGAGACTTCAATAATCTTCTCCTCAATTTAAGATGGCTTTCTTGGCGATGGTGCCCTTTCG
AGTTTGTGGCAACAACTTTCATCTGgtcaatttaactgcttctgtaccttccatagctatgatactgagaaatggatggcgtggaa
cfaatcaaaagtatgatacgagctctcaactctgataaagtaattctctgctcttctcaatcttccccataaaactgcttct
taatataatatttttgt**ag**GTGGCAAGTAACTAAAAAGTACTAGATCTCAGTAAATGCAACAACTTAATGAGAACGCCTGATTTAT
CTATGTTGGTGTCTTGGAGAGATTGATTTCTGAAGGTTGTCGCAATTTAATTAATTAATGACCCCTTCCATCGGAAATTAAGCTCCT
CAATGCTTTAACTTGAATGGATGTCACGCTCTTGAAGGATTGCCTTACGAAAATAGGATGTTTGTAGTGTCTTGGACGAGATTCTCATG
CCTCATCTGCATATGCACTTCAAACCTCCGAGAGCAGATTGCTAATTCGAAGTCAATTTGACCCCTGATGATCACATAGAGATTGCGCACTAAT
GCATGCTACCAGACTCGATTGGTAATTTTGAAGTCAATTTGACCCCTGATGATCACATAGAGATTGCGCACTAAT
TGGAGGCTGGGAAACTTAGGCGGTTGAATTTATTAAAGTGTAAAGATAAAGAAACTTCCGGACTCGGTGGGCAATTACAATCA
TTAGTCAGATTGGATTTGTCATCAACAAGTTTGGGACACTTACTGTATTCAATCGGCAATCTAAGCAATTTGACAGTACTTGGATGA
GCCACATAATTGGGATAACAAAATTTGCCAAGACGATTGGTGTAGGAAAGCTTGAAGAGTTAGAAGTACTGATGGATGTCACCATT
GACCGCGGAGACTCCCTGAAGAAATTTGGGAGGTTGTCATCTTAAAGTACTTAGACTTTGTCATTCAACAGTATATCCAGATTACCATC
ACTGTGCATTACTTCTAATCTCAAAACACTTAATCTAGAATCATGTTCTAAGCTTAATGAGTTGCCACAGCTTCCCCAAGTTTGA
GTTGTCTAAGGTGGACCTTTAATACTCTCAAGTTTGTGGAAATAGGACATATATGCTAGAGCGGGACTTGTCTGTCACTCTAGCAT
GCTACCCCTATATGAGGTGGAAGAAGACAGCGATTTTGGCCCGACCTCCGCTCTTCTAGTTTAAAGAGATTGGAAGTTGGAATCTG
TCTATTAGACACGGTGGGATTTTCCGTTTGGAGCAACTGTGAACCTTGTGCGATTACAAAGTTGTCATGATCGAATGCTCTTGTG
TGAGCCCTGAGGTGGTCTGAGCGCTTGTCTTGAACAAAAAGGCAATTGGGTGTGGTCCGTTTAAAAATGTGCCAAGAGCAGT
TCTTCTATTAGCAGATGAAGATTCTGGAGAGGTAATCTGTTTTGGATGCTCCTTGTAGTTCAGATCTGGGGCTTGA**ag**aattggg
atccctttgtctccatcgatgtgtgattgcaattccatgaaagagatgcaaaccttcgaaatttgagaagctaaaggaagctcag
agtaggagaatgccccaaagctaaagaagctggaaaggtctaaactgggaattcttaacagaaacttggaaacttcocaaactgtgtg
tcgttgagagattggctgacacctcaaatttgctggttagaatgtccacaatgaacattgagagcagttgcccaactcgtaatagt
actgtagatgcatgacacagagattcaggttactgagcctgaccacctaaagtatttttagcagcttacactgagctcgacct
gacaaagctccggtgagatgtagtgaacacctgatatatttgtgctcagtaaggaagtaagatcaagctttatgttgggttatt
gtctccctttgttttggtttttttctctcctgtaatttggattcggatgctgaaagaggtctgaaagaggtttccactcaactgtgtg
agtacgggccattaaagattgaaatccaatttgaagatgagatgtagagagaaagtagtgagagattggcccaatgacctcgac
caagaatgaccagaattgagtagcttccatccccattctctttgttatcgtaaagtagatagaattttaccgggctcgtaacatatt
ctcgtcattggctggatagcactttgttataaccatggagtttccgtccagcatgcttgctgtagcatgcaattctcactcaactgtgtg
atctcctggtaactcagctcggctctgcttctcttaatttgaatgaagcagagggcccagccttagcctcaactgtgtgaa
aatcggctgccattgaggtcttataatctctgacttctcgtgctgctttgtggtaactcaggatctgaaactgtgtgttactttt
gcaactaaagctccctcgtctgaaatcacactcttgatgaaaagcctgaacggagaattcctcactactaaactattgatcgaatt
tggtcattataggcctatttgaagaaagtgagctagattcattggttgcgattatgtaatgctggcgagaacctcattgataaaat
tgaactcgatcccgaatgtgttttctgtaaatcatatga 3'

>CT14 (09 éxons e 09 íntrons)

5' taggattgagttggccaaaacttcaaccccttagattgagcaccacatggcgagagtgaaatgctgcttgggtggcagcatgagagct
ctccaaagtttttccatggttggcgaatgtaccttcagtttctcagatggtttgactgtagatgagaagggacattcctatttta
ttcttttggatgaatgagaattgttctacactcgtattgttggatgtagcacaaccgataatgtagttgatcatgtagaaagcattgtt
atttatcagatggttagattttattgcttctcgttaattgttggatgctgtgggaatagtttgggttagtttagaagttttggt
agcttcataagaaatgattttgcatatgacaacatcattataataatattttagttaggatttttggaggatgatggaaacgtagtg
ccttttggattctgaaaccttacttgaatatctcattgatagatgggtcgtgcaaatagggatgcaataatgataggcatttaatt
atctgtgcatgataatcagtagtgtt**ag**AACGATGGCTCAGCAACTAAGAATTGAATTTGCTGATGACCGCTCTCAATTTGCAAT
ATTGAATATGAAGATTCATATGTGCTGGAAGCAAAAGCACTAGAAGCTTTTGGCAATGTTATGGAAGATCTAGAGGATGCTGTG
CGCCACCTTGGCTAGTAGAGGACTGTCTATCGGCATTGGAGCGAAAATAGACATTGGCTATGCTAGTCTAGCTCAGAAAGGGCCT
TGAGCTTAAGCGCAACCGTATTGATCTGTTCAATATGTGGAAAGCAAAGGATCAAGTCCCTTATGCAAGGATCAAGAGTTAACCA
ACACAAGAGAAGAAACCTCCTCAAGTGAAGAAATGTAAGGTAAGGTTTACAATTTGACAAAGGCTCAATGCTTGGCAGC
TGTGTGAGCCAAAGAAGTAAACACCTCTTATACATTTGAGAACTTTGATTATGTGTCAAGTTCTGAATATTGGCTGAATCCAATC
CTTTGTGACTGTAGATTACAGGAAAGACAGACCATT**g**taaggaaggatagaagatccttctgtaaatcccgataatccaactagaact
atgtggatttattgtgggaataaacttcagagctgaagttaaagggatttggcacctgccaattgaaactgctggtgtgagcaccct
ttctacatgagttccatattgctccagagattcgtcgaatattatgctcgtcgtggtggtttagcttgggttagttagttagttagct
tccataataatggttagattttggatgatacactcactatgattttgggtcacttctggaagggtttattgtactaaatgttga
ctgttgtgagatgcaaatggttatcccttctggttcaactctctactttatagataatgagtaggaatgaggtgagatgctgactgg
tccattttggcacaacatagaatagactagcaagaggggcttgttgggcaatttgaataagtagattttgcatatgtagcaat
gctagaagggaaacagcaaggaacatttggaagggcaaaccttaacttccctcgtgcaatcctcagaaactctgatttggatgg
ttcaatgaatatgagaggaaggaatggggtggttctcactttataatgattctactgtattcggatattgcttggatttc
tcataaatctgaagcataaggtgtttcaaaa**ag**TTTATGAACTGGTAGAGAATCAATTAGACAAGAAAATAAAAGTATGAGATCCG
ATCGAGGTCAGAAATTTATCTGATGAGTTCAGAAAATTATGTGATGAAAAGGAATAGAAGACAGTTGCTATTCCATATACCTC
TCAACAAAATGGTGTTCGGGAGAGAAGAACAGAACCTTACTGGAAT**G**gttaggtccatgtagggcctcaactcacttatcact
tttttggagtgatgtgtggttaactattgcctcactaacttaaccgggtgcctcacaattagttattccactccataaagttatgg
ataggtgaaaaaccgacttaagtttctaaagtcatggggtgtgctgcctataatgaaactccctccacaatttgggaaataggct
ccaaatgaaagaagagtttttaatgagatactgcaaacctgaaatggatgtggttccatagggtgagcaggaaagtgggaaataa
ctgaatttgaactcagggatgacatcctttagtggatgaatttccaaaggggcaaaataggagaagattttccctatttcaaa
ccttggatcaagataatgatattagtgagggtcgtccaagtgaggtaaatagagaagtagaattgaaatcactcattctcaat
acaaccacatgagatgagatgagctattctaatctaagttgggagcataatggggaatgagtgctgcaaggtacattcctccata
tagcgaacaagtggcaagatccccatcagcgtttgacattgaaggggggaaactttataatgtctccacagataaaggtgag
ccaagaaattaaatgaggtcattaaatgcccctagtaagggaaaattggatttagcaatgaaagggaaatgagtagcactcaatgaaacaa
accaagctggggaactggtgactctctaaa**ag**GGACGACAGACTATTGGGAACAAGTGGTTCTCAAAATAAAGCGAAAGGCTGATG
GCTCGATAGAAAGGCATAAGGCTCGCCTTATGGCGAAGGGGATAATCAACAGGAAAGATTTGATATGAAATGAAAGACTGCTTTC
AGTGAGATTTACCTTGAATTGCTGATTATTTCTGGCAATAATGGTTAGTATGGATTTGAGTTACATTTATGAAATGAAAGACTGCTTTC
TCAATTTGAAATTTAGAGAAGAAACATATATGGAACCACTGCTGTTTTCATAGTGAAGGGCAAGAAAGAA**ag**gtatgctgacttt
gggggtcgatataatagccctaaagcagctcgtcaagacaatggatatacgttttcaatgcccataatggcatatgagtttacaatgag
tgatgaggtcattgtgatataatcaaaagatccaaggatcaatttgtgatcattcatatattgtgtagatatactgataccaga
agcaatattggagttgggttaactgctcaagagatagttgtcttccaactttgagatgaatgatatgggatgtaaacccataaacact
cctattgcaaaaagggaaaggtgagccatgagctgtccaaagctccacaagagaaggttcaactaaatcctcttaactgata
gtgctgttgggagcttagatgtagctatgagtgtagcaaacaccagacatagcttggcagttggatggtgagtagatcaactccaa

tccaggctccagcacaatgaaaagccggttaaagaataactaaggatctaaaggggactgctgattatacgtgagttattatggaaag
gatctgcgactctaaggctattctgatgatgattggggaggagatttaaataaaaaggaaatctacctctaggtttgtctctaccga
ataatggcaccatagtgtggagcagtaagaaacaaaagtgatagccttgttcacgatataagttgagttcgtggcattatcagcagc
agtaataagaagggtttggcttaagagattttttggatcatttaggtgttattgaaagtgtgcagattcattgttggtaactaatg
atagccaagcagcaaaaagtgtaaccaaaagatccaaaatatcattggcaagaccaaaacatagataccaagtataatttcgtcaaa
aatggttgacgaaaagatggtgaacttataatacatactacgcatagaatggtagcagatcctatgacaaagccaatacctagagat
gtgtttgtggctatatgaaatctctagagattgctgtagggctgatgactgaatattgtaattccctaaagttccacaatttgatg
aatttgggtttttcatcattaatgacctatgaatctttgtttgatctttatgcatcttaagtctcagtgcatcttaagttaggaaa
gtatgtcggacagatagagattagcccactcacacgggtaatacactctattactgtgtgataaataagagatgagactgtttttaa
gcttattatctaggtgataatcgagagctcaagcttaaaacagctatgtcgccttaactgggtgttaagatgaggacataaacttac
tatgtccgagagtaaaaatccccacataatcttactttatctgtctatgatgcccattgtgagttctgataagtttagtgaattgagtag
atacgtgaactcaagtcagacattgggtatgtctgaactatcatctgtacggattgaaagggtgtagacatacgcctccatgggaaagg
agtgtgataccgtaatacgtatctcactacgtatgcatgagacgaccaataagagtgccaaaagtttgatctcaacttttttttt
atgtcgtgtgagactcttgaggaagaggttctcctcaatttttttagtccccccatgactcttacttattttcaagatttctatttagtg
tttgctatcttaggattgttgcctaatggtcatgagtttgattcgatcctaatggggcattttctagaaaagcaagctgaactaaaattgag
agttgaaaggaagaggaagatgattttggagaatcacattgtagttttgtattcaccgggttaaccaattatgactgtttttgtgat
aatcataattgtgaatacaatatatcattgtttgaaagagatcaatagagataaaaatgtgatcaatcgatctaggggtactgcatact
aaatctactcggagtgatgctcattatgagctgctatatatctcctaac**ag**ATGTATAGCAACGAGCTCTCAAAGTATGCTGGTCCG
ACGGATTATTCACT**Ggt**atgaaatgtgaaagagaggaagaaatctgttcggcccacatgttaagaaaagctatgatggattt
ttttctgatgaggtgagtagggctgtccgccatgttggttttgtaccaccatgtgcgagtgaggagatgatggatttatgagatg
ggcctgagttggcccgtccaccaagttgggcccacaaaagccggcccacgggaatagggcccgtggatgggggaaggtataagggag
gagagagagagagagagagagagaagcaccctttggcacaatcaacgtacgctctctcccgcgtgtttctctcccga**ag**AACAG
TAGGCATACGGCTTCCAATTTCTTCCCCTCAAGGCTTCATCGACCGGATTTCTCTTCCCTCCATTGACGCGTGAATTAATCTGTGGCTA
ACAAG**Ggt**acgctcgactccgtggtgtgctttacgatcgggtgatacgtgtttttcccgagcttcgtcttccgcattgatt**ag**GGGTTC
GATTCGGTGTGCAATCCGGTATTTCCGGGATCAGTCATTTCCCAACAAAACCTAAACGTCAGATCTTCTTCTTCTGCTCT**Ggt**gac
catctcctcggatataagagtggttttcaogttttcttttccatttctgttctttgtt**ag**GGTACTTGTTTTCCCTCGACATAGTGG
CCACCATGGAGAGGCACAACAATGAGATGGCCGACTGTGTATGGCCAGGGCTAAGAAGATCTGCCATGATTTTGGTGAG**gt**gagttc
ttgttgaggcataaatttccaagtttcgataagtgaaaatgtctagggatcaatgggattgagtgacccgggtgtggttgcatttgtt
ggaggat**ag**GTCGCGGTGGAGACGAGAGTGGAGAATGGGGATGCAAGAGATGTCATCTGCAGCGTGGCACAGAAAGTTGGTCCCTGATG
TTCTGGTGATGGGAACCATGGTTACGGTCTTATAAAAAG**gt**acttcaattactcattttgtgtaattactcaaaactctgtacatcg
gctctaataatcaatcacctcagaggggaagacatgtattatgtagccttaacttgagataatgttaacatgaaagatgggtaataca
attttatgatgaacggggataggtcgtacattaagaatttagaacctaccatgtctattttccataacaaggccaattt aaagtac
cataactattgttcgaaaaaatgtgctggaactatttgggaatattgataaaggatagatggatctcctgaacttttgagttcaa
aatccatcttagtctaactatattttttctcacttatcccgggtgaaggttgccttacctacatgtggggcatttgaagttctctc
agactgatccaatattcaatgatttattcgaattagtagaatgagggatttactcttaattgaaaacctccttcgctttctatgggt
agattgtggcatctgtatcgatacttaattgtgatttgggtgaaatgtcac**ag**GGCCTTTCTGGGAGCGTGAGCTATCACTGCGCCAA
GAACGTCAAAGTGTCTGTGCTGGTAGTGAAGAGGCCAAAACCAACAGGAACGACTGAaactcgatctatacttctcgtgttgggtcc
ttgtcctaggttgtcttctacaggattatcgagttttatgtgtatataatcctacagtggtgcatattctttgttttctgccttgtaa
agggattggatcttaataatagaatttgattgtcaaaaaggacaggggtgcttgacctcatcattaaagcagcatacaataaaaataaaat
aaaaatgtgaatgagcttatcgaagtttccatcagctgtcgacagagactggcacaatcaaaagagagaagagcagatagagaatgatg
gcattatcttaaacatatagcagtagatccagactctcaacaagtgcgattgatctgaaaatggacgtttgccttcgttctgaaaa
ctcgagtcagtcacaaatggatagggatgcaatcagtcgctcaaaagctctgggatcagccttcgtacagttctctccctctctaaagaa
atgtcctacatctatattgtaagatagaaaaataaaaattatgaaggcccaagcttgagatctctatagctcacctaaat 3'

MOTIVO TIR-1

176 YEVLNFRGADTR 188

```

          * : : * * . . * . * *
AAQ93075  --NYDLFLSFRGEDTRN---
CAC95124  --AYDVFLSFRGEDTRK---
AAU04763  --RYDVFLSHRAKDHRANND
P_174037  RLKYDVFLSFRGADTRDN--
AAG13418  -LKYDVFLSFRGADTRDN--
→ CT13    --SYEVLNFRGADTRNG--
CAC35330  --EYEVFLSFRGLDVRKT--
ruler    |.....180.....190

```



MOTIVO TIR-3

276 ILPIFLDVEPVDVK 289

```

          : * * : *   : : *   . * :   .
AAQ93075  |GRHVLPIFYHVDP SHVRKQDGI
CAC95124  |GQIVLPIFYDIDPSDVRKQNGI
AAU04763  |DQLVLPIFYKVDPGDVRKQEGI
P_174037  |DRRILPIFYHVDP SHVRKQSDI
AAG13418  |DRRILPIFYHVDP SHVRKQSDI
→ CT13    --EILPIFLDVEPVDVKLKTN
CAC35330  |QHIILPVFLFVDPRDVRHTESI
ruler    |.....280.....290.

```



Anexo 4 - Sequências de aminoácidos do agrupamento contíguo CT13 representando os motivos TIR-1, TIR-2, TIR-3, TIR-4, P-loop, Quinase2a, Quinase3a, HD (hidrofóbico), motivo-1 (LRR), motivo-2 (LRR) e motivo-3 (LRR). As sequências foram alinhadas utilizando o programa *ClustalX* (1.83).

MOTIVO TIR-2

225 IGAELLQAIENSKIYIPIFSANYASSCWCLRELAYMAEC 263

```

:      *          .: *      : ::: * . * * : * :      :
AAQ93075 DDLNRGEEIKEELFRAIEGSRISIIIVFSKRYADSSWCLDELVKIMECR--
CAC95124 DELPRGEEISDHFLRAIQESKISIAVFSKGYASSRWCLNELVEILKCKK-
AAU04763 EDEEDGGKPLTEKMKAVDESRSIVVFSSENYG-SWVCMKEIRKIRMCQ--
P_174037 EGMERGDEISSSLKAGMEDSAASVIVISRNYSGSRWCLDELAMLCKMK--
AAG13418 EGMERGDEISSSLKAGMEDSAASVIVISRNYSGSRWCLDELAMLCKMK--
→ CT13   ESLHVGEEIGAELLQAIENSKIYIPIFSANYASSCWCLRELAYMAECTLK
CAC35330 EELEKGETIGPSLIRAITESKIYIPIILTQNYASSKWCLQELAKMVECWKS
ruler   ...220.....230.....240.....250.....260....

```



63

MOTIVO TIR-4

325 WQKALIAVG 333

. . . * : : : : . . . : *

AAQ93075 ERVKQWKKALTEAANLSGHDLRI
CAC95124 KLVKEWRKALEEAGNLSGWNLND
AAU04763 EEVKKWRKSMNKVGNLSGWHLQD
P_174037 EKVQEWREALTLVGNLAGYVCDK
AAG13418 EKVQEWREALTLVGNLAGYVCDK
→ CT13 SEAESWQKALIAVGKIKGWNLKK
CAC35330 ETVLEWKEALQEVGRMKGYHVTE
ruler 320.....330.....340



MOTIVO P-LOOP

408 FLVIHGMGGVGKTTLTkVVFN 428

. . . : : ** * : * * * * * : : . : .
AAQ93075 SS-GGSNVIMVGIWGMGGLGKTTAAKAIYNQIHH
CAC95124 SA-ATDDVRIVGIHGMPGIGKTTIAQAVFNQLCY
AAU04763 GI-GLDDIRFVGIWGMGGIGKTTLARIYRSVSH
P_174037 DTSSSGVQVLGLYGMGGIGKTTLAKAFYNKIVG
AAG13418 DTSSSGVQVLGLYGMGGIGKTTLAKAFYNKIVG
→ CT13 DMESDD-VKFLVIHGMGGVGKTTLTkVVFNRFSS
CAC35330 NLDSSTSEKIIGIHGMGGLGKTTLAKAVYDKVST
ruler .400.....410.....420.....430



MOTIVO QUINASE 2A

491 KVLIVLDDVD 500

: : : * : : .
AAQ93075 RRVLVIMDNIDEV
CAC95124 KRVLVVADDVAHL
AAU04763 IKALIILDDVDHL
P_174037 KKIIIVLDDVDHI
AAG13418 KKIIIVLDDVDHI
→ CT13 KVLIVLDDVDEN
CAC35330 HKLLIVLDDVDEK
ruler 490.....500



Anexo 4 - Continuação

MOTIVO QUINASE 3A

518 SSRIITTRY 527

```
          : . . : ***      *
AAQ93075 SRIIITTRDEHLLK
CAC95124 SRVIITTRDSNLLR
AAU04763 SRIIVTTRNEHLLV
P_174037 TLIVITTRDSEILS
AAG13418 TLIVITTRDSEILS
→ CT13   SRIIITTRYKSVLM
CAC35330 SRFLITTRDARGLE
ruler   .520.....530
```



MOTIVO HIDROFÓBICO

591 GKLPALAEVTGSSLSGR 607

```
          . ***  : : *  * :
AAQ93075 CGGLPLALEVLGS
CAC95124 CGGLPLALEVIGA
AAU04763 SGDLPLAIEVLGS
P_174037 SGLLPLAVEVFGS
AAG13418 SGLLPLAVEVFGS
→ CT13   LGKLPLAEVTGS
CAC35330 AAGLPLYIKVIGS
ruler   590.....600
```



MOTIVO 1 (LRR)

930 LM RTPDLSMLVSLERLILEGCR 950

```
          * :   :           .       ** * : *
AAQ93075 -----NLKTLDLSESRSLQKSPDFSQVPNLEELILYNCK-
CAC95124 -----MLQSPKFLQY-----VIYIYILEKLNKLGCS-
AAU04763 -----RLKTVNLSDSQFISKTPDFSGVPNLERLILSGCV-
P_174037 -----DENLKVVILRGCHSLEAIPDLSNHEALEKLVFEQCT-
AAG13418 ILTVKTFLCFFQVDENLKVVILRGCHSLEAIPDLSNHEALEKLVFEQCT-
→ CT13 -----KLKVLDSLNCNNLM RTPDLSMLVSLERLILEGCR-
CAC35330 -----KLKAVTLERC FNLNKVPDFSHCRDLEWLD FDECRN
ruler .....910.....920.....930.....940.....950
```



MOTIVO 2 (LRR)

1010 LDALTEILMPHSSYAFKL 1027

 . . : : :
AAQ93075 DFYKSKSVEALLNGCLILRELHEDIGE
CAC95124 SIGNVKSLETNLISGCSQLEKLPESMGL
AAU04763 SIS-LESLIVLSLSNCSLKNFPNIVGN
P_174037 DVSGLKLEKLFSLGCSDLVLPENIGA
AAG13418 DVSGLKLEKLFSLGCSDLVLPENIGA
→ CT13 EIGCLDALTEILMPHSSYAFKLPETIGN
CAC35330 GISKLSLEFLTLALN-DPYKSDFTEMI
ruler1010.....1020.....1030



MOTIVO 3 (LRR)

1037 LLTLDVSHSWIGMLPDSIGNLKSLTTL 1064

 .* * . : : .* : *.
AAQ93075 EDIGEMISLRTLEAEYTDIREVPPS-IVRLKNLTRLSSLSSV
CAC95124 ESMGDMESELIELLADGIENEQFLSS-IGQLKHVRRLSLRGY
AAU04763 NIVGNMKNLTELHLDGTSIQELHPS-IGHLTGLVLLNLENC
P_174037 ENIGAMTSLKELLLDGTAIKNLPES-INRLQNLEILSLRGC
AAG13418 ENIGAMTSLKELLLDGTAIKNLPES-INRLQNLEILSLRGC
→ CT13 ETIGNSKSLTLDVSHSWIGMLPDS-IGNLKSLTLD--VS
CAC35330 FTEMLPTSLMSLLISNDTQKSCPDTSENLQRLPNLSNLIN
ruler .1030.....1040.....1050.....1060.....

