

Universidade de Brasília  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Fitopatologia

**CARACTERIZAÇÃO DE GENES ANÁLOGOS DE RESISTÊNCIA (RGAS) EM  
*Elaeis guineensis* e *Elaeis oleifera* CONTRASTANTES EM RESISTÊNCIA A *Fusarium  
oxysporum* f.sp. *elaeidis***

**Flávio Romero Palma**

Orientador: Prof. Dr. Robert Neil Gerard Miller

**CARACTERIZAÇÃO DE GENES ANÁLOGOS DE RESISTÊNCIA (RGAS) EM  
*Elaeis guineensis* e *Elaeis oleifera* CONTRASTANTES EM RESISTÊNCIA A *Fusarium  
oxysporum* f.sp. *elaeidis***

Dissertação apresentada ao Programa de Pós-Graduação *stricto sensu* em Fitopatologia da Universidade de Brasília, como requisito para a obtenção do Título de Mestre em Biologia, Área de Concentração Fitopatologia.

**Brasília 2011**

Universidade de Brasília  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Fitopatologia

Dissertação de autoria de Flávio Romero Palma, intitulada “Caracterização de genes análogos de resistência (RGAs) em *Elaeis guineensis* e *Elaeis oleifera* contrastantes em resistência a *Fusarium oxysporum* f.sp. *elaedis*”, apresentada como requisito parcial para a obtenção do grau de Mestre em Biologia, Área de Concentração Fitopatologia, defendida e aprovada em 12 de Dezembro de 2011, pela banca examinadora abaixo assinada:

---

Prof. Dr. Robert Neil Gerard Miller  
Orientador

---

Prof. Dr. Cleber Furlanetto  
Examinador Interno

---

Prof. Dr. Lúcio Flávio de Alencar Figueiredo  
Examinador Externo

**Brasília 2011**

*“Que a tua vida não seja uma vida estéril. Sé útil. Deixa rastro.”*

Josemaria Escrivá de Balaguer

Dedico este trabalho à família que Deus escolheu para mim.

## AGRADECIMENTOS

Embora apenas o meu nome apareça na capa desta dissertação, um grande número de pessoas contribuíram para torná-la possível. A elas eu devo minha gratidão.

Agradeço ao professor Robert Neil Gerard Miller, por sua orientação durante a minha pesquisa e estudo na Universidade de Brasília. Seu entusiasmo, sua acessibilidade e disposição, além, é claro, do seu exemplo - um pesquisador de absoluta competência - foram fundamentais para o meu progresso. Fui incrivelmente afortunado por ter um orientador como o Robert, que me deu a liberdade para pesquisar por conta própria e, ao mesmo tempo, a orientação para tornar este trabalho o melhor possível.

Ao Dr. Georgios Pappas da Embrapa Cenargen pela ajuda com as análises de bioinformática.

À Universidade Católica de Brasília pelo apoio material e pela cessão das suas instalações.

Aos professores do Departamento de Fitopatologia da UnB, pelos ensinamentos e disponibilidade.

À Cristiane Teixeira pela grande ajuda e pelo treinamento. Seu acompanhamento e amizade foram fundamentais para o progresso no trabalho laboratorial e me motivaram a persistir.

Aos amigos Flávia Leonel, Marco Ninomia e Gláucia Midorikawa pela ajuda e disponibilidade.

A Renato Vieira e Rafael Stanziona. A paciência e carinho destes grandes amigos foram a maior motivação que eu poderia receber.

A Raphael Rezende, Sérgio Consolmagno, Guilherme Aminger, Homero Piccolo, Ives Gandra, Stefano Miglorelli, Décio Piva, Henrique Marconi, Roberto Minadeo, João Pedro e André Guerreiro pelo carinho e cuidados.

A Henrique Menegaz, Bruno Azevedo, André Jaborandy, João Felipe Belo, Felipe Domingos, Pedro Dias, Carlos Eduardo, Gustavo Furia, Nélio Pova e Fabiano Garcia que me ajudaram tanto e de tantas formas.

Aos amigos da Biologia Ana Carolina, André Cunha, Franklin Júnior, Hugo Cruz, João Paulo, Marcela Maly e Nayara Almeida pelo apoio e por tornarem estes dois últimos semestres mais divertidos e interessantes.

À minha família de sangue, Telma, Josué, Bruno, Nicole e Tharcílio, pela educação que deram e valores que me transmitiram.

Agradeço a Deus pela minha vida e pela vida de todas as pessoas aqui citadas. A minha Mãe pelo zelo e alento e a “Mi Padre” por me ensinar a importância e o valor de um trabalho que se faz com amor.

## RESUMO

O dendê, *Elaeis guineensis* Jacq., espécie de origem africana, e o caiaué, *Elaeis oleifera* (Kunth) Cortés, estão entre culturas oleaginosas mais importantes do mundo. No entanto, o progresso no melhoramento genético destas culturas perenes tem sido relativamente lento em comparação com outras culturas oleaginosas anuais. O melhoramento genético visando a resistência a doenças é ainda limitado, sendo a identificação de novas fontes de resistência às principais doenças uma prioridade. Os objetivos deste estudo foram isolar RGAs NBS-LRR e explorar sua diversidade em acessos de *E. guineensis* e *E. oleifera* utilizados em programas de melhoramento locais, testando plantas que representam o centro de origem de cada espécie e que contrastam em resistência à murcha vascular. Esta doença, causada pelo fungo *Fusarium oxysporum* Schlechtend.:Fr. f.sp. *elaeidis* Toovey, é considerada uma das mais destrutivas em dendê. Usando *primers* degenerados desenhados para amplificar RGAs NBS-LRR em genomas de monocotiledôneas, TIR e não-TIR, relatamos a primeira análise em grande escala de RGAs NBS-LRR no gênero *Elaeis*. De um total de 486 seqüências de DNA de alta qualidade, que foram geradas a partir de plasmídeos recombinantes contendo insertos de produtos de PCR, 365 apresentaram similaridade significativa para genes de resistência NBS-LRR e seqüências de RGAs conhecidos. A montagem de grupos de seqüências de alta qualidade gerou um total de 60 seqüências contíguas não redundantes, entre *clusters* e *singletons*. Foram identificados 54 RGAs NBS-LRR não-TIR distintos, os quais apresentaram o domínio NB-ARC em teste feito utilizando o Genewise, e codificavam proteínas com os motivos de aminoácidos esperados para o domínio NBS. Um total de 39 seqüências continham ORFs contíguas. Foram determinadas as similaridades entre as seqüências e considerável polimorfismo foi observado. Em análises feitas por Biolayout, foi verificado o agrupamento de diversas seqüências em dois *clusters* principais sendo que um deles reuniu apenas *contigs* que possuíam seqüências provenientes de tratamentos com acessos resistentes. A construção de um dendrograma pelo método Neighbor Joinig não permitiu a verificação de tendência nos agrupamentos. Os dados gerados neste estudo enriquecem os recursos genômicos para o gênero *Elaeis* e podem ser explorados para a descoberta de genes candidatos de resistência e para o desenvolvimento de marcadores moleculares. O mapeamento genético de tais marcadores para QTLs pode abrir oportunidades para a seleção assistida por marcadores, com introgressão de genes através do melhoramento convencional ou por técnicas de transformação genética.



**Palavras-chave:** dendê, *Elaeis guineensis* Jacq., *Elaeis oleifera* (Kunth) Cortés, *Fusarium oxysporum* Schlechtend.:Fr. f.sp. *elaedis* Toovey, murcha vascular de *Fusarium*, seqüenciamento de DNA, análogos de gene de resistência, NBS-LRR, genes candidatos.

## ABSTRACT

The African Oil Palm, *Elaeis guineensis* Jacq., and the American Oil Palm, *Elaeis oleifera* (Kunth) Cortés, are amongst the world's most important oil-bearing crops. Progress in genetic improvement in these perennial crops has, however, been relatively slow, in comparison with other annual oil crops. Genetic improvement in relation to disease resistance has been limited, with identification of new sources of resistance to major disease regarded as a priority. The objectives of this study were to isolate NBS-LRR RGAs and explore their diversity across *E. guineensis* and *E. oleifera* accessions used in local breeding programs, together with material representing the centre of origin of each species and material contrasting in resistance to oil palm vascular wilt. This disease, caused by the fungal pathogen *Fusarium oxysporum* Schlechtend.:Fr. f.sp. *elaedis* Toovey, is today considered one of the most destructive in oil palm. Using degenerate primers for targeted NBS-LRR RGA amplification across monocotyledon genomes, together with universal TIR and non-TIR NBS-targeting primers, we report the first large scale analysis of NBS-LRR RGAs in the genus *Elaeis*. From a total of 486 high quality DNA sequences which were generated from PCR product insert-containing recombinant plasmids, 365 showed significant similarity to known NBS-LRR R-genes and RGA sequences. Clustering assembly of high quality sequence data generated a total of 60 non-redundant contiguous sequences and singletons. Fifty four distinct non-TIR-NBS-LRR RGAs were identified, all of which tested positive for the expected NB-ARC domain using Genewise, and encoded proteins with the expected NBS domain amino acid motifs. A total of 39 sequences contained contiguous ORFs. Sequence similarities between sequences were determined, with considerable polymorphism observed. Biolayout analysis showed the clustering of several sequences in two main clusters, with one of these grouping only contigs derived from treatments with resistant accessions. The construction of a phylogenetic tree using Neighbor Joining method revealed considerable diversity in the NBS-LRR RGAs obtained, but did not show any obvious trend in the clusters, with specific clades grouping sequences from various treatments. The data generated in this study enrich the genomic resources for the genus *Elaeis*, which can be exploited for candidate resistance gene discovery and molecular marker development. Genetic mapping of such markers to quantitative trait loci could open up opportunities for marker assisted selection, with gene introgression through conventional breeding or genetic modification approaches.

**Key-words:** Oil palm, *Elaeis guineensis* Jacq., *Elaeis oleifera* (Kunth) Cortés, *Fusarium oxysporum* Schlechtend.:Fr. f.sp. *elaedis* Toovey, Fusarium vascular wilt, DNA sequencing, Resistance Gene Analogs, NBS-LRR, candidate genes.

## LISTA DE FIGURAS

- Figura 1:** Morfologia de *Elaeis guineensis* Jacq. e *Elaeis oleifera* (Kunth) Cortés
- Figura 2:** Formas básicas dos frutos de *Elaeis guineensis*
- Figura 3:** A produção em MT e Int\$ dos principais produtores de dendê em 2009
- Figura 4:** As duas síndromes causadas por *Fusarium oxysporum* f.sp. *elaeidis*
- Figura 5:** As seis diferentes famílias de genes de resistência
- Figura 6:** Alguns motivos conservados dentro do domínio NBS
- Figura 7:** Motivos conservados em dominios NBS e LRR alvos de primers degenerados
- Figura 8:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2 µl dos DNAs genômicos totais de *E. oleifera* e *E. guineensis*
- Figura 9:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2 µl dos produtos inespecíficos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos primers P1A e P3A
- Figura 10:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2 µl dos produtos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos primers P1B e RNBS-D
- Figura 11:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2 µl dos produtos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos primers 1F e P3B
- Figura 12:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2 µl dos produtos inespecíficos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos primers 3F2 e 13R1
- Figura 13:** Motivos conservados e seqüência consenso de RGAs NBS-LRR de *E. guineensis* e *E. oleifera*
- Figura 14:** Eficiência das combinações de primers P1B/RNBS-D e 3F2/13R1 na amplificação de RGAs
- Figura 15:** Imagem de BioLayout 2D mostrando as relações de similaridade derivadas de Blastx ( $E$ -value  $1E^{-05}$ ) entre RGAs da família NBS-LRR de *E. guineensis* e *E. oleifera* e genes de resistência ou RGAs de *A. thaliana*
- Figura 16:** Dendrograma gerado por Neighborn joining a partir das seqüências de aminoácidos entre os motivos *kinase-2* e RNBS-D dos RGAs NBS-LRR

## LISTA DE TABELAS

**Tabela 1:** *Ranking* dos 20 principais países produtores de dendê

**Tabela 2:** Valores de caracteres agronômicos de *E. guineensis*, *E. oleifera* e híbrido interespecífico

**Tabela 3:** Espécie, acesso e resistência do material vegetal utilizado

**Tabela 4:** Sequências dos *primers* degenerados usados no isolamento de RGAs em *Elaeis* e seus motivos alvo

**Tabela 5:** Reagentes e concentrações utilizados para a amplificação de RGAs NBS-LRR por PCR

**Tabela 6:** Nomes dos tratamentos baseados em espécie, acesso e combinação de *primers* degenerados

**Tabela 7:** Quantificação em ng/μl de DNA total dos acessos de *E. oleifera* e *E. guineensis* com os respectivos valores de absorvância e razões medidos pelo Nanodrop ND-1000

**Tabela 8:** Lista de *primers* degenerados testados, etapa concluída para cada combinação e seus respectivos motivos de uso ou não uso

**Tabela 9:** Estatísticas dos *amplicons* obtidos para todos os tratamentos usando *primers* degenerados direcionados a NBS-LRR RGAs e genes de resistência

**Tabela 10:** Motivos conservados encontrados nos diferentes acessos de *Elaeis* com duas combinações de *primers*

**Tabela 11:** Resultado das análises por Blastx para identificação de domínios de interesse (NBS-LRR) e similaridades entre os RGAs e sequências de nucleotídeos não redundantes (nr) do Genbank

## LISTA DE ABREVIACÕES

AF: Amarelecimento fatal

ATP : Adenosina tri-fosfato

Avr: Avirulência

BAC: *Bacterial Artificial Chromosome* (cromossomo artificial de bactéria)

CC: *Coiled coil* (cauda espiralada)

DNA: *Deoxyribonucleic acid* (Ácido desoxirribonucléico)

e-PCR: *Electronic PCR* (PCR eletrônico)

ESTs: *Expressed Sequence Tags* (Sequências expressas)

ETI: *Effector-triggered immunity* (Imunidade ativa por efetores)

ETS: *Effector-triggered susceptibility* (Suscetibilidade ativa por efetores)

FAO: *Food and agriculture organization*

Gene Avr: *Avirulence gene* (gene de avirulência)

GTP: Guanosina tri-fosfato

HR: Reação de hipersensibilidade

IL: Interleucina

LRR: *Leucine rich repeat*

MAMP: *Microbe-associated molecular pattern* (Padrão molecular associado a microorganismos)

MAP: *Mitogen-associated protein* (Proteína associada a mitógeno)

Mb: Milhões de pares de base

MT: milhões de toneladas

NBS: *Nucleotide binding site*

ORF: *Open Reading Frame* (Quadro aberto de leitura)

PAMP: *Pathogen-associated molecular pattern* (Padrão molecular associado ao patógeno)

Pb: Pares de base

PTI: *PAMP-triggered immunity* (Imunidade ativa por PAMP)

PCR: *Polymerase Chain Reaction* (Reação em cadeia da polimerase)

QTLs: *Quantitative trait loci* (Loci de caracteres quantitativos)

RAPD: *Random amplification of polymorphic DNA* (DNA polimórfico amplificado randomicamente)

RFLP: *Restriction fragment length polymorphism* (Polimorfismo do tamanho do fragmento de restrição)

RGA: *Resistance Gene Analog* (Gene análogo de resistência)

RLK: *Receptor like Kinase*

RLP: *Receptor Like Protein*

SAR: *Systemic Acquired Resistance* (Resistência Sistêmica Adquirida)

STK: Serina/Treonina quinase

SISGEN: Sistema Genoma

TF: Fator de transcrição

TIR: *Toll interleucine receptor* (Receptor Toll interleucina)

YAC: *Yeast artificial chromosome* (Cromossomo artificial de levedura)

## SUMÁRIO

<b>1. Introdução</b>	<b>19</b>
1.1 O Gênero <i>Elaeis</i> : classificação botânica, morfologia e diversidade.....	19
1.1.1 Classificação botânica.....	19
1.1.2 Morfologia de <i>Elaeis</i> .....	19
1.2 O uso e a produção de <i>Elaeis</i> .....	21
1.2.1 A produção de <i>Elaeis</i> .....	21
1.2.2 O uso de <i>Elaeis</i> .....	23
1.3 Doenças em <i>Elaeis</i> .....	24
1.3.1 Antracnose .....	24
1.3.2 Mancha foliar de <i>Cercospora</i> .....	24
1.3.3 Necrose foliar.....	25
1.3.4 Amarelecimento fatal.....	25
1.3.5 Anel vermelho.....	26
1.3.6 Marchitez sopressiva.....	26
1.3.7 Podridão do colmo.....	27
1.3.8 Murcha vascular.....	27
1.4 Resistência de plantas às doenças.....	30
1.4.1 Interação planta-patógeno.....	30
1.4.2 Tipos de patógenos.....	30
1.4.3 Mecanismos de defesa.....	31
1.4.4 Bases citológicas e histológicas da resistência.....	31
1.4.5 Bases genéticas da resistência.....	32
1.4.5.1 <i>PAMP-triggered immunity</i> : a resistência não-hospedeiro.....	32
1.4.5.2 <i>Effector-triggered immunity</i> : a resistência raça-específica.....	33
1.4.5.2.1 O modelo gene-a-gene.....	34
1.4.5.2.2 O modelo ziguezague.....	34
1.4.5.2.3 O modelo de guarda.....	34
1.4.5.2.4 O modelo <i>decoy</i> .....	35
1.5 Genes de resistência.....	35
1.5.1 Família NBS-LRR: <i>Nucleotide binding site</i> (NBS) e <i>Leucine rich repeat</i> (LRR)...	36
1.5.1.1 A organização dos genes NBS-LRR no genoma das plantas.....	37
1.5.1.2 Os motivos conservados no domínio NBS.....	38
1.5.2 Família RLP: <i>Receptor-like proteins</i> .....	38



1.5.3 Família RLK: <i>Receptor-like kinases</i> .....	39
1.5.4 Família RSTK: Serina/treonina quinase citoplasmática sem domínio LRR.....	39
1.5.5 Família com domínio <i>coiled coil</i> ancorado à membrana.....	40
1.5.6 Genes de resistência que codificam proteínas de resistência diversas.....	40
1.6 Análogos de genes de resistência (RGAs).....	40
1.6.1 Utilização de RGAs.....	41
1.6.1.1 Seleção e análise de clones BAC e mapeamento físico.....	41
1.6.1.2 Análise de expressão gênica diferencial.....	42
1.6.1.3 Marcadores moleculares e seleção assistida por marcadores.....	43
1.7 Melhoramento e recursos genômicos em <i>Elaeis</i> .....	44
<b>2. Objetivos</b> .....	<b>47</b>
2.1 Objetivo geral.....	47
2.2 Objetivos específicos.....	47
<b>3. Material e Métodos</b> .....	<b>48</b>
3.1 Seleção de material de <i>Elaeis</i> .....	48
3.2 Extração do DNA genômico de <i>Elaeis</i> .....	48
3.3 Amplificação de RGAs por PCR.....	49
3.4 Purificação e clonagem de produtos de PCR.....	52
3.5 Extração de DNA plasmidial.....	53
3.6 Seqüenciamento de clones recombinantes.....	54
3.7 Análise de bioinformática das seqüências.....	54
<b>4. Resultados</b> .....	<b>56</b>
4.1 Desenho experimental e coleta de material vegetal.....	56
4.2 Extração do DNA genômico.....	56
4.3 Amplificação de RGAs por PCR.....	57
4.4 Análise de bioinformática das seqüências.....	60
4.4.1 Processamento das seqüências.....	60
4.4.2 Análise filogenética.....	70
<b>5. Discussão</b> .....	<b>72</b>
5.1 Importância de <i>Elaeis</i> .....	72
5.2 Necessidade de melhoramento de <i>Elaeis</i> .....	72

5.3	Objetivos do isolamento de RGAs.....	73
5.4	Caracterização de RGAs.....	74
5.5	Pseudogenes.....	75
5.6	Conjunto de <i>primers</i> e a eficiência apresentada.....	75
5.7	Origem das seqüências de RGAs NBS-LRR.....	76
5.8	Motivos conservados.....	77
5.9	Análise de similaridade.....	77
5.9.1	Biolayout.....	77
5.9.2	Neighbor joining.....	78
<b>6. Conclusão</b>		<b>79</b>
<b>7. Perspectivas</b>		<b>81</b>
7.1	Construção de bibliotecas BAC.....	81
7.2	Análise de expressão gênica.....	81
7.3	Marcadores genéticos e co-segregação com QTLs.....	82
7.4	Confirmação da função <i>in planta</i> .....	83
7.5	Avanços no conhecimento de imunidade em plantas.....	83
<b>8. Referências Bibliográficas</b>		<b>85</b>
<b>9. Apêndices</b>		<b>98</b>
9.1	Apêndice A: Lista de seqüências de nucleotídios de <i>contigs</i> e <i>singletons</i> .....	98
9.2	Apêndice B: Lista de seqüências protéicas de <i>contigs</i> e <i>singletons</i> .....	110

## 1. INTRODUÇÃO

### 1.1 O gênero *Elaeis*: Classificação botânica, morfologia e diversidade

#### 1.1.1 Classificação botânica

O dendezeiro, *Elaeis guineensis* Jacq., espécie de origem africana, e o caiaué, *E. oleifera* (Kunth) Cortés, espécie de origem americana, são palmeiras oleaginosas e pertencem à família Arecaceae (Palmeaceae) (Figura 1) (Chia *et al.*, 2009).

Descrito pela primeira vez em 1570, por Matias Lopes, com o nome de *Nucula indica*, o dendezeiro só foi encontrar seu nome botânico definitivo quando introduzido na Martinica. Jacquin classificou a palma africana, em 1763, como pertencente ao gênero *Elaeis*, do grego “*Elaion*”, que significa óleo, e como nome específico *E. guineensis*, atribuindo sua origem à costa da Nova Guiné (Camarero, 1981).

Na classificação botânica atual, o gênero *Elaeis* pertence à classe Liliopsida (Monocotiledônea), ordem Arecales (Palmales), família Arecaceae (Palmae), subfamília Arcoideae, tribo Cocoseae (Cocoinaea), subtribo Elaeidinae. Embora sejam citadas três espécies nesse gênero, apenas *E. guineensis* Jacq., denominada dendezeiro ou palma africana e *E. oleifera* (Kunth) Cortés, denominada caiaué ou palma americana (Adam *et al.*, 2005; Dransfield *et al.*, 2005) são de interesse agrônômico e econômico. A espécie *E. odora* não é cultivada e pouco se sabe sobre a mesma (Chia, 2008).



Fonte: <http://www.pacsoa.org.au/palms/Elaeis>

**Figura 1:** Morfologia de *Elaeis guineensis* Jacq. (A) e *Elaeis oleifera* (Kunth) Cortés (B)

#### 1.1.2 Morfologia de *Elaeis*

*E. guineensis* e *E. oleifera* são morfologicamente similares. As duas espécies do gênero *Elaeis* são plantas alógamas e apresentam cariótipo  $2n = 32$  (Conceição e Müller, 2000). O sistema radicular, fasciculado, é constituído de 8.000 a 100.000 raízes adventícias primárias,

com 4 a 10 mm de diâmetro, que emergem, principalmente, da base do caule. Estas se localizam na faixa de 20 a 60 cm de profundidade (Ferwerda, 1975).

O caule, ereto e sem ramificações, é do tipo estipe. Alcança até 30 m de altura, porém, em condições de florestas, é raro atingir mais de 20 m. O diâmetro médio do caule é de 22 a 75 cm, com variações em função da espécie, do cultivar e do ambiente (Gonçalves, 2001).

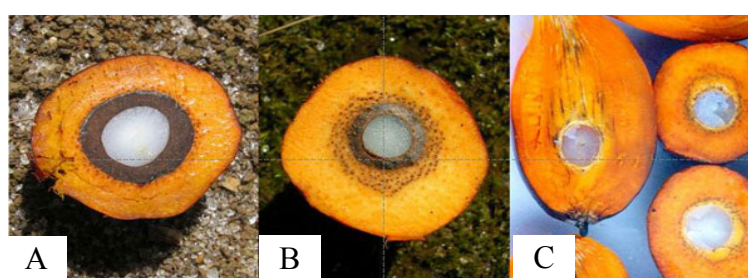
No topo da estipe encontram-se de 35 a 50 folhas pinadas que medem de 5 a 7 m e pesam de 5 a 8 kg cada. A folha é composta por um eixo central, subdividido em parte basal, que se liga ao estipe, denominada pecíolo, cujo comprimento é de aproximadamente 1,5 m. Há centenas de espinhos na parte distal, denominada ráquis, que suporta os folíolos. Repartidos dos dois lados do ráquis encontram-se de 250 a 350 folíolos, com variados ângulos de inserção. Possuem a cutícula espessa e grande quantidade de tecido lignificado (Surre e Ziller, 1969).

As plantas do gênero *Elaeis* são monóicas, ou seja, as flores masculinas e femininas são produzidas na mesma planta, mas separadas em inflorescências masculina e feminina, no caso deste gênero. Estas inflorescências são protegidas por duas brácteas fibrosas denominadas espadas e são compostas por espigas formadas na axila de cada folha. Cada esboço floral possui os dois sexos, porém um deles fica, quase sempre, rudimentar (Conceição e Müller, 2000).

O fruto é uma drupa séssil tendendo para o esférico. Tem de 2 a 5 cm de comprimento e 3 a 30 g de peso. Quando imaturo possui coloração que varia de violeta-escura a preta e a metade inferior é marfim. A metade superior apresenta cor marrom durante o amadurecimento, podendo variar com o grupo (Gonçalves, 2001). Segundo Surre e Ziller (1969), o corte do fruto permite distinguir, da superfície para o interior, um exocarpo cutinizado, liso, brilhante e fino; um mesocarpo de coloração amarela ou alaranjada, muito oleoso e com estreitas fileiras de fibras cujos feixes se tornam mais compactos à medida que estão mais próximos do centro do fruto; um endocarpo esclerificado, muito duro, de cor negra, envolvido por fibras aderentes; e um endosperma ovóide que ocupa toda cavidade do endocarpo.

Em *E. guineensis* há três tipos de formas básicas do fruto: o Dura, o Pisifera e o híbrido entre eles, Tenera, que é o mais utilizado no fornecimento de sementes selecionadas (Figura 2). Cada um destes tipos possui diferenças significativas na espessura do endocarpo, principal característica comercial do fruto. Um gene principal controla esta espessura e possibilita três fenótipos. Plantas ( $sh^+ sh^+$ ), homozigotas para este gene, apresentam o fenótipo endocarpo com espessura superior a 2 mm e com 35 a 55 % de polpa. Estes são os

frutos do tipo Dura. Plantas do tipo Pisífera (sh- sh-), também homocigotas para este gene, produzem frutos sem endocarpo e, geralmente, abortivos. Os indivíduos heterocigotos (sh+ sh-), denominados Tenera, apresentam endocarpo com menos de 2 mm e com 60 a 90 % de polpa. Do cruzamento entre Dura e Pisífera obtêm-se 100% de indivíduos heterocigotos (Tenera). O fruto do tipo Tenera tem casca de 0,5 a 4 mm de espessura, tamanho médio e com sua amêndoa bem desenvolvida. Apesar de ser menor que os frutos do tipo Dura, tem maior rendimento. Também as plantas do tipo Tenera são mais produtivas, tanto em número de cachos quanto no número de frutos por cachos (Barcelos *et al.*, 2000).



Fonte: [www.cirad.fr](http://www.cirad.fr)

**Figura 2:** Formas básicas dos frutos de *Elaeis guineensis*: Dura (A), Tenera (B) e Pisífera (C).

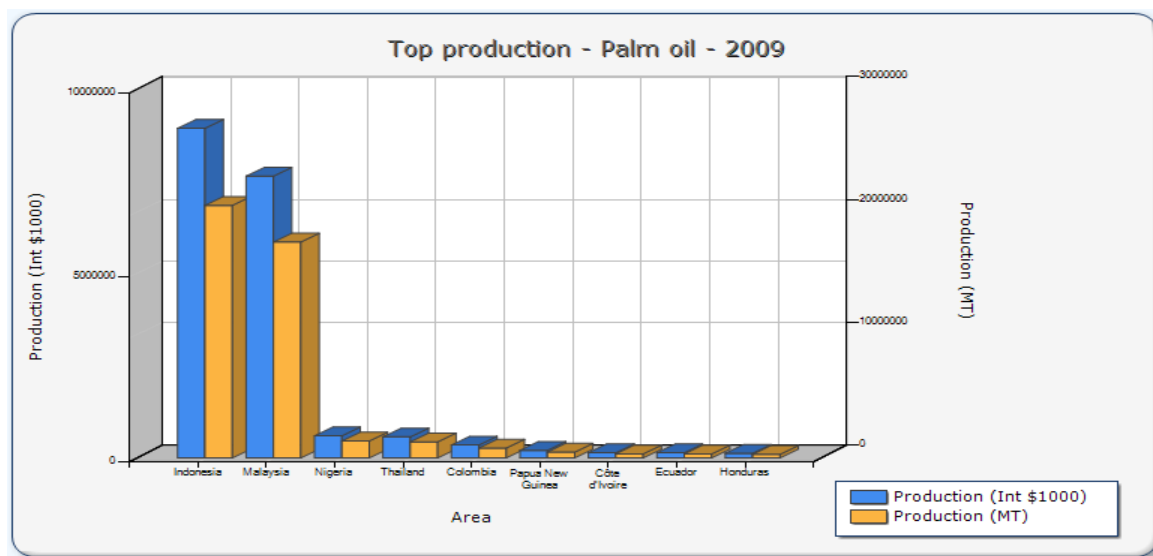
## 1.2 A produção e o uso de *Elaeis*

O dendezeiro já é cultivado comercialmente desde o início do século XX, destacando-se principalmente por ser a oleaginosa com maior produtividade de óleo. O caiaué, com produtividade de óleo inferior, não é cultivado comercialmente, mas apresenta características importantes para uso em programas de melhoramento genético do dendezeiro, como resistência a pragas e doenças e reduzido crescimento vertical do tronco e óleo mais insaturado (Chia, 2008). O dendezeiro é, atualmente, uma das principais fontes de óleo vegetal no mundo. A cultura ocupa o segundo lugar entre as culturas de produção de óleo do mundo e o primeiro no mercado internacional de óleos vegetais e gorduras. O dendezeiro é cultivado em toda a região tropical úmida (Zeven, 1964) e a espécie *E. oleifera* tem sido incorporada em programas de melhoramento (Le Guen *et al.*, 1991).

### 1.2.1 A produção de *Elaeis*

Os maiores produtores de *E. guineensis* são a Indonésia, com produção estimada de

20.550.000 MT/ano, Malásia, 17.564.900 MT/ano, e Nigéria, 1.380.000 MT/ano. O Brasil ocupa a décima posição com produção de 265.000 MT/ano (FAO, 2009) (Figura 3; Tabela 1)



Fonte: www.fao.org

**Figura 3:** A produção em MT e Int\$ dos principais produtores de dendê em 2009

**Tabela 1:** Ranking dos 20 principais países produtores de dendê

Posição	País	Produção (\$1000)	Produção (MT)
1	Indonésia	8940421	20550000
2	Malásia	7641732	17564900
3	Nigéria	600378	1380000
4	Tailândia	569924	1310000
5	Colômbia	349089	802400
6	Papua Nova Guiné	204476	470000
7	Costa do Marfim	141393	325000
8	Equador	139653	321000
9	Honduras	126166	290000
10	Brasil	115290	265000
11	China	100063	230000
12	Costa Rica	82990	190757
13	República Democrática do Congo	80485	185000
14	Camarões	79180	182000
15	Guatemala	78310	180000
16	Gana	56557	130000
17	Filipinas	39155	90000
18	Venezuela	35761	82200
19	México	28278	65000
20	Angola	23928	55000

### 1.2.2 O uso de *Elaeis*

O processamento industrial do dendê tem como propósito fundamental a obtenção de dois produtos: óleo de palma ou óleo de dendê, que é extraído do mesocarpo do fruto, e o óleo de palmiste, que é extraído da amêndoa. O beneficiamento do dendê gera, no entanto, uma variedade muito grande de outros produtos (co-produtos) que, evidentemente, são aproveitados pela indústria. As porcentagens médias de todos os produtos obtidos na extração são de 20% para o óleo de palma bruto, 1,5% para o óleo de palmiste, 3,5 % para a torta de palmiste, 12% para fibras e 5% para cascas (Hartley, 2000).

Devido ao seu baixo custo de produção, boa qualidade e ampla utilização, o óleo de palma é um dos mais requeridos como matéria-prima para diferentes segmentos nas indústrias oleoquímicas, farmacêuticas, de sabões e cosméticos. Seu principal uso é na alimentação humana, responsável pelo consumo de 80% da produção mundial. Suas principais aplicações na alimentação humana são: margarinas, gorduras sólidas, óleo de cozinha, maionese, panificação, leite e chocolate artificiais, fritura industrial, entre outras (Franz, 2000). Ademais, um novo uso tem sido buscado: biocombustível. O óleo de dendê apresenta características muito semelhantes ao diesel, fazendo com que maiores atenções sejam voltadas para este produto (Alptekin *et al.*, 2008).

O óleo de palmiste é rico em ácido láurico e apresenta-se na forma pastosa quando a temperatura ambiente, apresentando uma coloração esbranquiçada. Quando o óleo de palmiste é aquecido ele torna-se um produto líquido de cor levemente amarelada. É utilizado para fins alimentícios e, em especial, na produção de chocolates onde pode substituir a manteiga de cacau. O grande crescimento do óleo de palma no mercado mundial também impulsiona o consumo e a produção do óleo de palmiste já que a extração dos dois produtos é feita a partir do mesmo fruto (Franz, 2000).

As fibras e as cascas de dendê, co-produtos da cultura, podem ser utilizadas para a alimentação de caldeiras como combustíveis. Estima-se que 1,0 ha de dendê, aos quinze anos, tenha seqüestrado 35,87 toneladas de carbono ou 90 toneladas de matéria seca, além de permitir a co-geração de energia em função do potencial energético dos resíduos da extração do óleo, representados por: 1271,5 kg de cachos vazios, 710,5 kg de fibra e 222,5 kg de casca, para cada tonelada de óleo produzida (Viegas e Müller, 2000).

A torta de palmiste, também um co-produto do dendê, pode ser utilizada como fertilizante (adubo orgânico) ou componente de ração para animais e o engaço, suporte fibroso que sustenta o fruto, que constitui 20% a 25% do cacho de frutos frescos, podem ser

utilizados no campo como *mulch* ou são parcialmente desidratados até próximo de 40% de umidade e utilizados como energia extra em caldeiras. Podendo funcionar como fonte de matéria orgânica fornecendo ao solo quantidades consideráveis de nutrientes, podendo, assim, ser utilizado como um complemento de fertilizantes (Viegas e Müller, 2000).

### 1.3 Doenças em *Elaeis* sp.

Aproximadamente 30 doenças são encontradas no gênero *Elaeis*. As principais são as manchas foliares causadas por *Cercospora elaeidis* Steyaert; podridões no colmo causadas por *Ganoderma boninense* Pat., *Ceratocystis paradoxa* Moreau e *Armillaria mellea* Kummer; podridões na coroa causadas por *Marasmius palmivorus* Sharples; podridões no broto causadas por *Erwinia* spp.; antracnose causada por *Botryodiplodia palmarum* (Cooke) Petr. & Syd., *Melanconium elaeidis* Beeli e *Glomerella cingulata* (Stonemam) Spauld & Schrenk; ferrugens causadas por *Pythium splendens* Braun e *Rhizoctonia lamellifera* Small; necroses foliares causadas por *Thanatephorus cucumeris* (Frank) Donk; tombamento de mudas causadas por *Curvularia eragrostidis* Henn; o anel vermelho causado por *Bursaphelenchus cocophilus* (Cobb) Baujard; a marchitez de *Phytomonas* sp.; Potyvirus, Foveavirus e a murcha vascular causada por *Fusarium oxysporum* f.sp. *elaedis*, objeto de estudo deste trabalho (Aderungboye, 1977).

#### 1.3.1 Antracnose

A antracnose do dendezeiro é causada pelos fungos *Botryodiplodia palmarum*, *Glomerella cingulata* ou *Melanconium elaeidis*. *B. palmarum* e *M. elaeidis* provocam sintomas similares, ao passo que de *G. cingulata* produz lesões delimitadas pelas nervuras. Estas infecções estão largamente distribuídas e relacionadas com o manejo das plantas. Os sintomas aparecem em plantas com até 3 meses de idade, podendo, em condições favoráveis, ocorrer também em plantas mais velhas. Inicialmente, os sintomas caracterizam-se por pequenas manchas claras que evoluem para marrom-escuro, circundadas por um halo marrom-pálido e, mais externamente, por uma zona amarelada. Com o passar do tempo, as lesões aumentam de tamanho e suas partes centrais tornam-se acinzentadas (Bergamin Filho *et al.*, 1997).

#### 1.3.2 Mancha foliar de *Cercospora*

Em muitos países, esta é uma das mais importantes doenças foliares do dendezeiro, na



fase de viveiro, devido à marcante suscetibilidade da espécie *E. oleifera* e de seus híbridos com *E. guineensis*. A doença é causada por *Cercospora elaeidis*. A penetração do fungo ocorre através dos estômatos e é favorecida por períodos de alta umidade. Plantas com deficiência nutricional são mais predispostas à infecção. Os sintomas podem aparecer na fase de pré-viveiro ou viveiro e caracterizam-se por pequenas manchas descoloridas, circundadas por um halo verde-pálido. À medida que crescem, as manchas tornam-se deprimidas e de cor marrom. Aproximadamente um mês após o aparecimento da lesão primária, pode-se observar uma série de manchas adicionais, ao seu redor, que coalescem, formando extensa área necrosada. A necrose aumenta rapidamente e o limbo foliar torna-se acinzentado a marrom. Sobre as lesões observa-se abundante esporulação do agente causal. Algumas práticas culturais como a eliminação de folhas velhas com sintomas e a adubação adequada das plantas, auxiliam o controle da doença. No entanto, quando as condições de ambiente são favoráveis ao patógeno, deve-se recorrer ao controle químico (Bergamin Filho *et al.*, 1997).

### 1.3.3 Necrose foliar

É uma doença que ocorre em vários países onde o dendezeiro é cultivado e sua importância varia de região para região e é causada pelo fungo *Corticium solani*, também chamado de *Thanatephorus cucumeris*, cuja fase imperfeita é *Rhizoctonia solani*. Ocorre nas fases de viveiro e pré-viveiro. Os sintomas variam com a idade da folha no momento da infecção. Normalmente, os sintomas começam nas folhas mais jovens (folhas ainda não totalmente abertas), podendo ocorrer também na base da flecha dos *seedlings* jovens. Neste caso, a lesão inicialmente é irregular, verde-olivácea, limitada por bordos violeta-marrom. A lesão torna-se proeminente com a expansão da flecha. Inicialmente, o tecido lesionado tem aspecto úmido e quando a flecha se expande, a lesão seca rapidamente tornando-se acinzentada com margem marrom brilhante. Para o controle, as plantas afetadas devem ser removidas do viveiro. Deve-se evitar o excesso de água e sombra (Bergamin Filho *et al.*, 1997).

### 1.3.4 Amarelecimento fatal

Esta doença foi constatada pela primeira vez em 1974, no município de Benevides, Estado do Pará. É a doença mais importante da cultura na América Latina. Inicialmente, observa-se amarelecimento dos folíolos basais das folhas mais jovens. Em seguida, surgem

necroses, que evoluem da extremidade para a base, até causar a seca completa da folha. Plantas com 2 anos de idade já mostram sintomas. O mesmo quadro sintomatológico ocorre nas flechas. É muito comum folhas e flechas quebrarem-se na base. As partes internas do estipe não apresentam alterações, mas há redução na emissão de raízes. Os cachos formados antes do aparecimento dos sintomas chegam a ser colhidos, mas cachos mais jovens bem como inflorescências formados em plantas sintomáticas abortam e secam. Embora não se conheça ainda o patógeno do AF, tem-se evidência de que a espécie *E. oleifera*, não é afetada. Assim, vem sendo desenvolvido um intenso programa de melhoramento genético, visando incorporar a resistência de *E. oleifera* na espécie comercial *E. guineensis*. Outra medida recomendada para o controle da doença é a inspeção periódica da plantação, com frequência semanal ou quinzenal, para identificar e eliminar as plantas com sintomas iniciais (Bergamin Filho *et al.*, 1997).

#### 1.3.5 Anel vermelho

É uma das mais importantes doenças do dendezeiro na América Latina. Causada pelo nematóide *Bursaphelenchus cocophilus* cujo vetor é *Rhynchophorus palmarum*. A planta apresenta, inicialmente, uma redução no crescimento das folhas centrais, com nítida tendência de agrupamento das mesmas. Em seguida, estas folhas centrais tornam-se amareladas e seu pecíolo pode exibir tonalidade alaranjada. As inflorescências das plantas afetadas normalmente abortam. Os frutos dos cachos já formados apodrecem antes da maturação. Internamente, a planta apresenta tecidos necrosados que em corte transversal têm o formato de um anel. Algumas vezes, a necrose é observada apenas nos pecíolos. Nas partes necrosadas quase sempre encontra-se o agente patogênico. Em geral, os sintomas ocorrem em plantas com mais de 5 anos. Não se conhece nenhum método capaz de salvar uma planta de dendezeiro afetada pelo anel vermelho. Assim, plantas enfermas devem ser eliminadas. O controle de besouros por armadilhas é uma medida adicional que deve ser adotada (Bergamin Filho *et al.*, 1997).

#### 1.3.6 Marchitez sorpressiva

Esta doença foi relatada apenas na América do Sul e ataca plantas com mais de dois anos de idade. É causada pelo protozoário flagelado *Phytomonas*. Caracterizam-se por uma coloração amarronzada nas extremidades dos folíolos das folhas mais velhas, progredindo

para a base e provocando uma rápida seca de toda a folha. É comum os folíolos mostrarem-se totalmente secos, enquanto a ráquis ainda exibe uma coloração verde-pálida, secando posteriormente. A seca atinge rapidamente as folhas medianas, enquanto as flechas apodrecem nas plantas em estágio avançado da doença flechas normalmente quebram na base. As inflorescências abortam e os cachos imaturos apodrecem. Para o controle, recomenda-se o controle químico dos insetos vetores. Aparentemente, a espécie *E. oleifera* não é afetada pela marchitez sorpressiva, podendo ser empregada na obtenção de híbridos com *E. guineensis* (Bergamin Filho *et al.*, 1997).

### 1.3.7 Podridão do colmo

A maior ameaça para a produção sustentável de dendê no Sudeste Asiático é a podridão causada pelo fungo *Ganoderma boninense*. A podridão basal é a manifestação mais comum da doença. É caracterizada por uma deterioração do colmo, produção de múltiplas lanças e de corpos de frutificação na base da estipe. Em casos mais graves, a planta pode tombar. Perdas mais graves por esta doença ocorrem na Indonésia e Malásia, com incidência também na África, Papua Nova Guiné e Tailândia. Em áreas costeiras da Malásia foram relatadas perdas que variavam de 50% a 80%. Em função da doença há um declínio na produção de frutas frescas. Programas de melhoramento buscando resistência a *Ganoderma* estão em andamento em diversos centros de pesquisa (Flood *et al.*, 2011).

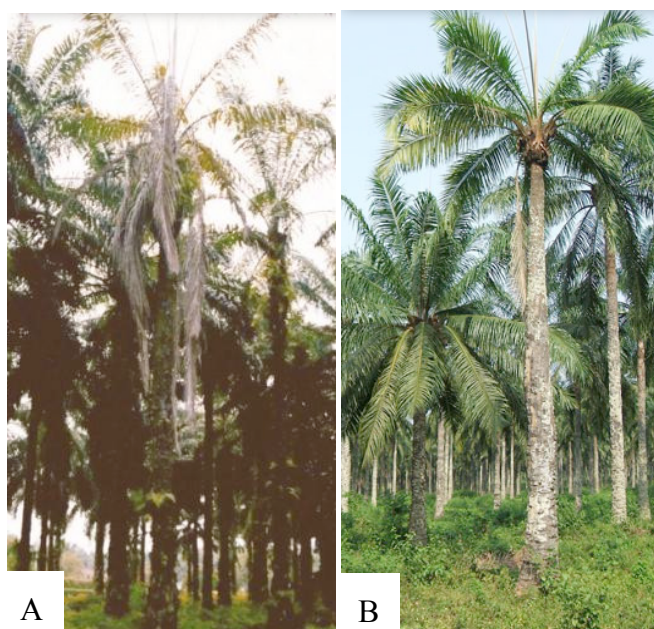
### 1.3.8 Murcha vascular

A murcha vascular do dendê, também chamada fusariose, é causada pelo patógeno *Fusarium oxysporum* Schlechtend.:Fr. f.sp. *elaedis* Toovey, um fungo do solo que produz macro e microconídios e clamidósporos, sendo estes últimos capazes de sobreviver em restos culturais e no solo. O principal hospedeiro é *E. guineensis* e a inoculação artificial com isolados do patógeno em *E. oleifera* é efetiva (Renard *et al.*, 1980).

A doença foi descrita pela primeira vez na República Democrática do Congo (Wardlaw, 1946) e, posteriormente, na Costa do Marfim, Nigéria, Gana, Camarões (Flood, 2006), Brasil (Van de Lande, 1984) e Equador (Renard, 1989). Hoje, a murcha vascular é a doença mais destrutiva do dendezeiro na África (Flood, 2006).

Duas síndromes são observadas em plantas adultas (Figura 4). Na primeira, chamada de *acute wilt*, as folhas secam e morrem rapidamente, mantendo a sua posição ereta até a

queda da folha. A doença progride rapidamente e as plantas morrem dentro 2 a 3 meses. Na segunda síndrome, chamada de *chronic wilt*, as plantas permanecem vivas vários meses e até mesmo anos, mas sofrem enfezamento. As folhas mais velhas tornam-se desidratadas e ficam penduradas em torno da coroa. Estes sintomas progridem gradualmente e as folhas mais jovens são afetadas. As folhagens novas que são produzidas na coroa permanecem e se tornam cloróticas e atrofiadas. O resultado é um achatamento da coroa. O ápice do tronco também pode ter seu diâmetro reduzido. A planta pode permanecer vários anos nesta condição (Corley *et al.*, 2003).



Fonte: [www.cirad.fr](http://www.cirad.fr)

**Figura 4:** As duas síndromes causadas por *Fusarium oxysporum* f.sp. *elaeidis*: *acute wilt* (A) e *chronic wilt* (B)

Segundo Flood (2006), vários estágios intermediários entre as formas aguda e crônica podem ocorrer e nenhuma diferença entre isolados das duas síndromes foi detectada. Plantas jovens também mostram encurtamento progressivo das folhas mais jovens, dando a aparência de achatamento da coroa, bem como dessecação e morte das folhas mais velhas.

Estudos detalhados sobre a fisiologia da doença feitos por Mepsted *et al.* (1995) indicaram que ocorre um déficit nutricional da planta devido a uma diminuição no tamanho do pecíolo e da lâmina foliar, causada quase inteiramente por uma redução na divisão celular. Isto ocorre por causa de alterações no nível ou na atividade de giberelinas.

Internamente são observados escurecimento do sistema vascular tanto em palmeiras

adultas como em mudas (Corley *et al.*, 2003). O bloqueio dos vasos do xilema com calosidades e tiloses pode ser facilmente observado ao microscópio de luz (Mepsted, 1992) e, mesmo em plantas altamente doente, a maior parte das raízes não mostram sinais de infecção (Prendergast, 1957).

A doença começa com a invasão do fungo por meio das raízes e seu crescimento nos elementos de vaso do xilema. A colonização do xilema pelos microconídios ocorre no fluxo de transpiração. O fungo cresce através da parede, colonizando, assim, o vaso adjacente. Durante o tempo que o patógeno tem para crescer através da parede de um vaso, o hospedeiro tem tempo para limitar sua propagação através da produção de gomas, géis e tiloses que impedem a transpiração (Paul, 1995).

Em material resistente, este mecanismo de defesa começa a funcionar rapidamente e é acompanhado pelo acúmulo de compostos antifúngicos. Assim, a oclusão dos vasos restringe ainda mais a colonização do fungo e permite o acúmulo de antifúngicos a níveis inibitórios.

Isso ocorre tanto nos vasos invadidos como nos vasos adjacentes. No material suscetível, calosidades e géis também são produzidos, mas sua produção é mais lenta, assim como o acúmulo de compostos antifúngicos. Isso permite que o patógeno consiga colonizar outros vasos, levando a uma maior oclusão vascular e, finalmente, os sintomas externos (Corley *et al.*, 2003; Paul, 1995).

A infecção ocorre pode ter seu início por contato da raiz com tecido infectado de outra planta. Restos culturais podem servir como fontes nutricionais para o patógeno, aumentando a quantidade de inóculo inicial e o risco de infecção (Renard, 1989).

Por ser uma doença causada por um patógeno presente no solo, o controle é, ainda, bastante difícil. Práticas culturais, tais como a remoção e queima de plantas infectadas têm sido defendidas e utilizadas, mas são impraticáveis a nível de hectares de plantação. Além disso, devido à interligação das raízes das plantas adjacentes, muitas palmeiras também precisariam ser removidas ao ser adotada esta técnica (Flood, 2006). O uso de *Brachiaria* como cultura de cobertura tem sido estudada (Renard, 1983) ao invés de culturas tradicionalmente utilizadas, tais como *Pueraria* ou *Calapogonium caeruleum* (leguminosas fixadoras de nitrogênio), que tendem a aumentar a incidência da doença (Corley, 2003). A doença também pode ser controlada através de um melhor manejo da área plantada. Estudos de Renard *et al.*, (1991) mostraram que a incidência da doença pode ser reduzida pela metade com o plantio de novas plantas com distância mínima de dois metros umas das outras.

No entanto, apesar de alguns sucessos com o manejo da doença em campo, a busca de resistência continua sendo o único método prático de longo prazo para o controle da

Fusariose. Programas de melhoramento, como os da Costa do Marfim, têm sido bem sucedidos por causa da introdução de variedades resistentes. Estes programas conseguiram reduções de perdas de 20-30% para menos de 3% (Franqueville *et al.*, 1990). No entanto, programas de melhoramento de culturas perenes são de longo prazo. De 4 a 7 anos são necessários para selecionar material com base em ensaios de campo com infecção natural.

## **1.4 Resistência de plantas às doenças**

### 1.4.1 Interação Planta-Patógeno

As plantas precisam se defender contra o ataque de vírus, microorganismos, invertebrados e até outras plantas. Como não possuem um sistema circulatório de defesa, cada célula deve possuir um mecanismo pré-formado e/ou induzido de defesa, diferentemente do que ocorre com o sistema imunológico dos vertebrados (Walbot, 1985).

Após a redescoberta do trabalho de Mendel, os melhoristas de plantas reconheceram que a resistência à doença era muitas vezes herdada como um gene dominante ou de dominância incompleta (Keen, 1990). Desde então, houve um considerável acúmulo de conhecimento baseado em pesquisas nas áreas bioquímica, biologia molecular e genética no que diz respeito à resistência de plantas às doenças e o uso de cultivares resistentes se tornou a estratégia mais valiosa para controlar as doenças das plantas cultivadas (Crute *et al.*, 1996).

### 1.4.2 Tipos de patógenos

Os fitopatógenos podem ser classificados em três tipos de acordo com a relação que estabeleçam com a planta. Patógenos que mantêm vivo seu hospedeiro e se alimentam de tecidos vegetais vivos são chamados biotróficos. Os chamados necrotróficos freqüentemente produzem toxinas ou enzimas que degradam o tecido infectado e promovem a rápida liberação de nutrientes. Alguns patógenos são biotróficos durante os primeiros estágios da infecção, mas tornam-se necrotróficos posteriormente. Estes patógenos são chamados hemibiotróficos (Bergamin Filho *et al.*, 1997)

A maioria dos patógenos biotróficos e hemibiotróficos só podem causar doença em um grupo relativamente pequeno de plantas hospedeiras por causa do conjunto ligeiramente diferente de genes especializados e mecanismos moleculares necessários para cada interação hospedeiro-patógeno. Quando um patógeno é capaz de causar doença em uma espécie em

particular, dois resultados são possíveis: uma resposta compatível, que é uma interação que resulta em doença, e uma resposta incompatível, que é uma interação que resulta em uma forma leve da doença ou nenhuma (Bergamin Filho *et al.*, 1997).

#### 1.4.3 Mecanismos de defesa

As plantas possuem mecanismos de defesa pré-determinados e induzíveis. O primeiro é o principal mecanismo no caso de resistência não específica, em que as plantas sintetizam peptídeos, proteínas e metabólitos secundários, que restringem a infecção por patógenos (Heath, 2000). Este tipo de defesa inclui muitas barreiras pré-formadas, tais como paredes celulares, tricomas e cutículas mais espessas. Estes componentes não só protegem as plantas contra a invasão, mas também dão rigidez a elas (Freeman e Beattie, 2008).

As plantas podem ativar um arsenal muito eficaz de respostas de defesa induzível, composta pela morte geneticamente programada de células infectadas (a reação de hipersensibilidade, HR), bem como reforçar e produzir antibióticos no local da infecção (Hammond-Kosack *et al.*, 1996). Estas respostas locais podem, por sua vez, desencadear uma resposta duradoura sistêmica - resistência sistêmica adquirida – (*systemic acquired resistance*, SAR), que prepara a planta, em termos de resistência, contra um amplo espectro de patógenos (Métraux, 2001). Esta resposta requer uma disposição substancial de recursos celulares, incluindo a reprogramação genética extensa e uma realocação metabólica (Somssich e Hahlbrock, 1998). Assim, as defesas são mantidas sob controle genético e são ativadas apenas se a planta detecta um invasor em potencial.

#### 1.4.4 Bases citológicas e histológicas da resistência

Ao se fazer uma análise citológica e histológica da parede celular na pré-infecção, pode-se observar componentes que contribuem grandemente para a proteção da planta como cutina, carboidratos, géis, tiloses e deposição de lignina.

Segundo Wang e Pinckard (1973), a cutina atua como uma barreira química e física na germinação e penetração de alguns patógenos – como os fungos – dentro das células. A espessura da cutina tem sido correlacionada com níveis de resistência da planta para o patógeno que penetra diretamente dentro dos tecidos.

Quando o patógeno começa a penetrar a parede celular, podem ser sintetizados carboidratos, principalmente calose e celulose, em maior quantidade. Estes reforçarão o

interior da parede celular, dificultando a penetração do patógeno. Esta deposição pode continuar após a penetração do patógeno, formando, assim, uma papila (Aist, 1976).

Os géis são formados pelas plantas aproximadamente 48 horas após o ataque do patógeno. Plantas suscetíveis e resistentes possuem a habilidade de produzir estes géis, embora eles sejam degradados mais rapidamente em plantas que apresentem suscetibilidade ao microorganismo (Matiello *et al.*, 1999).

Protusões da parede celular das células paravasculares por meio das pontuações do xilema, formam as tiloses, um outro mecanismo de defesa induzível. Após a formação das tiloses, ocorre o espessamento da parede celular da célula, dificultando o transporte do patógeno (Bell, 1981).

A capacidade de uma planta induzir rapidamente a lignificação de um tecido também é um mecanismo de defesa e ocorre na própria célula afetada e ao redor das lesões. Este tipo de reação só ocorre em cultivares resistentes (Reed *et al.*, 1979).

#### 1.4.5 Bases genéticas da resistência

A incompatibilidade entre planta e patógeno resulta na resistência genética. Porém, a evolução do patógeno, devido a pressão de seleção, mutações, alterações nas frequências alélicas e processos adaptativos, pode levar a uma quebra da resistência, seja parcial, seja total (Hayden, 2003).

O desenvolvimento de novas abordagens para a proteção das culturas exige avanços contínuos na compreensão dos mecanismos moleculares que regulam a imunidade da planta. Durante as duas últimas décadas, ferramentas moleculares e genômicas têm possibilitado nossa compreensão do sistema imunológico vegetal, com dois ramos fundamentais atualmente reconhecidos: *PAMP-triggered immunity* e *effector-triggered immunity*

##### 1.4.5.1 *PAMP-triggered immunity*: a resistência não-hospedeiro

Um ramo do sistema imunológico vegetal, conhecido como *PAMP-triggered immunity* (PTI), a defesa inata, defesa basal ou não-auto-reconhecimento, é baseado em interações entre padrões de reconhecimento de *receptors-like kinases* (PRRs) (Shiu e Blecker, 2001; Altenbach e Robatzek, 2007) na matriz extracelular de plantas ou na membrana plasmática (Boller e Felix, 2009), que reconhecem padrões moleculares conservados associados a patógenos (PAMPs) (Nürnberg e Kemmerling, 2009). Os PAMPs são epítomos microbianos,



tais como a flagelina bacteriana, fator Tu de alongamento bacteriano, lipopolissacarídeos, peptidoglicanos,  $\beta$ -glucanos, ergosterol ou quitinas fúngicas (Boller e Felix, 2009), que, originalmente chamado de elicitores, são definidos como moléculas conservadas em uma classe de patógenos e considerados essenciais na função de aptidão ou sobrevivência. Dado que estas moléculas também podem ocorrer em microrganismos não patogênicos, o termo padrões moleculares associados a microrganismos (MAMPs) também pode ser empregado (Boller e Felix, 2009). Produtos da degradação da parede celular da planta hospedeira, que são liberados como resultado da presença de enzimas hidrolíticas (Vorwerk *et al.*, 2004), também servem como elicitores endógenos. Tais interações provocam a PTI, que envolve a ativação de uma proteína mitógeno associada (MAP), cascatas quinase e ativação de fatores de transcrição WRKY (TF), impedindo a colonização do hospedeiro, e conferem resistência à maioria dos patógenos em potencial. PTI é considerada responsável pela resistência do tipo não-hospedeiro. As plantas normalmente apresentam respostas semelhantes a PTIs quando do reconhecimento de PAMPs diversas.

#### 1.4.5.2 *Effector-triggered immunity*: a resistência raça-específica

Um segundo ramo do sistema imunológico vegetal atua quando os patógenos têm sucesso na superação da PTI. Genes da avirulência do patógeno (genes Avr) evoluíram segundo a espécie, raça, ou mesmo efetores de virulência específicos de estirpes (anteriormente referido como fatores de avirulência), que, após a secreção e translocação dentro da célula hospedeira, suprimem ou interferem na sinalização da PTI, tornando as plantas suscetíveis ao patógeno e tendo a doença como um resultado desta *effector-triggered susceptibility* (ETS) (Göhre e Robatzek, 2008; Chisholm *et al.*, 2006). Assim, os efetores diferem dos PAMPs, pois atingem especificamente os mecanismos de defesa do hospedeiro. Bactérias patogênicas podem transferir estes efetores para as células do hospedeiro suscetível através de sistemas de secreção do tipo III. As plantas, em resposta, co-evoluem formando receptores específicos que reconhecem os sinais moleculares dos efetores de um patógeno (tal que a proteína de virulência é agora classificada como uma proteína de avirulência) e ativar uma segunda linha de defesa conhecido como *effector-triggered immunity* (ETI) (Jones e Dangl, 2006), também conhecida, anteriormente, como resistência raça-específica ou gene-gene. A ETI parece ter evoluído depois da PTI, com a aquisição de proteínas de resistência necessárias. Em contraste com a PTI, que ocorrerá em todos os membros de uma determinada

espécie vegetal, a ETI é restrita a determinados genótipos dentro de uma espécie de planta, refletindo o polimorfismo dos genes R de resistência (Xiao *et al.*, 2008).

Foram elucidados quatro modelos importantes dentro da ETI: o modelo gene-a-gene, o modelo zigzague, o modelo guarda, e, mais recentemente, o modelo *decoy*.

#### 1.4.5.2.1 O modelo gene-a-gene

Segundo este modelo, que foi originalmente descrito por Flor para explicar o patossistema *Melampsora lini*-linho, quando a detecção de um patógeno potencial, as plantas expressam grandes conjuntos de genes chamados de genes R de resistência. Os genes de resistência codificam receptores putativos que respondem aos produtos dos chamados genes Avr (genes de avirulência) expressos pelo patógeno durante a infecção. Estudos mais recentes de genética molecular com genes Avr clonados têm mostrado que muitos genes Avr promovem a virulência em hospedeiros que não expressam o gene de resistência correspondente. Isto sugere a co-evolução do patógeno e do hospedeiro, onde o primeiro evolui em termos de genes de virulência e o segundo, por sua vez, evolui em genes de resistência que reconhecem proteínas de virulência como sinais de invasão. Assim, uma proteína de virulência pode tornar-se uma proteína de avirulência se uma proteína de resistência no hospedeiro reconhecê-la (McDowell e Woffenden, 2003).

#### 1.4.5.2.2 O modelo zigzague

No modelo zigzague, uma vez que um fator Avr ou em efector do patógeno é reconhecido, se diminui o *fitness* deste. A pressão de seleção pode, então, resultar na perda ou alteração de efetores, ou no ganho de outros efetores (como novos fatores de virulência) através da evolução da população. Conseqüentemente, para evitar a extinção população de plantas, novos *loci* de genes de resistência devem evoluir para reconhecer estes novos efetores, a fim de recuperar a resistência (Jones e Dangl, 2006).

#### 1.4.5.2.3 O modelo de guarda

Este modelo prevê que as proteínas de resistência ativam esta resistência quando elas interagem com outra proteína vegetal (uma proteína *guardee*) que é modificada pelo patógeno na tentativa de se criar um ambiente favorável. A resistência é acionada quando a proteína de

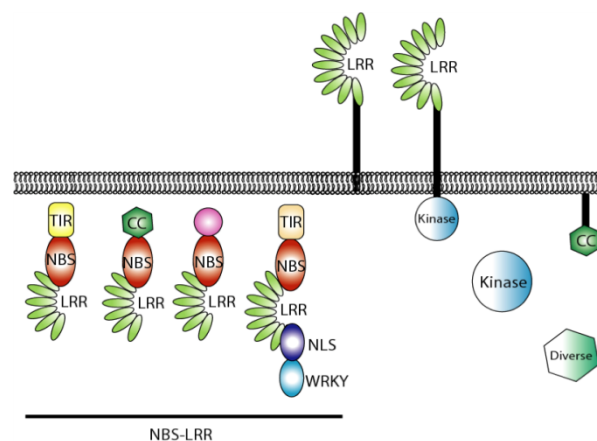
resistência detecta uma tentativa de atacar sua proteína *guardee*, o que não envolve necessariamente a interação direta entre as proteínas de resistência e Avr (Van der Hoorn *et al.*, 2002). As proteínas do patógeno, embora tradicionalmente denominados fatores de avirulência, atuam como fatores de virulência, e são chamadas de efetores (Bent e Mackey, 2007).

#### 1.4.5.2.4 O modelo *decoy*

No modelo *decoy*, a proteína de resistência não interage diretamente com proteínas Avr, mas monitora modificações em proteínas *decoy* do hospedeiro. Esta proteína *decoy*, embora sem função de proteção, imita um componente de defesa das plantas, e, como tal, atrai a proteína Avr (van der Hoorn *et al.*, 2008).

### 1.5 Genes de resistência

Genes de resistência a doenças têm sido clonados e caracterizados para monocotiledôneas e dicotiledôneas (Hammond-Kosack e Jones, 1996). Até o momento, mais de 50 genes de resistência foram clonados para uma grande variedade de espécies (Wenkai *et al.*, 2006). Estes genes codificam proteínas que podem ser agrupadas em famílias, com base na homologia de seus domínios e função bioquímica (Dangl e Jones, 2001). Até o momento, existem seis diferentes famílias de genes de resistência já identificadas, sendo cinco com base em sua estrutura e uma família adicional de genes que não se conformam com as cinco primeiras (Figura 5).



**Fonte:** Emediato, 2009

**Figura 5:** As seis diferentes famílias de genes de resistência

### 1.5.1 Família NBS-LRR: *Nucleotide binding site* (NBS) e *Leucine rich repeat* (LRR)

A família NBS-LRR é a maior classe de genes de resistência conhecida até então. É também a família mais importante, sendo todos os genes conhecidos ligados, de alguma forma, à resistência. Genes desta família conferem resistência a bactérias, fungos, vírus, nematóides e afídeos (Timmerman-Vaughan *et al.*, 2000). Nenhuma outra função a não ser esta foi relatada para esta grande família de genes (Bai *et al.*, 2002).

Os produtos dos genes desta família são compostos por um conservado N-terminal (região NBS) e um domínio LRR carboxi-terminal variável em tamanho, composto por 10 a 40 motivos curtos de LRR (Cannon *et al.*, 2002). O domínio NBS é capaz de se ligar ao ATP e está envolvido em vias de transdução de sinal, acionadas pela presença do patógeno (Ellis e Jones, 1998; van der Biezen e Jones, 1998). Esta interação resulta numa resposta de defesa e morte celular. Este domínio compreende duas subclasses: TIR (*Toll interleucine receptor*) NBS e não-TIR NBS, baseados em aminoácidos localizados no N-terminal que mostram homologia com domínios de sinalização intracelular de *Drosophila* (Toll) e uma interleucina (IL) basal de mamíferos. O receptor Toll controla a resposta imune a bactérias gram-positivas e fungos através da ativação de uma via de sinalização. O grupo TIR está presente em dicotiledôneas, enquanto o grupo não-TIR ocorre tanto em monocotiledôneas quanto em dicotiledôneas. O NBS também pode estar ligado a um coiled coil (CC) (cauda espiralada) no N-terminal ou então pode estar ligado a um motivo TIR no N-terminal e um NLS e WRKY no C-terminal. Têm-se, portanto, 4 subfamílias (Miller *et al.*, 2009):

1. NBS-LRR com motivo TIR no N-terminal (por exemplo, o gene *Bs4* que confere ao tomate – *Lycopersicon esculentum* – resistência a *Xanthomonas caepralis* pv. *vesicatoria*)
2. NBS-LRR com motivo CC no N-terminal (por exemplo, o gene *Bs2* que confere ao pimentão – *Capsicum annuum* – resistência a *Xanthomonas caepralis* pv. *vesicatoria*)
3. NBS-LRR com N-terminal sem motivo CC (por exemplo, o gene *RGC2* que confere ao gênero *Musa* resistência a *Fusarium oxysporum* f.sp. *cubensis*)
4. NBS-LRR com motivo TIR no N-terminal, um domínio WRKY (proteína ativadora transcricional) e um domínio NLS (sinal de localização nuclear), ambos no C-terminal (por exemplo, o gene *RRS1* que confere a *Arabidopsis thaliana* resistência a *Ralstonia solanacearum*)

O domínio LRR está presente em numerosas proteínas e estão relacionados à ligação

proteína-proteína, ligações peptídeo-ligantes e interações carboidrato-peptídeo (Dangl e Jones, 2001). Nas proteínas NBS-LRR, a superfície LRR de repetições de unidades  $\alpha$ -hélice-loop e  $\beta$ -strand-loop com resíduos diferentes de leucina parece ser a responsável pelas interações de reconhecimento específico de elicitores (proteínas Avr) produzidos pelo patógeno (Kobe e Deisenhofer, 1994; Jones e Jones, 1997).

Vários genes de resistência pertencentes a esta família já foram clonados e bem caracterizados. São exemplos os genes de resistência de *Arabidopsis thaliana*, linho, tabaco, tomate e arroz (Bent *et al.*, 1994; Mindrinos *et al.*, 1994; Whitham *et al.*, 1994; Grant *et al.*, 1995; Lawrence *et al.*, 1995; Milligan *et al.*, 1998; Yoshimura *et al.*, 1998).

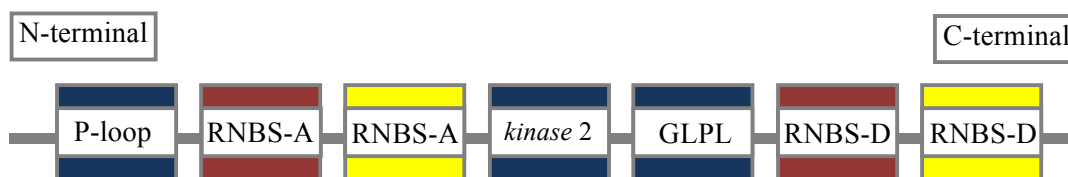
#### 1.5.1.1 A organização dos genes NBS-LRR no genoma das plantas

Os *loci* NBS-LRR podem ser encontrados como genes isolados (*singletons*) ou em *clusters* (Holub, 2001). Em alguns casos, estes *clusters* contém cópias dos genes NBS-LRR de clados filgeneticamente diferentes (Hulbert *et al.*, 2001; Leister *et al.*, 1996). A partir do sequenciamento do genoma de *Arabidopsis* - 125 Mb no total - pode-se analisar a diversidade e organização das sequências de genes de resistência desta família. Aproximadamente 150 sequências mostraram homologia à família NBS-LRR. Estes genes de resistência estão desigualmente distribuídos nos cromossomos, sendo 49 no cromossomo I, 2 no cromossomo II, 16 no cromossomo III, 28 no cromossomo IV e 55 no cromossomo V. Embora haja muitos destes genes de resistência em famílias multigênicas, há 46 homólogos de genes de resistência em *singletons*, 25 *doublets*, 8 *loci* com três cópias, e *loci* individuais com quatro, cinco, sete, oito e nove genes codificantes de NBS-LRR. Há mais genes TIR-NBS-LRR (94) que genes não-TIR-NBS-LRR (55) (Meyers *et al.*, 2003). No caso do *Oryza sativa* L. (var. Nipponbare) – com um genoma de 420 Mb – há aproximadamente 535 sequências codificadoras de NBS, incluindo 480 genes não-TIR-NBS-LRR (Goff *et al.*, 2002; Zhou *et al.*, 2004). A distribuição cromossômica de genes codificadores de NBS em arroz varia de 20 no cromossomo 9 a 133 no cromossomo 11. Estes 480 genes não-TIR-NBS-LRR representam cerca de 1% de todas as ORFs previstas no genoma do arroz (Goff *et al.*, 2002), enquanto no genoma de *A. thaliana*, TIR e não-TIR-NBS-LRR representam 0,43% do total previsto de ORFs. Assim, tanto o número absoluto quanto o relativo de representação de genes NBS-LRR no genoma de arroz são claramente mais elevados do que no genoma *Arabidopsis* (Bai *et al.*, 2002). 263 genes codificadores de NBS residiam em 44 *clusters* e o número médio de genes em cada *cluster* era seis. Há 15 grupos com quatro cópias, 13 com cinco, 4 com seis, 4 com sete, 3 com oito, 3

com dez, 1 com 15 e 1 com 17 genes codificadores de NBS . Além destes *clusters*, havia 40 doublets e 17 de triplets. Portanto, um total de 394 genes residiam em um *cluster* ou em estavam em tandem (Zhou *et al.*, 2004).

### 1.5.1.2 Os motivos conservados no domínio NBS

O domínio NBS, TIR e não-TIR, em genes de resistência ou análogos de genes de resistência possuem motivos amplamente conservados em função da diversidade de patógenos (Figura 6). Em NBS TIR e não-TIR há motivos P-loop, *kinase 2* e GLPL. Os motivos RNBS-A e RNBS-D estão presentes em ambos, TIR e não-TIR, mas apresentam diferenças entre eles.



Fonte: Emediato, 2009

**Figura 6:** Alguns motivos conservados dentro do domínio NBS. Nos retângulos azuis estão os motivos presentes em ambos, TIR e não-TIR; nos retângulos vermelhos os motivos exclusivos não-TIR; nos retângulos amarelos os motivos exclusivos TIR.

O motivo P-loop, também chamado *kinase-1*, está envolvido na ligação com fosfatos e íons Mg. A *kinase-2*, que participa de reações de transferência de grupos fosfato, contém quatro aminoácidos hidrofóbicos consecutivos seguidos por resíduo conservado de aspartato. O motivo GLPL possui a sequência glicina, leucina, prolina e leucina de aminoácidos (Liu *et al.*, 2007). Os motivos RNBS-A e RNBS-D apresentam grande variabilidade na sua composição.

### 1.5.2 Família RLP: Receptor-like proteins

Os genes de resistência desta família codificam proteínas que funcionam como receptores (*receptor-like proteins* – RLPs) e possuem um domínio LRR extracelular no N-terminal ancorado à célula com pequenos domínios transmembrana. Não existem domínios

sinalizadores intracelulares. Em tomate (*Lycopersicon esculentum*), os receptores LRR estão ancorados na membrana e são codificados por um complexo de genes *cf* que conferem resistência a *Cladosporium fulvum* (Wulff *et al.*, 2009).

RLPs também estão envolvidos no desenvolvimento das plantas. Em *Arabidopsis* há o gene *Tmm* (Yang e Sack, 1995; Nadeau e Sack, 2002), que está envolvido na padronização estomática, e o gene *Clavata2* (*Clv2*) (Jeong *et al.*, 1999) e o seu ortólogo em milho (*Zea mays*) *Fasciated Ear2* (*Fea2*) (Taguchi-Shiobara *et al.*, 2001), que estão envolvidos no desenvolvimento de meristema.

### 1.5.3 Família RLK: Receptor-like kinases

As proteínas codificadas pelos genes desta família possuem um domínio LRR extracelular, um domínio serina/treonina quinase intracelular e um domínio transmembrana. O domínio LRR está relacionado à recepção de estímulos externos e a serina/treonina quinase, às cadeias de fosforilações na transdução de sinais do citoplasma ao núcleo celular (Song *et al.*, 1995).

As proteínas desta família de genes podem ser divididas em duas categorias principais (Shiu *et al.*, 2004). A primeira categoria funcional consiste em quinases envolvidas no crescimento e desenvolvimento celular. Em *Arabidopsis* há genes envolvidos na abscisão de órgãos florais (*Haesa*) (Jinn *et al.*, 2000.), na regulação do crescimento celular (*Bri1*) (Li e Chory, 1997) e no desenvolvimento do meristema (*Clavata1*) (Clark *et al.*, 1997). Na segunda categoria estão as proteínas envolvidas na interação planta-patógeno e respostas de defesa, como o gene *Xa21*, que confere resistência às raças de *Xanthomonas oryzae* pv. *oryzae* em arroz (Song *et al.*, 1995) e o gene de percepção da flagelina em *A. thaliana*, *Fls2* (Gomez-Gomez e Boller, 2000).

### 1.5.4 Família serina/treonina quinase citoplasmática sem domínio LRR (RSTK)

Esta família codifica proteínas que não contém domínios NBS-LRR. Possuem uma quinase que fosforila resíduos de serina e treonina (Tang *et al.*, 1996). O exemplo mais conhecido, o gene *AvrPto*, confere resistência a algumas estirpes de *Pseudomonas syringae* pv. *tomato* em tomate. Em experimentos *in vitro*, *Pto* se mostrou capaz de autofosforilação em vários sítios. A função do gene *Pto* requer mediação do gene *Prf*, membro da classe NBS-LRR (Jones *et al.*, 1997) que tem ligação próxima ao gene *Pto* (Salmeron *et al.*, 1996). A

resistência conferida pelo gene *Pto* está associada à reação de hipersensibilidade e à morte celular programada.

#### 1.5.5 Família com domínio coiled coil ancorado à membrana

Esta família de genes de resistência foi encontrada em *Arabidopsis*. O principal representante, o gene *RPW8*, codifica uma redutase localizada na membrana, com uma estrutura helicoidal, sem semelhança alguma com qualquer outro gene de resistência (Kobe e Kajava, 2001). O gene *RWP8* ativa a produção de ácido salicílico como estratégia de defesa e promove o acúmulo local de H<sub>2</sub>O<sub>2</sub>, presumivelmente para restringir o crescimento do haustório reduzindo ao mesmo tempo o dano oxidativo às células do hospedeiro (Wang *et al.*, 2009), conferindo um amplo espectro de resistência (Xiao *et al.*, 2003).

#### 1.5.6 Genes de resistência que codificam proteínas diversas

São agrupados nesta família, os genes que codificam R proteínas que não se alinham com aquelas das 5 primeiras famílias. Vários genes foram descritos e conferem resistência a uma grande variedade de patógenos. Alguns exemplos são o gene *HS1*, que confere resistência a *Heterodera schachtii* em beterraba (*Beta vulgaris*) (Cai *et al.*, 1997), o gene *Mlo*, que codifica uma proteína da membrana confere resistência a *Erysiphe graminis* f.sp. *hordei* em cevada (*Hordeum vulgare*) (Büschges *et al.*, 1997) e o gene *RTM2*, que confere resistência ao *tobacco etch potyvirus* (TEV) em *Arabidopsis* (Whitham *et al.*, 2000).

### 1.6 Análogos de genes de resistência (RGAs)

Uma alternativa para encontrar genes de resistência em plantas é a busca de análogos de genes de resistência (RGAs). Os RGAs podem ser identificados por similaridade de sequência: certos domínios funcionais, NBS e LRR, por exemplo, são altamente conservados mesmo entre genes de resistência remotamente relacionadas. Estudos genômicos revelaram que as famílias NBS-LRR, RLP e RLK estão presentes em quase todas as plantas (Young, 2000).

Conhecendo bem a estrutura dos genes de resistência, bem como quais são os domínios conservados e os motivos de cada um, é possível desenhar *primers* degenerados para a busca desses genes em uma determinada espécie.



RGAs podem ser amplificados por PCR e/ou usando informações genômicas e, então, mapeados geneticamente. Se um RGA se relaciona a um *locus* de resistência previamente definido, sua funcionalidade pode ser testada. Isso geralmente é feito para determinar se um gene candidato confere resistência quando expresso como um transgene em uma linhagem que é, normalmente, suscetível à doença (Hulbert, 2001).

Usando análises baseadas em PCR em busca de domínios conservados para a família NBS-LRR de genes de resistência, já foram fornecidos genes candidatos de resistência em diversas culturas como a batata (Leister *et al.*, 1996), soja (Kanazin *et al.*, 1996 ; Yu *et al.*, 1996; Peñuela *et al.*, 2002; He *et al.*, 2003), milho (Collins *et al.*, 1998), girassol (Gentzbittel *et al.*, 1998), alface (Shen *et al.*, 1998), *Brassica* (Joyeux *et al.*, 1999), arroz (Mago *et al.*, 1999), feijão (Rivkin *et al.*, 1999), citros (Deng *et al.*, 2000), café (Noir *et al.*, 2001), grão de bico (Huettel *et al.*, 2002), uva (Donald *et al.*, 2002), trigo (Lacock *et al.*, 2003), chicória (Plocik *et al.*, 2004) e sorgo (Totad *et al.*, 2005).

Lee *et al.* (2003), trabalhando com dois cultivares de maçã, *Malus domestica* cv. Fuji e *M. domestica* cv. Hong-ok e duas espécies selvagens, *M. prunifolia* e *M. baccata*, identificou 11 distintos RGAs com características da família NBS de genes de resistência, usando PCR e *primers* degenerados direcionados a motivos conservados em NBS-LRR.

De forma semelhante, He *et al.*, (2003) indentificaram e caracterizaram 61 RGAs em algodão (*Gossypium hirsutum* L.) por meio de PCR e *primers* degenerados.

### 1.6.1 Utilização de RGAs

#### 1.6.1.1 Seleção e análise de clones BAC e mapeamento físico

Clones BAC (*Bacterial Artificial Chromosome*) são fatores F modificados que podem clonar grandes fragmentos de DNA - insertos de até 500 kb de comprimento – e, conseqüentemente, permitem a exploração de genomas baseada em bibliotecas de grandes genomas (Shizuya *et al.*, 1992; Zhang *et al.*, 1996; Tomkins *et al.*, 1999). A utilização de clones BAC em substituição aos clones YAC (*Yeast Artificial Chromosome*) se deu pela eficácia de transformação, por não apresentarem quimerismos e nem rearranjos dos insertos (Marra *et al.*, 1997; Yüksel e Paterson, 2005).

A técnica apresenta várias finalidades e aplicações, entre elas o isolamento de genes de interesse, o mapeamento físico, a ligação entre mapas genéticos e físicos, a identificação de regiões cromossômicas de interesse e o seqüenciamento e compração de genomas. No caso da

utilização para o mapeamento físico, são gerados mapas genéticos que estão distribuídos pelo genoma. Estes mapas fornecem maneiras para clonagem posicional, já que na essência são como andar sobre os cromossomos. O mapeamento físico possibilita a localização cromossômica de genes de interesse, o que auxilia no processo de melhoramento assistido (Proite, 2009).

Nas bibliotecas BAC, cada clone é armazenado individualmente e ordenadamente. Já existem bibliotecas BAC para diversas espécies vegetais de importância econômica, como amendoim (Yüksel e Paterson, 2005); milho (Messing *et al.*, 2004); soja (Danesh *et al.*, 1998; Tomkins *et al.*, 1999); arroz (Ammiraju *et al.*, 2006); trigo (Akhunov *et al.*, 2005; Devos *et al.*, 2005); café (Noir *et al.*, 2004), entre vários outros.

Em arroz, vários RGAs foram identificados em bibliotecas BAC e tiveram o mapeamento feito. Um exemplo é o gene *RPR1*. Identificado primeiramente por meio de *primers* degenerados, este gene codifica uma proteína NBS-LRR e o seu mapeamento revela sua localização no cromossomo 11 (Song *et al.*, 1997).

#### 1.6.1.2 Análise de expressão gênica

A análise da expressão gênica diferencial é uma importante ferramenta molecular e pode contribuir para a caracterização de resistências, pois possibilita a identificação de genes-chave de redes pela comparação da expressão gênica durante o desenvolvimento de uma estrutura sob condições normais e em organismos carregando uma mutação ou submetidos a algum tipo de estresse (Guimarães *et al.*, 2005).

Para o entendimento da função de um gene, é fundamental o estudo de onde e quando ele é expresso. Recentemente, várias estratégias foram desenvolvidas para permitir o estudo da função de vários genes simultaneamente, entre elas: a genética reversa (geração de mutações específicas em genes de interesse), *screens* mutagênicos (geração de mutações randômicas e *screening* de um pool de mutantes para se identificar fenótipos de interesse) e bioinformática (a análise dos dados gerados por todas as estratégias acima mencionadas). As estratégias acima, associadas a projetos genoma, que têm como objetivo o seqüenciamento completo do genoma de vários organismos, têm possibilitado o estudo sistemático da expressão diferencial de genes. A análise da expressão diferencial de genes em resposta ao ataque de um patógeno constitui, portanto, ferramenta importante no isolamento de genes de resistência ou de fatores envolvidos com a interação patógeno-hospedeiro (Guimarães *et al.*, 2005).

Vários trabalhos tem buscado avaliar ESTs relacionadas a infecções por importantes patógenos. Ameline-Torregrosa *et al.* (2008), trabalhando com a leguminosa *Medicago truncatula* e os patógenos *Colletotrichum trifolii* e *Erysiphe pisi*, realizou uma análise BLAST de 25 seqüências de RGAs no banco de dados de ESTs de *M. truncatula*. Como resultado, foram encontradas ESTs para os RGAs identificados, com exceção de um.

#### 1.6.1.3 Marcadores moleculares e seleção assistida por marcadores

Segundo Milach (1998), marcadores moleculares são características de DNA que diferenciam dois ou mais indivíduos e são herdadas geneticamente. Os marcadores moleculares são, geralmente, herdados de forma codominante, não costumam exibir interações epistáticas ou pleiotrópicas, são detectados em tecidos jovens ou adultos e em qualquer fase do desenvolvimento da planta. A utilização de marcadores moleculares permite que a seleção e novos cruzamentos sejam realizados em uma mesma geração, o que aumenta consideravelmente a eficiência de um programa de melhoramento.

A seleção assistida por marcadores consiste na identificação precoce de plantas que apresentam fenótipo desejável pela análise molecular de seu genótipo. Três condições são necessárias para a utilização da seleção assistida por marcadores moleculares em programas de melhoramento. A primeira é que o marcador deve estar associado ao gene de interesse. A segunda é que a técnica deve ser baseada em PCR e ser eficiente o suficiente para fazer a avaliação de grandes populações. Por último, a técnica deve ser reproduzível, econômica e de fácil condução (Mohan *et al.*, 1997)

Neste contexto, em várias plantas, fragmentos de RGAs foram utilizados como marcadores moleculares para marcar *loci* de resistência a doenças. Alguns exemplos são *Arabidopsis thaliana* (Aarts *et al.*, 1998), melão (Mas *et al.*, 2001), caupi (Gowda *et al.*, 2002), *Lycopersicon* (Zhang *et al.*, 2002), feijão (Lopez *et al.*, 2003), cacau (Lanaud *et al.*, 2004) e algodão (Hinchliffe *et al.*, 2005).

Em arroz, mais de 15 genes e dezenas de *loci* de características quantitativas (QTL) associados à resistência foram identificados e localizados através do uso de marcadores moleculares como RFLP e AFLP (Song *et al.*, 2001). Alguns RGAs também foram clonados por PCR usando *primers* degenerados que foram projetados com base nas regiões conservadas dos genes de resistência e os resultados de estudos posteriores mostraram que alguns destes RGAs estão ligados firmemente a marcadores RFLP ou RAPD (Song *et al.*, 2001). Ramalingam *et al.* (2003) buscaram RGAs que pudessem ser utilizados como marcadores

moleculares para testar a associação de resistência em arroz para diversas doenças. Os 118 *loci* marcadores obtidos por PCR e RFLP. Agrupamentos de RGAs foram observados, principalmente no cromossomo 11, onde houve o reconhecimento de genes de resistência ao cretamento bacteriano e QTLs para queima das bainhas.

Em cana de açúcar (*Saccharum officinarum*), num trabalho feito por Rossi *et al.*, (2003), 88 RGAs expressos foram agrupados em *clusters*. Cinquenta RGAs, quando utilizados como marcadores RFLP, mostraram um total de 148 *loci* polimórficos *single-dose*. Um *cluster* de dois *loci* contendo RGAs NBS-LRR foi identificado como pertencente ao único gene de resistência conhecido nesta cultura e que confere resistência à ferrugem da cana.

Em milho, *Zea mays*, Collins *et al.* (1998) identificaram 11 NBS-LRR RGAs e a partir deles e de outro RGA obtido de trigo mapearam 20 *loci* por RFLP. A utilização destes RGAs como sondas possibilitou a identificação de regiões genômicas que possuem genes de resistência a vírus e fungos. Liu *et al.* (2011) desenvolveram conjuntos de marcadores moleculares que usam o polimorfismo de comprimento de íntrons em RGAs em milho. Foram selecionados 632 RGAs que possuíam ao menos um íntron que pudesse servir como marcador putativo para o polimorfismo do comprimento do íntron. Por meio de e-PCR foram desenvolvidos 1972 marcadores candidatos com pares de *primers* desenhados para regiões exônicas que ladeiam o íntron. Como resultado, 46,4% dos candidatos a marcadores possuíam a capacidade de discriminar polimorfismo.

## 1.7 Melhoramento e recursos genômicos em *Elaeis*

O interesse em germoplasma de *E. oleifera* é devido a algumas características como crescimento lento, qualidade do óleo e resistência a doenças, incluindo amarelecimento fatal, o principal problema de cultivo na América (Bergamin Filho *et al.*, 1997), bem como a murcha de *Fusarium*, a principal doença da cultura na África (Renard *et al.*, 1980;. Hardon *et al.*, 1985).

Devido a sua menor produtividade, o cultivo de caiaué ainda não é economicamente viável. No entanto, híbridos interespecíficos das duas espécies do gênero podem ser obtidos com rendimentos em torno de 90% da produção comercial da cultivar comercial, mas geralmente mostrando esterilidade parcial (Amblard *et al.*, 1995).

Os híbridos interespecíficos geralmente apresentam características intermediárias aos dois genitores (Tabela 2), revelando assim a ação predominantemente aditiva dos genes.

Dentre essas características, enfatiza-se a menor taxa de crescimento de tronco, o que aumentaria o período de exploração econômica dos plantios, produção de óleo com teor de ácidos graxos insaturados superiores ao do dendezeiro, sendo, portanto, mais líquido nas condições ambientais e apresentando-se mais apropriado ao consumo alimentar. Além de aspectos relacionados à resistência a doenças e pragas, às quais o dendezeiro é suscetível (Barcelos, 1986).

**Tabela 2:** Valores de caracteres agronômicos de *E. guineensis*, *E. oleifera* e híbrido interespecífico.

Característica	Unidade	<i>E. guineensis</i>	<i>E. oleifera</i>	Híbrido F1
Crescimento do estipe	cm/ano	30 - 75	5	15 - 25
Amarelecimento fatal	Mortalidade (%)	75	0	≤ 1
Fusariose	Índice de incidência	58 - 141	0 - 400	0 - 150
Grau de insaturação do óleo	%	40 - 60	60 - 83	62 - 69
Teor de óleo na polpa seca	%	67 - 76	35 - 49	59 - 68
Teor de óleo no cacho	%	18,3 - 25,5	1,7 - 4,4	3,8 - 17,0
Taxa de extração de óleo	%	20 - 24	≤ 9	8,9 - 18,8

Fonte: Chia, 2008

Apesar das características de grande interesse agrônômico dos híbridos, estes apresentam um aspecto limitante para sua exploração em plantações comerciais; a baixa produtividade em óleo, conseqüência de uma baixa relação óleo/mesocarpo e frutos normais/cacho. Ensaio conduzidos no Benim, Costa do Marfim e Indonésia demonstraram que as melhores progênies híbridas eram inferiores de 10 a 15% em relação aos materiais comerciais de dendezeiro. Além da baixa produção em óleo, existem problemas de fertilidade polínica que levam à necessidade da realização de polinizações assistidas para boa frutificação, sobretudo em áreas em que não existam dendezaís nas proximidades (Barcelos e Amblard, 1992; Amblard *et al.*, 1995).

A busca de RGAs em *Elaeis* conta, até o momento, com apenas 4 seqüências depositadas no Genbank. Há, portanto, grande necessidade em aumentar o número de pesquisas que visem achar estes genes análogos de resistência.

Ho *et al.* (2007) construíram seis bibliotecas de cDNA a partir de embriões zigóticos de dendê, células em suspensão, meristemas apicais, flores e raízes. Foram gerados um total de 14.537 (ESTs). Várias outras iniciativas buscando a montagem de bibliotecas de cDNA e ESTs tem sido relatadas e estão em andamento.

A primeira biblioteca BAC de dendê foi construída usando a enzima *Hind III* em *E. guineensis*. Tentativas preliminares também têm sido reportadas para construção de uma biblioteca BAC na Malásia, com estimativa de cobertura de aproximadamente 2 vezes o tamanho do genoma de *E. guineensis*. Entretanto, nenhuma iniciativa, até o momento, foi feita para o desenvolvimento de uma biblioteca BAC para *E. oleifera*. Para dendê, 8.000 clones BAC, de 120 kb cada, equivalem a um genoma haplóide. O desenvolvimento de mapas físicos para plantas de interesse comercial, que na sua maioria apresentam grandes genomas, tem uma importância fundamental, pois facilitam novas formas de pesquisa na sua genética e melhoramento. Uma vez gerado, os mapas físicos poderão prover virtualmente um número ilimitado de marcadores de DNA para realização de gene *tagging*, manipulação de genes e outros estudos genéticos.

Estudos sobre marcadores moleculares em *Elaeis* têm focado principalmente cultivares comerciais de *E. guineensis*, apesar de haver algumas pesquisas já relatadas em *E. oleifera* (Moretzsohn *et al.*, 2002; Singh *et al.*, 2008). Marcadores microssatélites de *E. oleifera* tem sido desenvolvidos com o objetivo de caracterizar coleções de germoplasma provenientes da Colômbia, Costa Rica, Panamá e Honduras. No total, dez marcadores microssatélites, a maioria deles apresentando transferibilidade para *E. guineensis*, foram desenvolvidos e caracterizados a partir de 1500 seqüências da biblioteca genômica *E. oleifera*.

Os sistemas de transformação genética utilizados mais rotineiramente (transformação mediada por *Agrobacterium* e o sistema de biobalística) têm sido empregados com sucesso na transformação de espécies monocotiledôneas. Para *E. guineensis*, o processo biobalístico já foi utilizado para a produção de plantas transgênicas onde foram testados os genes *GSU* de resistência a antibióticos e o gene *GFP* que codifica uma proteína fluorescente (Parveez *et al.*, 2000).

## 2. OBJETIVOS

### 2.1 Objetivo geral

Caracterizar análogos a genes de resistência (RGAs) em diferentes materiais de *E. guineensis* e *E. oleifera* contrastantes em resistência à *Fusarium oxysporum* f.sp. *elaedis*, buscando, de forma preventiva, fontes de resistência no gênero.

### 2.2 Objetivos específicos

Amplificar de forma específica por PCR, RGAs em materiais de *E. guineensis* e *E. oleifera* contrastantes para resistência a *Fusarium oxysporum* f.sp. *elaedis* utilizando primers degenerados desenhados a partir da família NBS-LRR de genes de resistência;

Analisar seqüências de DNA por meio de técnicas de bioinformática e identificar seqüências com similaridade a genes de resistência em bancos de dados de nucleotídeos e proteínas;

Determinar a diversidade de RGAs em *E. guineensis* e *E. oleifera* por meio da análise filogenética das seqüências obtidas.

### 3. MATERIAL E MÉTODOS

#### 3.1 Seleção de material de *Elaeis*

Foram selecionadas acessos de *E. guineensis* e *E. oleifera* contrastantes em resistência a *F. oxysporum* f.sp. *elaeidis*. No total, oito plantas, sendo duas cultivares de *E. oleifera* e seis de *E. guineensis* foram utilizadas. A Tabela 3 mostra as espécies utilizadas, acessos e informações sobre a resistência da planta a *F. oxysporum* f.sp. *elaeidis* e sua origem. Todos os acessos utilizados fazem parte do programa de melhoramento de *Elaeis* sp. da Embrapa Amazônia Ocidental e provém de diferentes regiões do continente africano, centro de origem de *E. guineensis*, e amazônicas, centro de origem de *E. oleifera*.

**Tabela 3:** Espécie, acesso e resistência do material vegetal utilizado

<b>Espécie</b>	<b>Acesso</b>	<b>Resistência a <i>F. oxysporum</i> f.sp. <i>elaeidis</i></b>
<i>E. oleifera</i>	Manicoré	Resistente (Centro de origem)
<i>E. oleifera</i>	Tefé	Resistente (Centro de origem)
<i>E. guineensis</i>	Nigéria	Resistente (Centro de origem)
<i>E. guineensis</i>	LM2T	Resistente (Centro de origem)
<i>E. guineensis</i>	Bahia 1	Não informado
<i>E. guineensis</i>	C 10	Resistente
<i>E. guineensis</i>	C 23	Moderadamente resistente
<i>E. guineensis</i>	C 25	Suscetível

Os acessos Manicoré e Tefé, de *E. oleifera*, são provenientes da região amazônica, de municípios com o mesmo nome, provável centro de origem da espécie. O acesso Nigéria provém do provável centro de origem de *E. guineensis*. Estes acessos, por terem co-evoluído com o patógeno, são resistentes. LM2T é proveniente de La Mé, Costa do Marfim, e apresenta resistência a *Fusarium*. Bahia 1 não possui informações quanto a resistência ao patógeno e é proveniente do estado da Bahia. Os acessos C10, resultante do cruzamento DA115D (Deli) x LM2T (La Mé), C23, resultante do cruzamento [LM269D (La Mé) x DA115D (Deli)] x LM2T, e C25, resultante do cruzamento [DA3D (Deli) x D5D (Deli)] x LM2T, são, respectivamente, resistentes, moderadamente resistentes e suscetíveis.

Todas as análises laboratoriais foram feitas nas dependências na Universidade Católica de Brasília, Campus II.

#### 3.2 Extração do DNA genômico de *Elaeis*

O DNA genômico foi extraído de folhas congeladas, armazenadas em freezer a -80°C, segundo o protocolo modificado de CTAB (Grattapaglia e Sederoff, 1994). Tiras de folhas de



150 mg foram maceradas em tubos de microcentrífuga de 1,5 mL com o auxílio de nitrogênio líquido. Foi adicionado ao tecido recém macerado 1 mL de tampão de extração CTAB (brometo de cetil trimetil amônio – CTAB 5%; NaCl 5M; Tris-HCl 1M pH 8,0; Água destilada; 2-mercaptoetanol; Polivilpirrolidona), solução na qual foi ressuspendido. Os tubos foram incubados em banho-maria a uma temperatura de 65°C por 30 minutos. Foi feita a primeira extração com solvente orgânico adicionando 600 µl de uma solução de clorofórmio e álcool isoamílico na proporção de 24:1. Os tubos foram agitados em vórtex, homogeneizados e centrifugados em microcentrífuga a velocidade de 12000 rpm durante 5 minutos. A fase superior, aquosa, foi transferida para um novo microtubo de 1,5 mL e à ela foi adicionado 1/10 de seu volume de uma solução 10% CTAB, 1,4 M NaCl. Foi feita uma nova extração com o mesmo solvente orgânico e nova centrifugação a velocidade de 12000 rpm durante 5 minutos. Novamente retirou-se a fase aquosa superior, a qual foi transferida para um novo microtubo de 1,5 mL. Foi adicionada à amostra uma solução de isopropanol a -20°C em um volume de 2/3 do total do extrato. Deixou-se precipitar o DNA *overnight* a 4°C. Os tubos foram centrifugados a 7500 rpm durante 5 minutos para a formação de um pellet. Foi descartado o sobrenadante e o pellet foi lavado duas vezes em 1 mL de etanol 70%. Após centrifugação a 7500 rpm durante 3 minutos, o pellet foi lavado uma vez em 1 mL de etanol 95%. Retirou-se o etanol e o pellet foi ressuspendido em 50 mL de tampão TE (Tris-HCl 10 mM pH 8,0; EDTA 1 mM) com RNase (10 µg/mL). O DNA genômico total foi quantificado utilizando o espectrofotômetro Nanodrop ND-1000 (Thermo Scientific) e também em gel de agarose 1,5%, corado com 5 µg/mL de brometo de etídeo a 0,15 µg/mL, utilizando como marcador 4 µL de Low DNA Mass Ladder (Invitrogen, EUA). Devido a grande quantidade de RNA visualizado no gel de agarose, todos os DNAs foram tratados com 2 µL de TE + RNase (10 µg/mL) e novamente quantificados. Todas as amostras de DNA foram ajustadas para uma concentração final de 20 ng/µL.

### 3.3 Amplificação de RGAs por PCR

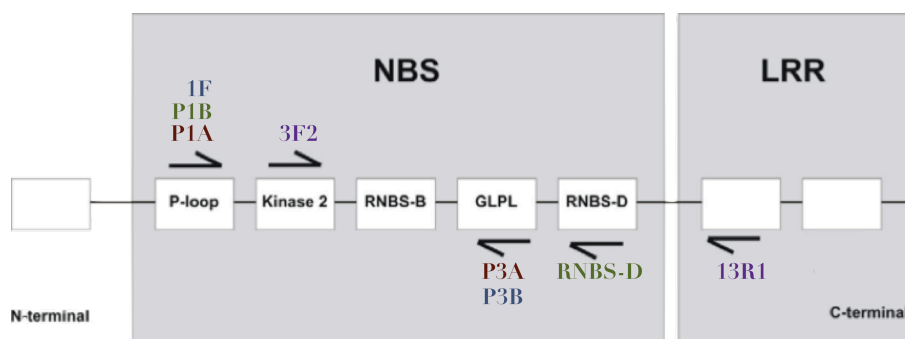
Foram testadas quatro combinações de *primers* degenerados com o objetivo de obter produtos pertencentes à família NBS-LRR de genes de resistência. A primeira combinação, denominada combinação A, foi formada pelos *primers* P1A e P3A. A combinação B foi formado pelos *primers* P1B e RNBS-D, a combinação C, pelos *primers* 1F e P3B e combinação D, pelos *primers* 3F2 e 13R1. A Tabela 4 mostra a sequência, o motivo no qual há anelamento e o autor do *primer*. A Figura 7 mostra um esquema dos motivos conservados

e a região de anelamento dos *primers*.

**Tabela 4:** Seqüências dos *primers* degenerados usados na amplificação de RGAs em *Elaeis* e seus motivos alvo

Comb.	Primer degenerado	Seqüência do primer (5'a 3')*	Motivo protéico alvo	Grupo de origem	Autor
A	P1A (forward)	GGIATGCCIGGIIIIIGGIAARACIAC	P-loop	Dicotiledônea	Bertioli <i>et al.</i> , 2003
	P3A (reverse)	AIITYIRIIRYIAGIGGYAAICC	GLPL	Dicotiledônea	Bertioli <i>et al.</i> , 2003
B	P1B (forward)	GGIATGGGIGGIIIIIGGIAARACIAC	P-loop	Dicotiledônea	Bertioli <i>et al.</i> , 2003
	RNBS-D (reverse)	GGRAAIARISHRCARTAIIRAARC	RNBS-D não-TIR	Mono e Dicotiledônea	Penuela <i>et al.</i> , 2002
C	1F (forward)	GGCGGGGTGGGCaaracnacnht	P-loop	Dicotiledônea	Miller <i>et al.</i> , 2008
	P3B (reverse)	AIITYIRIIRYIAGIGGIAGICC	GLPL	Dicotiledônea	Miller <i>et al.</i> , 2008
D	3F2 (forward)	GAGGTACTTCCTGGTGCTGgayayrtbtgg	Quinase 2	Monocotiledônea	Miller <i>et al.</i> , 2008
	13R1 (reverse)	CGGCCAAGTCGTGCAyvakrtcrtgca	LRR	Monocotiledônea	Miller <i>et al.</i> , 2008

\*Código degenerado de bases: I = inosine; R = A/G; Y = C/T; M = A/C; K = G/T; W = A/T; S = C/G; B = C/G/T; D = A/G/T; H = A/C/T; V = A/C/G; N = A/C/ G/T



Fonte: Emediato, 2009

**Figura 7:** Motivos conservados em domínios NBS e LRR alvos de *primers* degenerados. As setas mostram as posições aproximadas de anelamento dos *primers*.

As condições de PCR foram otimizadas em termos de temperatura de anelamento de *primer* e concentração de reagentes. As reações foram realizadas conforme a Tabela 5. As reações foram realizadas com volume final de 50 µL para a combinação B, a fim de conseguir maior concentração de DNA, e 25 µL para as combinações A, C e D, contendo, para cada combinação, 20 ng/µL de DNA, 2,6 mM de MgCl<sub>2</sub>, 0,2 mM de dNTPs, 1 µM do primer a ser utilizado, 5 U/µL de DNA polimerase *Taq* Platinum (Phoneutria, Belo Horizonte, MG), e 10X de tampão da *Taq* Platinum (Phoneutria, Belo Horizonte, MG). As temperaturas dos ciclos de PCR foram de acordo com o seguinte programa: 94°C por 3 minutos; 35 ciclos de 94°C por 1 minuto, 48°C por 30 segundos – para a combinação B – e 50°C por 30 segundos – para a combinação D –, 72°C por 1 minuto; e 10 minutos para extensão final a 72°C.

**Tabela 5:** Reagentes e concentrações utilizados para a amplificação de RGAs NBS-LRR por PCR

Reagente	Comb. B (µL)	Comb. D (µL)
Tampão 10x	5,0	2,5
MgCl <sub>2</sub> (2,6 mM)	1,0	0,75
DNTPs (0,2mM)	4,0	2,0
Primer forward (1µM)	4,0	2,5
Primer reverse (1µM)	4,0	2,5
Taq polimerase (5U/µL)	0,5	0,4
DNA (20ng/µL)	5,0	2,0
Água (Milli-Q)	26,5	12,35
Volume Total	50,0	25,0

As combinações A e C, tratamentos 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 e 31, foram descartadas, pois se mostraram inespecíficas. A Tabela 6 mostra a combinação de *primers* utilizada por espécie e cultivar, conferindo siglas para cada um dos tratamentos.

**Tabela 6:** Nomes dos tratamentos baseados em espécie, acesso e combinação de *primers* degenerados

Tratamento	Espécie	Acesso	Combinação de <i>primers</i>	Sigla
1	<i>E. oleifera</i>	Manicoré	P1A / P3A	EO_MAN_P1A/P3A_CO
2	<i>E. oleifera</i>	Manicoré	P1B / RNBS-D	EO_MAN_3F2/13R1_CO
3	<i>E. oleifera</i>	Manicoré	1F / P3B	EO_MAN_1F/P3B_CO
4	<i>E. oleifera</i>	Manicoré	3F2 / 13R1	EO_MAN_3F2/13R1_CO
5	<i>E. oleifera</i>	Tefé	P1A/P3A	EO_TEF_P1A/P3A_CO
6	<i>E. oleifera</i>	Tefé	P1B / RNBS-D	EO_TEF_P1B/RNBSD_CO
7	<i>E. oleifera</i>	Tefé	1F / P3B	EO_TEF_1F/P3B_CO
8	<i>E. oleifera</i>	Tefé	3F2 / 13R1	EO_TEF_3F2/13R1_CO
9	<i>E. guineensis</i>	Nigéria	P1A / P3A	EG_NIG_P1A/P3A_CO
10	<i>E. guineensis</i>	Nigéria	P1B / RNBS-D	EG_NIG_P1B/RNBSD_CO
11	<i>E. guineensis</i>	Nigéria	1F / P3B	EG_NIG_1F/P3B_CO
12	<i>E. guineensis</i>	Nigéria	3F2 / 13R1	EG_NIG_3F2/13R1_CO
13	<i>E. guineensis</i>	LM2T	P1A / P3A	EG_LM2T_P1A/P3A_R
14	<i>E. guineensis</i>	LM2T	P1B / RNBS-D	EG_LM2T_P1B/RNBSD_R
15	<i>E. guineensis</i>	LM2T	1F / P3B	EG_LM2T_1F/P3B_R
16	<i>E. guineensis</i>	LM2T	3F2 / 13R1	EG_LM2T_3F2/13R1_R
17	<i>E. guineensis</i>	Bahia 1	P1A/P3A	EG_BA1_P1A/P3A_NI
18	<i>E. guineensis</i>	Bahia 1	P1B / RNBS-D	EG_BA1_P1B/RNBSD_NI
19	<i>E. guineensis</i>	Bahia 1	1F / P3B	EG_BA1_1F/P3B_NI
20	<i>E. guineensis</i>	Bahia 1	3F2 / 13R1	EG_BA1_3F2/13R1_NI
21	<i>E. guineensis</i>	C 10	P1A / P3A	EG_C10_P1A/P3A_R
22	<i>E. guineensis</i>	C 10	P1B / RNBS-D	EG_C10_P1B/RNBSD_R
23	<i>E. guineensis</i>	C 10	1F / P3B	EG_C10_1F/P3B_R
24	<i>E. guineensis</i>	C 10	3F2 / 13R1	EG_C10_3F2/13R1_R
25	<i>E. guineensis</i>	C 23	P1A/P3A	EG_C23_P1A/P3A_MR
26	<i>E. guineensis</i>	C 23	P1B / RNBS-D	EG_C23_P1B/RNBSD_MR
27	<i>E. guineensis</i>	C 23	1F / P3B	EG_C23_1F/P3B_MR
28	<i>E. guineensis</i>	C 23	3F2 / 13R1	EG_C23_3F2/13R1_MR
29	<i>E. guineensis</i>	C 25	P1A / P3A	EG_C25_P1A/P3A_S
30	<i>E. guineensis</i>	C 25	P1B / RNBS-D	EG_C25_P1B/RNBSD_S
31	<i>E. guineensis</i>	C 25	1F / P3B	EG_C25_1F/P3B_S
32	<i>E. guineensis</i>	C 25	3F2 / 13R1	EG_C25_3F2/13R1_S

Seguido da amplificação, os produtos de PCR foram visualizados sob luz U.V. em géis de agarose 1.5% corados com 5  $\mu$ L de brometo de etídio a 0,15  $\mu$ g/mL. Os marcadores de peso molecular 1Kb Plus DNA Ladder (Invitrogen, EUA) e Low DNA Mass Ladder (Invitrogen, EUA) foram incluídos em todos os géis (2  $\mu$ L de cada).

### 3.4 Purificação e clonagem de produtos de PCR

Os produtos de PCR que possuíam o tamanho esperado de pares de base, aproximadamente 650 pb para a combinação P1B/RNBS-D e 700 pb para a combinação 3F2/13R1, foram purificados com o Qiagen QIAquick PCR purification kit (Qiagen, Valencia, Ca, E.U.A.). Posteriormente foram ligados em vetor pGEM-T Easy (Promega, Madison, WI, EUA), atendendo as condições descritas pelo kit.

A diálise foi feita para a retirada do excesso de sal. Foram alicotadas as amostras em membranas Millipore VSWP02500 (0,025  $\mu$ m) sobre água *Milli-Q* autoclavada. Deixou-se a diálise ocorrer por 40 minutos.

Para a confecção de células competentes, uma colônia de *Escherichia coli* DH5 $\alpha$  foi colocada em 5 mL de meio LB (Luria-Bertani Broth) líquido sem antibiótico em tubo falcon de 15 mL e deixada no *shaker overnight* a 37°C e 250 rpm. 1 mL deste pré-inóculo foi incubado em frasco Erlenmeyer contendo 199 mL de meio LB líquido durante, aproximadamente, 5 horas (colocado em *shaker* a 37°C e 250 rpm) para o crescimento da bactéria. Os 200 mL do inóculo foram distribuídos em 4 tubos Falcon de 50 mL gelados (50 mL em cada) e colocado por 30 minutos no gelo. Após este período, os tubos foram centrifugados a 3660 rpm a 12°C por 25 minutos. O sobrenadante foi retirado e 50 mL de água destilada autoclavada foram acrescentados aos tubos. As amostras foram centrifugadas a 4226 rpm por 20 minutos a 12°C e, posteriormente, o sobrenadante foi descartado. O *pellet* foi ressuscitado em 25 mL de glicerol 10% gelado e o conteúdo dos quatro tubos foram juntados dois a dois. Foi feita uma nova centrifugação a 4226 rpm por 15 minutos a 12°C, com o posterior descarte do sobrenadante. Adicionou-se 15 mL de glicerol 10% gelado e o conteúdo dos dois tubos foram transferidos para um só. Este tubo foi centrifugado a 4226 rpm por 15 minutos a 12°C e o sobrenadante foi descartado. O *pellet* foi ressuscitado em solução GYT (10% de glicerol; 0,125% de extrato de levedura e 0,25% de triptona) e distribuídos em tubos *ependorf* (50  $\mu$ L).

As células competentes de *E. coli* DH5 $\alpha$  foram transformadas por eletroporação utilizando o GenePulser II (Bio-Rad, Hercules, CA, E.U.A.) fixado em 2,5 Kv e 200 W. Um

volume de 3  $\mu\text{L}$  de ligação foram adicionados a um tubo *ependorf* contendo 50  $\mu\text{L}$  de célula competente. A suspensão foi transferida para uma cuveta e imediatamente levada ao eletroporador. Todo o procedimento foi feito com o material mantido no gelo. Após a eletroporação, foi acrescentado 1 mL de meio SOC (Super Optimal Broth) na suspensão, sendo imediatamente incubada por 1 hora a 37°C. Em seguida, 100 e 200  $\mu\text{L}$  do transformado foram plaqueados em meio LB Agar contendo 40  $\mu\text{L}$  do substrato de Xgal (20 mg/mL) (5-bromo-4-cloro-indolil-galactopiranosida), 4  $\mu\text{L}$  do indutor do gene Lac, IPTG (0,1M) (isopropil  $\beta$ -D-1-tiogalactopiranosida), e 500  $\mu\text{L}$  de ampicilina (100  $\mu\text{g}/\text{mL}$ ). O material foi incubado a 37°C *overnight*.

### 3.5 Extração de DNA plasmidial

Para a obtenção do pré-inóculo, colônias brancas de bactérias recombinantes foram selecionadas manualmente com até 96 clones preservados por combinação de *primer* e acesso. Os transformantes foram inoculados em placa ELISA (Corning-costar 3799) com 130  $\mu\text{L}$  de meio Circle Grow (CG) com ampicilina (100  $\mu\text{g}/\text{mL}$ ) e incubados a 37°C por 18 horas. Para a armazenagem das placas, 50  $\mu\text{L}$  de glicerol 20% foram adicionados a cada pré-inóculo.

A produção do inóculo foi feita utilizando placas Deep Well (Sorenson 23130). 30  $\mu\text{L}$  do pré-inóculo foram alicotados em 1 mL de meio CG com ampicilina (100  $\mu\text{g}/\text{mL}$ ). O inóculo foi incubado a 37°C, 275 rpm, por cerca de 18 horas.

Para a extração de DNA plasmidial foi utilizado o método adaptado de lise alcalina (Sambrook, *et al.*, 2001). Verificada a uniformidade do crescimento do inóculo, as placas foram centrifugadas a 2080 rcf por 6 minutos. O sobrenadante foi removido e 240  $\mu\text{L}$  de solução GET (Glucose 50 mM, EDTA 10 mM, Tris-HCl 25 mM, pH 8), mantida a 4°C, foi adicionada em cada poço. As células foram ressuspensas e a placa foi novamente centrifugada a 2080 rcf por 6 minutos. O sobrenadante foi retirado e 80  $\mu\text{L}$  de GET foram adicionados a cada poço. Após a ressuspensão das células, 60  $\mu\text{L}$  da solução foram transferidos para uma placa de depolipropileno de fundo em U (Conrning-costar 3365) contendo 2  $\mu\text{L}$  de RNase (10 mg/mL). 60  $\mu\text{L}$  de uma solução de lise contendo NaOH 0,2 M e SDS 1% foram adicionados em cada poço. A placa foi misturada por inversão por 20 vezes e incubada a temperatura ambiente por 7 minutos. 60  $\mu\text{L}$  da solução neutralizadora de KOAc foram adicionados em cada poço. As placas foram incubadas por 7 minutos a temperatura ambiente e em seguida centrifugadas por 12 minutos a 2080 rcf. O sobrenadante foi filtrado em uma placa Millipore (MAGV N22) fixada a uma microplaca depolipropileno de fundo em

V (Corning-costar 3363). Esse conjunto foi centrifugado a 2080 rcf por 6 minutos. A placa Millipore foi removida e 90 µL de isopropanol 100% foram adicionados em cada poço com o filtrado para a precipitação do DNA. A placa foi, então, homogeneizada e centrifugada a 2080 rcf para 45 minutos. Após este processo, o sobrenadante foi descartado e o precipitado foi lavado com 200 µL de etanol 70% a 4°C. As placas foram centrifugadas por 5 minutos a 2080 rcf. O sobrenadante foi, mais uma vez, descartado e o *pellet* foi deixado em secagem a temperatura ambiente por aproximadamente 2 horas e em seguida ressuspendido em 40 µL de água *Milli-Q* autoclavada. O DNA plasmidial foi visualizado sob luz U.V. em géis de agarose 1,5% corados com brometo de etídio a 0,15 µg/mL. Foram incluídos em todos os géis 2 µL do marcador de peso molecular Low DNA Mass Ladder (Invitrogen, EUA). As placas foram armazenadas a -4°C.

### 3.6 Seqüenciamento de clones recombinantes

Os insertos de DNA plasmidial de 768 clones, dos 1536 obtidos no total, foram seqüenciados unidirecionalmente no seqüenciador ABI 3130xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA), sendo 48 clones por tratamento.

Para o seqüenciamento utilizou-se 4 µL de amostra, 1 µL de água *Milli-Q*, 2 µL do *primer* M13 *reverse* e 3 µL de BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA).

### 3.7 Análise de bioinformática das seqüências

O processamento das seqüências de DNA foi realizado na plataforma SISGEN (Pappas *et al.*, 2008), pelos programas Cross-Match (<http://www.phrap.org/phredphrap/general.html>), SeqClean (Chen *et al.*, 2007), RepeatMasker (<http://www.repeatmasker.org>) e Lucy (Chou & Holmes, 2001).

Os dados gerados receberam tratamentos de mascaramento de vetor e regiões repetitivas, retirada de regiões de baixa complexidade, bem como o tratamento de qualidade com cortes de cauda PoliA/T. Os parâmetros utilizados para o controle de qualidade foram PHRED 15 e extensão mínima de 150 pb.

As seqüências pré-processadas foram montadas por tratamento, as redundâncias foram retiradas e foi desenvolvido um controle responsável pela diferenciação de polimorfismos e prováveis erros de seqüenciamento. Esse procedimento foi realizado com o auxílio dos

programas PHRAP v. 1.080812 e WCD (Hazelhurst *et al.*, 2008). Posteriormente as seqüências foram submetidas ao programa AMOS (<http://amos.sourceforge.net>) onde sofreram marcações de regiões que possuíam mais de duas assinaturas de *mis-assembly*, isto é, regiões com baixa cobertura, alta cobertura ou alta incidência de polimorfismos. As seqüências marcadas foram excluídas.

Após todos os cuidados relacionados à qualidade das seqüências, foi realizado o *clustering* utilizando o montador genômico MIRA v. 3RC4 (Chevreux, 2004).

A geração de uma biblioteca com informações protéicas de viridiplantae (grupo monofilético que compreende algas verdes e plantas terrestres) foi desenvolvida por meio do HMMER (Oliver *et al.*, 2008) utilizando dados de domínios protéicos originários do Pfam. As predições de genes, pseudogenes e domínios conservados foram feitos pelo programa Genewise (Birney *et al.*, 2004) utilizando o *data set* disponibilizado pelo HMMER. A confirmação de predição gênica foi feita pelo fGENESH (Slovyev *et al.*, 1994) e AUGUSTUS (Stanke *et al.*, 2004). Foi feita, então, a tradução em sequencia protéica pelo transeq, tradutor pertencente ao pacote emboss (Rice *et al.*, 2000).

As seqüências protéicas foram analisadas utilizando o programa Blastx (Altschul *et al.*, 1997) contra o banco de dados proteínas NR do Genbank. Uma análise de similaridade entre seqüências protéicas com NB-ARC R-genes e RGAs de *Arabidopsis thaliana* (L.) Heynh foi conduzido utilizando-se o programa BLASTx e o software BioLayout Express<sup>3D</sup> (<http://www.biobioinformatics.org/>). Foram utilizadas, também, as quatro seqüências de referencia de NBS-LRR RGAs de *E. guineensis* disponíveis no GenBank (novembro de 2011). O BioLayout é um programa que utiliza um algoritmo para a visualização biológica e que pode gerar, com base na similaridade entre seqüências proteicas, uma representação 2D ou 3D em forma de gráfico de rede (Enright e Ouzounis, 2001). Este tipo de apresentação de dados facilita a interpretação de características importantes dentro de grupos com alguma similaridade. Resultados de dados Blastx são convertidos em uma pontuação ponderada na forma de um E-value de Blast.

Uma análise filogenética entre RGAs NBS-LRR de *E. guineensis* e *E. oleifera* foi feito pelo método *Neighbor Joining*. O alinhamento múltiplo das seqüências de DNA foi feito por meio do ClustalW (Thompson *et al.*, 1997) e a edição das seqüências utilizadas no alinhamento foi feito utilizando o programa Jalview 2 (Waterhouse *et al.*, 2009).

## 4. RESULTADOS

### 4.1 Desenho experimental e coleta de material vegetal

O objetivo dos experimentos realizados e baseados em PCR, clonagem e seqüenciamento, foi o de maximizar a possibilidade de caracterização de diversos RGAs NBS-LRR no gênero *Elaeis*, por meio da avaliação das duas espécies cultivadas comercialmente, *E. guineensis* e *E. oleifera*, explorando a diversidade genética de material proveniente dos centros de origem de cada espécie, bem como observando molecularmente dados que permitam contrastar os diferentes acessos na característica resistência à murcha de *Fusarium*.

### 4.2 Extração e quantificação do DNA genômico

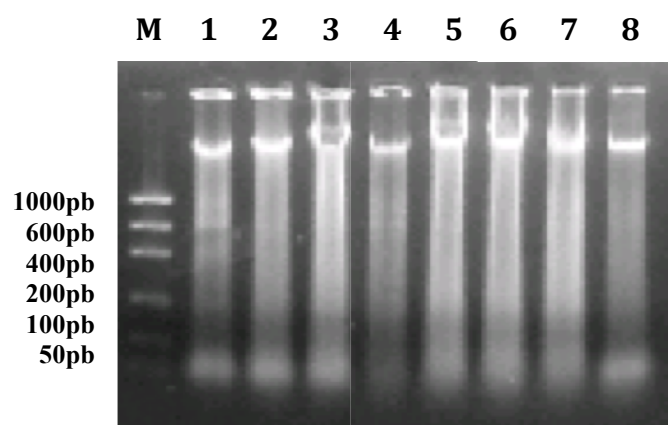
O material foliar serviu como uma fonte conveniente de extração de DNA genômico de alta qualidade. O DNA genômico total foi quantificado em gel de agarose 1,5%, corado com o corante quantitativo brometo de etídeo (Figura 8) e também por meio de espectrofotometria (NanoDrop – ND-1000) (Tabela 7).

O DNA genômico, assim como muitos contaminantes comuns encontrados em extratos de DNA, possuem absorvâncias entre 230nm e 320nm. O DNA de boa qualidade deve apresentar uma taxa de A260/A280 de 1,7 a 2,0 e uma taxa de A260/A230 maior que 1,5. A desvantagem do método é que também são feitas medições espectrofotométricas de contaminantes como RNAs e proteínas co-precipitadas, que exibem absorvância a 260 nm. Isto pode contribuir para o aumento da leitura A260 e uma superestimação da concentração de DNA.

Na eletroforese em gel de agarose e coloração com brometo de etídeo, pelo contrário, só as medidas de concentração de DNA são determinadas, já que este método não é afetado por esses contaminantes. A concentração é calculada por comparação da intensidade da banda da amostra com a de uma banda de um marcador de peso molecular cujo conteúdo de DNA é conhecido.

Como esperado, uma diferença entre as duas quantificações foi observada, sendo que as medidas feitas pelo Nanodrop apresentaram uma quantidade muito maior em relação a quantificação feita em gel de agarose.





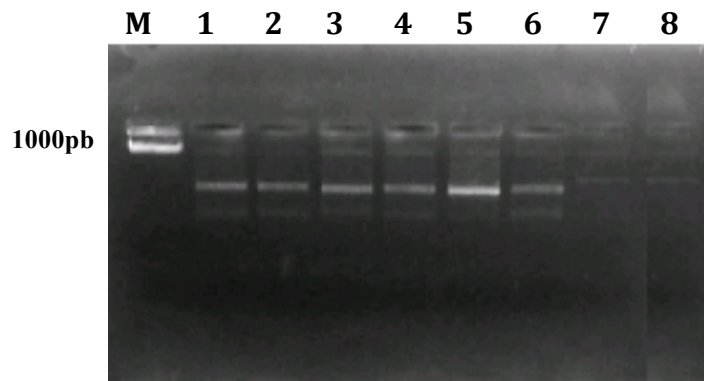
**Figura 8:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2  $\mu$ l dos DNAs genômicos totais de *E. oleifera* e *E. guineensis*. Usou-se como marcador para quantificação 4  $\mu$ l de *Low DNA Mass Ladder*. Amostras no gel: (M) Marcador de peso molecular (*Low DNA Mass Ladder*); (1) *E. oleifera*, acesso Manicoré; (2) *E. oleifera*, acesso Tefé; (3) *E. guineensis*, acesso Nigéria; (4) *E. guineensis*, acesso LM2T; (5) *E. guineensis*, acesso Bahia 1; (6) *E. guineensis*, acesso C10; (7) *E. guineensis*, acesso C23; (8) *E. guineensis*, acesso C25).

**Tabela 7:** Quantificação em ng/ $\mu$ l de DNA total dos acessos de *E. oleifera* e *E. guineensis* com os respectivos valores de absorvância e razões medidos pelo Nanodrop ND-1000.

Amostra	Espécie	Acesso	Concentração gel (ng/ $\mu$ l)	Concentração Nanodrop (ng/ $\mu$ l)	A260	A280	260/280	260/230
1	<i>E. oleifera</i>	Manicoré	50	666,05	13,32	7,71	1,73	1,24
2	<i>E. oleifera</i>	Tefé	100	1111,83	22,23	13,63	1,63	1,15
3	<i>E. guineensis</i>	Nigéria	50	3285,36	65,70	48,77	1,35	1,03
4	<i>E. guineensis</i>	LM2T	30	1675,43	33,50	18,66	1,80	1,45
5	<i>E. guineensis</i>	Bahia 1	80	1151,06	23,02	12,55	1,83	1,60
6	<i>E. guineensis</i>	C 10	80	1229,80	24,59	13,57	1,81	1,31
7	<i>E. guineensis</i>	C 23	100	1435,47	28,70	16,00	1,79	1,41
8	<i>E. guineensis</i>	C 25	80	4434,53	88,69	47,83	1,85	1,80

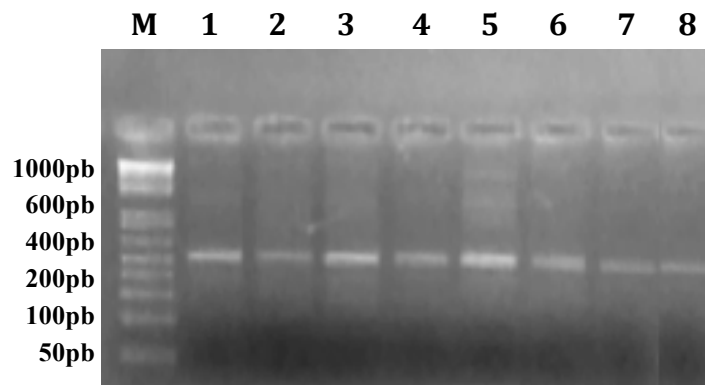
### 4.3 Amplificação de RGAs

A combinação A de *primers* abrange o domínio NBS, sendo o *primer* forward P1A capaz de se anelar ao motivo P-loop e o *primer* reverso P3A, ao motivo GLPL, gerando um produto de, aproximadamente, 550 pb. A especificidade dos *primers* desta combinação não foi satisfatória, apresentado co-amplificação de produtos de PCR de tamanho não esperado (Figura 9). Como consequência, esta combinação de *primers* não foi utilizada nos passos seguintes do processo de caracterização de RGAs.



**Figura 9:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2  $\mu$ l dos produtos inespecíficos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos *primers* P1A e P3A. Usou-se como marcador para quantificação 4  $\mu$ l de *Low Mass Ladder*. Amostras no gel: (M) Marcador de peso molecular (*Low DNA Mass Ladder*); (1) *E. oleifera*, acesso Manicoré; (2) *E. oleifera*, acesso Tefé; (3) *E. guineensis*, acesso Nigéria; (4) *E. guineensis*, acesso LM2T; (5) *E. guineensis*, acesso Bahia 1; (6) *E. guineensis*, acesso C10; (7) *E. guineensis*, acesso C23; (8) *E. guineensis*, acesso C25).

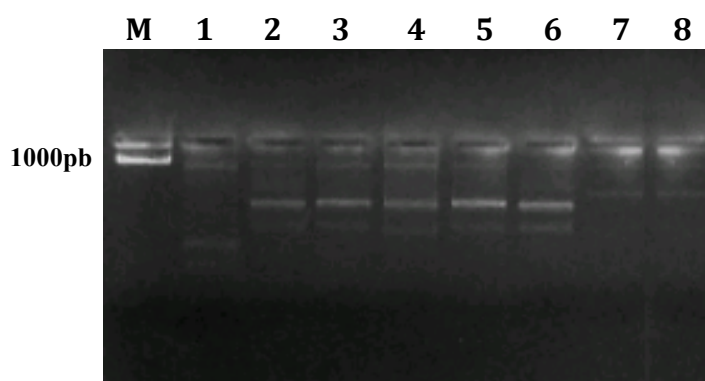
A combinação B, com os *primers* P1B e RNBS-D, se mostrou específica (Figura 10). P1B (*forward*) se liga ao motivo P-loop, enquanto RNBS-D (*reverse*) se liga a um motivo de mesmo nome. O produto de PCR gerado contém, aproximadamente, 650 pb.



**Figura 10:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2  $\mu$ l dos produtos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos *primers* P1B e RNBS-D. Usou-se como marcador para quantificação 4  $\mu$ l de *Low Mass Ladder*. Amostras no gel: (M) Marcador de peso molecular (*Low DNA Mass Ladder*); (1) *E. oleifera*, acesso Manicoré; (2) *E. oleifera*, acesso Tefé; (3) *E. guineensis*, acesso Nigéria; (4) *E. guineensis*, acesso LM2T; (5) *E. guineensis*, acesso Bahia 1; (6) *E. guineensis*, acesso C10; (7) *E. guineensis*, acesso C23; (8) *E. guineensis*, acesso C25).

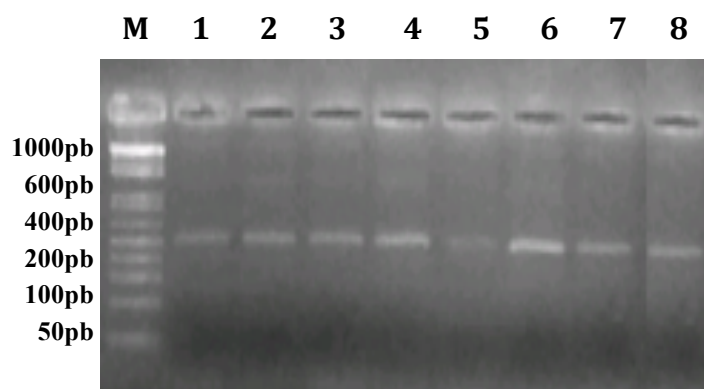
A combinação C é capaz de gerar um produto de, aproximadamente, 550 pb. Similarmente à combinação A, as regiões de anelamento são os motivos P-loop (*primer*

*foward* 1F) e GLPL (*primer reverse* P3B). Como houve inespecificidade de anelamento dos *primers* (Figura 11), o uso desta combinação foi descartado.



**Figura 11:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2  $\mu$ l dos produtos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos *primers* 1F e P3B. Usou-se como marcador para quantificação 4  $\mu$ l de *Low Mass Ladder*. Amostras no gel: (M) Marcador de peso molecular (*Low DNA Mass Ladder*); (1) *E. oleifera*, acesso Manicoré; (2) *E. oleifera*, acesso Tefê; (3) *E. guineensis*, acesso Nigéria; (4) *E. guineensis*, acesso LM2T; (5) *E. guineensis*, acesso Bahia 1; (6) *E. guineensis*, acesso C10; (7) *E. guineensis*, acesso C23; (8) *E. guineensis*, acesso C25).

A combinação D abrange os domínios NBS e LRR. O *primer* 3F2, que se liga ao motivo kinase 2, e o *primer* 13R1, que se liga a uma região do domínio LRR, geraram um produto de, aproximadamente, 700 pb, como esperado (Figura 12).



**Figura 12:** Visualização em gel de agarose 1,5% corado com brometo de etídeo de 2  $\mu$ l dos produtos inespecíficos de PCR de *E. oleifera* e *E. guineensis* gerados pela combinação dos *primers* 3F2 e 13R1. Usou-se como marcador para quantificação 4  $\mu$ l de *Low Mass Ladder*. Amostras no gel: (M) Marcador de peso molecular (*Low DNA Mass Ladder*); (1) *E. oleifera*, acesso Manicoré; (2) *E. oleifera*, acesso Tefê; (3) *E. guineensis*, acesso Nigéria; (4) *E. guineensis*, acesso LM2T; (5) *E. guineensis*, acesso Bahia 1; (6) *E. guineensis*, acesso C10; (7) *E. guineensis*, acesso C23; (8) *E. guineensis*, acesso C25).

A Tabela 8 resume os resultados da amplificação por PCR, mostrando as etapas atingidas para cada acesso e combinação de *primer*.

**Tabela 8:** Lista de *primers* degenerados testados, etapa concluída para cada combinação e seus respectivos motivos de uso ou não uso.

Combinação de <i>primers</i>	Etapas concluídas	Motivo
A (P1A / P3A)	Amplificação por PCR	Produtos inespecíficos
B (P1B / RNBS-D)	Amplificação por PCR, clonagem e seqüenciamento	Produtos específicos
C (1F / P3B)	Amplificação por PCR	Produtos inespecíficos
D (3F2 / 13R1)	Amplificação por PCR, clonagem e seqüenciamento	Produtos específicos

#### 4.4 Análise de bioinformática das seqüências

##### 4.4.1 Processamento das seqüências e montagem de *contigs*

Um total de 768 seqüências foram geradas a partir de *primers* degenerados direcionados à família de R genes NBS-LRR. A submissão destas seqüências ao Sistema Genoma (SISGEN) (Pappas, 2008) gerou um total de 486 seqüências de alta qualidade, sendo 71 seqüências do acesso Manicoré (*E. oleifera*), 32 seqüências originadas de Tefê (*E. oleifera*), 66 de Nigéria (*E. guineensis*), 61 de LM2T (*E. guineensis*), 69 de Bahia 1 (*E. guineensis*), 81 de C10 (*E. guineensis*), 44 de C23 (*E. guineensis*) e 62 de C25 (*E. guineensis*). No total, 365 seqüências apresentaram homologia a R-genes ou RGAs. Os tratamentos 4, 12, 16, 18, 22 e 24 apresentaram mais 70% de seqüências de alta qualidade e apenas as seqüências geradas relativas aos tratamentos 6, 26 e 30 não apresentaram homologia a genes de resistência ou RGAs (Tabela 9).

A partir das seqüências tratadas foram montados 44 *contigs* dos quais 38 apresentaram o domínio NB-ARC e 29 possuíam ORFs completas. Apenas sete *contigs* foram provenientes de um único acesso: os *contigs* 3, 40 e 44, provenientes de Bahia 1 (*E. guineensis*); o *contig* 14, proveniente de LM2T (*E. guineensis*); o *contig* 16, proveniente de Manicoré (*E. oleifera*); e os *contigs* 26 e 39, provenientes de Nigéria (*E. guineensis*). 16 *singletons* apresentaram presença do domínio NB-ARC e motivos conservados, com 10 possuindo ORFs completas. Estes *singletons* variaram quanto ao acesso do qual provêm. O acesso C25 não apresentou *singletons* com domínio NB-ARC e apenas quatro de suas seqüências contribuíram para a formação de *contigs*.

As seqüências foram traduzidas a fim de buscar os motivos de aminoácidos que caracterizam a região NBS em plantas. O alinhamento múltiplo realizado pelo programa

MUSCLE utilizando seqüências protéicas dos 44 *contigs* e 16 *singletons* derivados de *primers* direcionados à família NBS-LRR de R-genes revelou que a maioria destas proteínas são altamente conservadas em *E. guineensis* e *E. oleifera*.

Os motivos conservados encontrados nos *contigs* e *singletons* foram P-loop, *kinase-2*, GLPL, RNBS-C e RNBS-D (Figura 13). Todos os RGAs clonados continham motivos consistentes com a subfamília não TIR. O resíduo final do motivo *kinase-2* é útil para classificar as seqüências como TIR ou não TIR, com um triptofano (LLVLDDV**W**) presente apenas em seqüências não TIR (Meyers *et al.*, 1999). Apenas quatro seqüências, *contigs* e *singletons*, apresentaram o motivo P-loop, caracterizado pela seqüência de aminoácidos GVGKTT (glicina-valina-lisina-treonina-treonina). Das 60 seqüências, 22 apresentaram o motivo *kinase-2*, caracterizado pela seqüência de aminoácidos LLVLDDV (leucina-leucina-valina-leucina-aspartato-aspartato-valina) seguido por um triptofano (W), visto que as seqüências obtidas foram todas não TIR. O motivo GLPL, caracterizado pelos aminoácidos CGGLPLA (cisteína-glicina-glicina-leucina-prolina-leucina-alanina) estava presente em 50 seqüências.

Variações em seqüências de aminoácidos para os quatro motivos conservados foram aceitas, pois houve considerável divergência entre as seqüências. Apenas o *contig* 8 apresentou os quatro motivos em conjunto. Dos 60 *contigs* e *singletons*, 19 obtiveram códons de terminação precoce. A Tabela 10 resume os resultados obtidos.

**Tabela 9:** Estatísticas dos *amplicons* obtidos para todos os tratamentos usando *primers* degenerados direcionados a NBS-LRR RGAs e genes de resistência.

Trat.	Combinação de <i>primers</i>	Espécie	Acesso	Motivo alvo dos <i>primers</i>	Domínio	Seqüências de alta qualidade	Seqüências com homologia a R-genes ou RGAs
2	P1B / RNBS-D	<i>E. oleifera</i>	Manicoré	P-loop e RNBS-D	NBS não-TIR	33 (68,7%)	30 (90,1%)
4	3F2 / 13R1	<i>E. oleifera</i>	Manicoré	Kinase 2 e LRR	NBS-LRR não-TIR	38 (79,2%)	34 (89,5%)
6	P1B / RNBS-D	<i>E. oleifera</i>	Tefé	P-loop e RNBS-D	NBS não-TIR	24 (50%)	0
8	3F2 / 13R1	<i>E. oleifera</i>	Tefé	Kinase 2 e LRR	NBS-LRR não-TIR	8 (16,7%)	8 (100%)
10	P1B / RNBS-D	<i>E. guineensis</i>	Nigéria	P-loop e RNBS-D	NBS não-TIR	31 (64,6%)	30 (96,8%)
12	3F2 / 13R1	<i>E. guineensis</i>	Nigéria	Kinase 2 e LRR	NBS-LRR não-TIR	35 (72,9%)	32 (91,4%)
14	P1B / RNBS-D	<i>E. guineensis</i>	LM2T	P-loop e RNBS-D	NBS não-TIR	21 (43,7%)	18 (85,7%)
16	3F2 / 13R1	<i>E. guineensis</i>	LM2T	Kinase 2 e LRR	NBS-LRR não-TIR	40 (83,3%)	37 (92,5%)
18	P1B / RNBS-D	<i>E. guineensis</i>	Bahia1	P-loop e RNBS-D	NBS não-TIR	37 (77,1%)	31 (83,8%)
20	3F2 / 13R1	<i>E. guineensis</i>	Bahia1	Kinase 2 e LRR	NBS-LRR não-TIR	32 (66,7%)	27 (84,4%)
22	P1B / RNBS-D	<i>E. guineensis</i>	C10	P-loop e RNBS-D	NBS não-TIR	41 (85,4%)	41 (100%)
24	3F2 / 13R1	<i>E. guineensis</i>	C10	Kinase 2 e LRR	NBS-LRR não-TIR	40 (83,3%)	37 (92,5%)
26	P1B / RNBS-D	<i>E. guineensis</i>	C23	P-loop e RNBS-D	NBS não-TIR	28 (58,3%)	0
28	3F2 / 13R1	<i>E. guineensis</i>	C23	Kinase 2 e LRR	NBS-LRR não-TIR	16 (33,3%)	13 (81,3%)
30	P1B / RNBS-D	<i>E. guineensis</i>	C25	P-loop e RNBS-D	NBS não-TIR	30 (62,5%)	0
32	3F2 / 13R1	<i>E. guineensis</i>	C25	Kinase 2 e LRR	NBS-LRR não-TIR	32 (66,7%)	27 (84,4%)

	10	20	30	40	50	60	70	80	90
Contig1/1-170	R Y F L V L D D V W	T A N V	W C D L L R	N T L Y G C A A G S R V	L V T T R N	E Q I A M Q M M	A V N T H H V D K L S	L G D G W S	L L C K K V V L T G E E G E I
Contig10/1-165	---	---	---	---	---	---	---	---	---
Contig11/1-163	---	---	---	---	---	---	---	---	---
Contig12/1-172	---	---	---	---	---	---	---	---	---
Contig13/1-189	R Y F L V L D D V W	D T K	W E A L C	N T L K S C A V G S R I	L V T T R N	E Q I A M Q M R	A V N T H H I N K L S	S K D S W L L F C K K V V L T G E E E E I	
Contig14/1-117	---	---	---	---	---	---	---	---	---
Contig15/1-172	R Y F L V L D D V W	D E T G S K	W K E L R	G V L T S G A K G S M V	L V T T Q S	P I V A E V M G	T M D P V E L K S	L Q E D D F W R L F E H C V F G E N N L K E E	
Contig16/1-163	---	---	---	---	---	---	---	---	---
Contig17/1-99	---	---	---	---	---	---	---	---	---
Contig18/1-47	---	---	---	---	---	---	---	---	---
Contig19/41-200	R F L L V L D D V W	K D E R S S R L V E E E R	W E R F L	S P L K A G K K G S K I	L M T T R L	A T V A E M V G	T M D S I N L A G L R D H D C W S	L I K E C T F D G V H P D K H	
Contig2/7-176	R Y F L V L D D V W	D T K	W E A L C	D Y L T Y A E K G S K I	L I T A R E	E K S V D V M G	R T V R K I F L E G L A Y R D F W L	F F E A C A F G N E R V S L F	
Contig20/1-154	---	---	---	---	---	---	---	---	---
Contig21/1-149	---	---	---	---	---	---	---	---	---
Contig22/1-130	---	---	---	---	---	---	---	---	---
Contig23/1-165	---	---	---	---	---	---	---	---	---
Contig24/1-140	---	---	---	---	---	---	---	---	---
Contig25/1-150	---	---	---	---	---	---	---	---	---
Contig26/1-120	---	---	---	---	---	---	---	---	---
Contig27/1-169	---	---	---	---	---	---	---	---	---
Contig28/1-164	R Y F L V L D D V W	N E D Q V K	W D E L K	A L L G T G G E G S R I	I V T A R S	E G V S R I M G	P L D A Y R L D C L N D H D S W E L F	F H C K A F G E A T L E E G	
Contig29/1-153	---	---	---	---	---	---	---	---	---
Contig3/1-127	R Y F L V L D D V W	---	W L S	D Q K A A G I Y R E A A A S	E D L T Q I I I L G	A I	---	---	
Contig30/15-14	---	---	---	---	---	---	---	---	---
Contig31/7-171	G Y F L V L D D V W	N P	W E E L K	E L L T A G A E G S C V	I V T T R S T T V V S S I I G	---	---	---	
Contig32/58-228	R F L L V L D D V R	N E E K R K W E N K E R	W R K L L	A P L Q F G E S G S R I	L V T T R M	E M V A E M L D	S M S P V F L K G L A R D E Y W L	L F F K K C A F G S E E F S R Y	
Contig33/1-164	---	---	---	---	---	---	---	---	---
Contig34/15-14	---	---	---	---	---	---	---	---	---
Contig35/41-213	R S L I V L D D V W	D E Q Q R L	W Q V L F	Q P L N C G L R G S K V	I M T T R N R S A L S E N V P	---	---	---	
Contig36/1-155	---	---	---	---	---	---	---	---	---
Contig37/56-200	R Y L I V L D D V W	H K E V	W D D V G	S A L F D N G C G S R V	I V T T R N	Q E A A S A V	D I R V M K L E P L Q E Y D A W T	L F R K E A F G R D N D G S C	
Contig38/61-229	R Y L I V L D D M W	R I N R C	F D E F S	S V L V D S R C R S R I	V T T W I	H D V A C L A A	E S R V M L E K P L Q E S D S W I	L F R C N K A F W K Y K D R S C	
Contig39/1-153	---	---	---	---	---	---	---	---	---
Contig4/1-170	R Y F L V L D D V W	N E E E A E	W E R L K	P L L R G G K K G S K I	I V T T R S	E R V A S I M G	A V N T H H V N K L S	L E D S W L L L C K K V V L T G E E G E I	
Contig40/14-75	---	---	---	---	---	---	---	---	---
Contig41/1-169	R Y F L V L D D V W	N E D P V R	W D E F K	D L L G S R E G S T I	L V T T Q S	E Q V S R I L R	T V P P Y S L S F L S D D S W T L	F R Q R A F S P G A G E R P	
Contig42/1-170	R Y F L V L D D V W	N E N Y E K	W D R L R	T V L I G S A R G S K V	I V T T R S	A R V A S I M G	T V A P H F L S D L S E E D C W L L	F E K R A F G L G C K E T	
Contig43/1-172	R Y F L V L D D V W	D E T G S A	W E K L R	S A L T S G A K G S M V	L V T T Q S	L R V A E V M G	T M D P I E L K S L Q E D D F W R L	F E H C A F G E N N L E E E	
Contig44/1-149	---	---	---	---	---	---	---	---	---
Contig5/1-141	---	---	---	---	---	---	---	---	---
Contig6/1-137	---	---	---	---	---	---	---	---	---
Contig7/53-212	K I A I M L D D I W A R L E	---	L K D I G	I P F R D D H K G C K I	L L T S R K	S E V C D L M E	T D V S V L D V L S E E D S W T L	F R N K A F W K Y K D R S C	
Contig8/15-177	R Y L I V L D D M W	H I N G W	F D E L S	P V L V D S K C R S R I	V T T R N	H D V A Y L A V	E S R V L E L Q P L Q E G D A W I	L F R N K A F W K Y K D R S C	
Contig9/1-172	R Y F L V L D D V W	D E T G S K	W K E L R	G V L T S G A K G S M V	L V T T Q S	P V V A E V M G	T M E P I E L K S L E E D D H W R L	F E H C A F G E A T L E E E	
Trat10_EG_BA1_3F2/13R1_NI_15/1-151	---	---	---	---	---	---	---	---	---
Trat10_EG_BA1_3F2/13R1_NI_5/1-82	---	---	---	---	---	---	---	---	---
Trat11_EG_C10_P1B/RNBSD_R_45/7-154	F M D I V I R D T W	M M T Y G T K	P G M D I K	Q H F R R D L Q E V E V M L	A T L T K	A N V A M D A D P D T D	S P P C E L I L H S P E E S W D L L R K S F T	---	
Trat11_EG_C10_P1B/RNBSD_R_7/46-188	S F L L V L D D V W	N E D E A K	W E R L K	T L L R C G K K G S K I	V T T T R S	D V V A R I M G	T V A P H K L Q G L T A A D C W T L	F K Q R A F G P G R E E E T	
Trat12_EG_C10_3F2/13R1_R_18/1-125	---	---	---	---	---	---	---	---	---
Trat14_EG_C23_3F2/13R1_MR_2/1-151	---	---	---	---	---	---	---	---	---
Trat14_EG_C23_3F2/13R1_MR_28/1-129	---	---	---	---	---	---	---	---	---
Trat14_EG_C23_3F2/13R1_MR_48/1-89	---	---	---	---	---	---	---	---	---
Trat1_EO_MAN_P1B/RNBSD_CO_7/1-125	---	---	---	---	---	---	---	---	---
Trat2_EO_MAN_3F2/13R1_CO_1/1-163	---	---	---	---	---	---	---	---	---
Trat2_EO_MAN_3F2/13R1_CO_2/1-141	---	---	---	---	---	---	---	---	---
Trat2_EO_MAN_3F2/13R1_CO_20/1-142	---	---	---	---	---	---	---	---	---
Trat4_EO_TEF_3F2/13R1_CO_21/1-171	R Y F L V L D D V W	D E T G S K	W E K L R	S A L T S G A K G S M V	L V T T Q S	P V V A K V M G	T M D P I E L E C L Q E D D Y W R L	F E H C A F G D Q I L E E E	
Trat5_EG_NIG_P1B/RNBSD_CO_6/45-204	R Y L V L D D V W	N E D R N Q	W E P I F	L P L M Y G E R G S K I	L I T T Q N	K N V A N I M D	M A S S Y C L E G L S N Q D C W L I	F K R I A F A N K N S S E H	
Trat8_EG_LM2T_3F2/13R1_R_20/1-138	---	---	---	---	---	---	---	---	---
Trat8_EG_LM2T_3F2/13R1_R_34/1-141	---	---	---	---	---	---	---	---	---



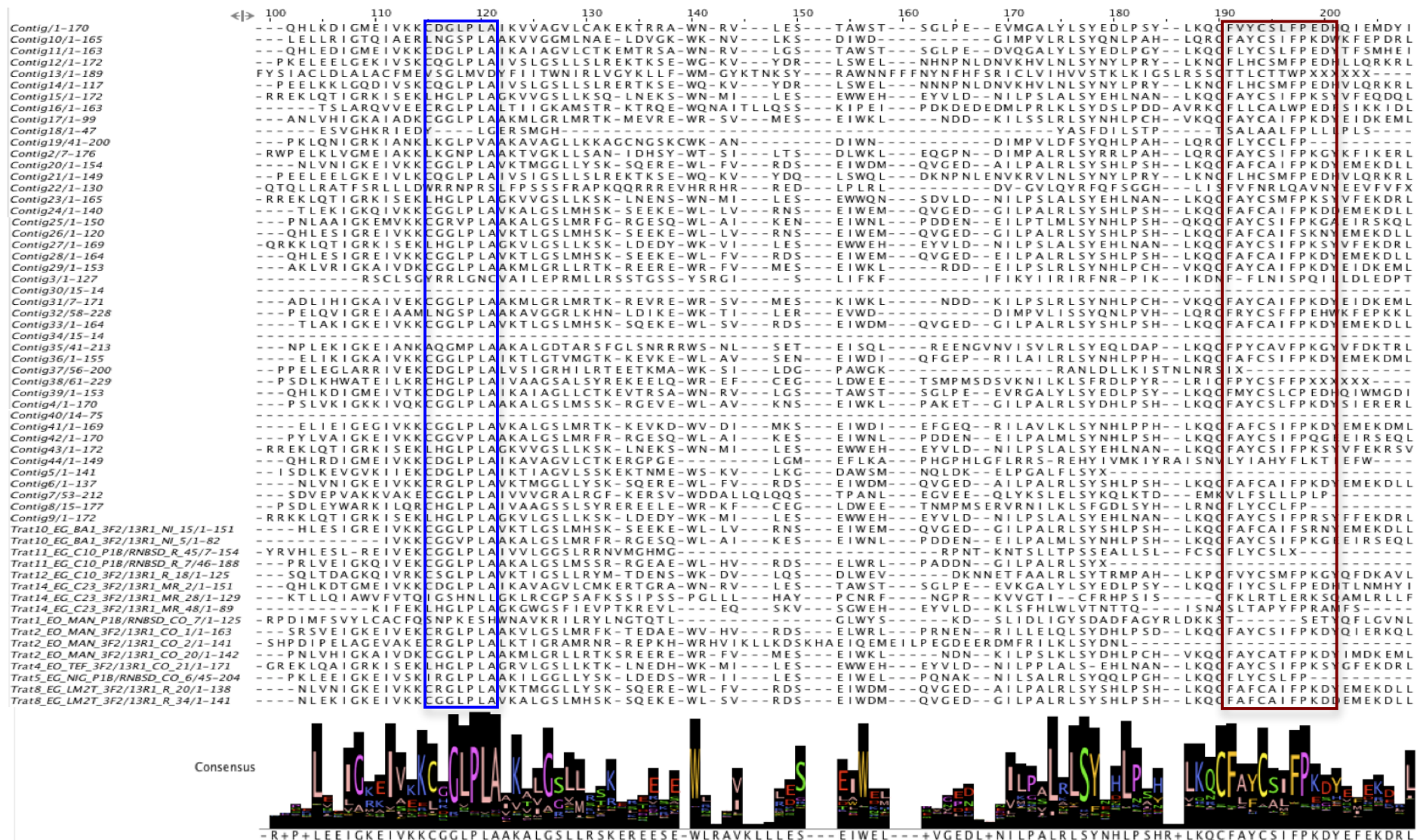


Figura 13: Motivos conservados e seqüência consenso de RGAs NBS-LRR de *E. guineensis* e *E. oleifera*. Em verde está o motivo *kinase-2*. Em azul, o motivo GLPL. Em vermelho, o motivo RNBS-D.

**Tabela 10:** Motivos conservados encontrados nos diferentes acessos de *Elaeis* com duas combinações de *primers*

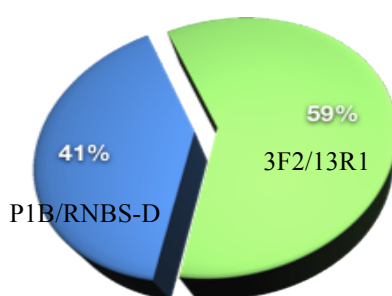
Seqüência	Domínio NB-ARC	P-loop	Kinase-2	GLPL	Códon de terminação precoce	Acesso de <i>Elaeis</i>	Combinação de <i>primer</i>	Tamanho do singleton ou contig (pb)	Quantidade de seqüências no contig ou singleton
c1	x		x	x		MAN, BA1, LM2T, C10, C23, C25	P1B/RNBS-D e 3F2/13R1	659	10
c2	x		x	x		MAN, NIG, BA1, LM2T, C10, C23, C25	P1B/RNBS-D e 3F2/13R1	716	17
c3			x		x	LM2T	P1B/RNBS-D e 3F2/13R1	411	2
c4	x		x	x		MAN, NIG, LM2T, C10, C25	P1B/RNBS-D e 3F2/13R1	698	13
c5	x			x	x	BA1	P1B/RNBS-D	426	2
c6	x			x		NIG, LM2T, C10	P1B/RNBS-D e 3F2/13R1	630	3
c7	x			x		MAN, NIG, BA1, LM2T, C10	P1B/RNBS-D	635	11
c8	x	x	x	x		MAN, NIG, BA1, LM2T, C10	P1B/RNBS-D e 3F2/13R1	740	64
c9	x		x	x		TEF, MAN, LM2T, C23	3F2/13R1	684	6
c10	x			x		TEF, MAN, NIG, BA1, LM2T, C10, C25	P1B/RNBS-D e 3F2/13R1	688	28
c11	x			x		NIG, BA1, LM2T, C10	P1B/RNBS-D e 3F2/13R1	643	5
c12	x		x	x		C10, C25	3F2/13R1	721	3
c13			x		x	MAN, NIG, BA1, LM2T, C10, C23, C25	P1B/RNBS-D e 3F2/13R1	633	22
c14	x			x		LM2T	P1B/RNBS-D e 3F2/13R1	476	2
c15	x		x	x		MAN, NIG, BA1, C10, C25	P1B/RNBS-D e 3F2/13R1	703	21
c16	x			x		MAN	3F2/13R1	658	4
c17	x			x	x	NIG, LM2T, C23	P1B/RNBS-D e 3F2/13R1	436	4
c18				x	x	MAN, BA1	P1B/RNBS-D	152	2
c19	x		x		x	NIG, BA1, C10, C23	P1B/RNBS-D e 3F2/13R1	612	4
c20	x			x		TEF, MAN, NIG, LM2T, C10, C25	P1B/RNBS-D e 3F2/13R1	675	14
c21	x			x		MAN, NIG, LM2T	3F2/13R1	624	3
c22	x			x	x	TEF, MAN	3F2/13R1	426	2
c23	x			x		MAN, NIG, BA1, LM2T, C23	3F2/13R1	663	7
c24	x			x		NIG, LM2T, C10, C25	P1B/RNBS-D e 3F2/13R1	619	5
c25	x					MAN, LM2T, C10	3F2/13R1	632	3
c26	x			x		NIG	P1B/RNBS-D e 3F2/13R1	532	2
c27	x			x	x	LM2T, C10	3F2/13R1	632	2
c28	x		x	x		MAN, BA1, LM2T	P1B/RNBS-D e 3F2/13R1	662	7
c29	x			x	x	TEF, MAN, NIG, BA1, LM2T	P1B/RNBS-D e 3F2/13R1	607	8
c30		x				MAN, NIG, BA1	P1B/RNBS-D	121	5
c31	x		x	x		MAN, BA1, C10, C23, C25	3F2/13R1	654	14



Seqüência	Domínio NB-ARC	P-loop	Kinase-2	GLPL	Códon de terminação precoce	Acesso de EfaeIs	Combinação de primer	Tamanho do singleton ou contig (pb)	Quantidade de seqüências no contig ou singleton
c32	x		x	x		MAN, NIG, BA1, LM2T, C10	P1B/RNBS-D e 3F2/13R1	874	8
c33	x		x	x		NIG, BA1, LM2T, C10	3F2/13R1	710	7
c34		x				MAN, BA1, LM2T	P1B/RNBS-D	121	5
c35	x		x	x		MAN, NIG, BA1, C10, C23	P1B/RNBS-D e 3F2/13R1	798	7
c36	x			x	x	NIG, LM2T, C10, C23, C25	3F2/13R1	676	10
c37	x		x	x		MAN, BA1, LM2T	P1B/RNBS-D	599	3
c38	x			x		MAN, NIG, BA1, LM2T, C10	P1B/RNBS-D	695	27
c39	x			x		NIG	3F2/13R1	621	3
c40		x			x	BA1	P1B/RNBS-D	419	3
c41	x		x	x		MAN, NIG, BA1, LM2T, C10	P1B/RNBS-D e 3F2/13R1	700	16
c42	x		x	x		LM2T, C25	3F2/13R1	671	2
c43	x		x	x	x	TEF, NIG, LM2T	P1B/RNBS-D e 3F2/13R1	685	6
c44	x			x	x	BA1	3F2/13R1	634	2
s1	x			x		BA1	3F2/13R1	435	1
s2	x			x	x	BA1	3F2/13R1	666	1
s3	x			x	x	C10	P1B/RNBS-D	493	1
s4	x		x	x		C10	P1B/RNBS-D	564	1
s5	x			x		C10	3F2/13R1	537	1
s6	x			x		C23	3F2/13R1	614	1
s7	x				x	C23	3F2/13R1	620	1
s8	x			x	x	C23	3F2/13R1	466	1
s9	x				x	MAN	P1B/RNBS-D	467	1
s10	x			x	x	MAN	3F2/13R1	672	1
s11	x			x		MAN	3F2/13R1	438	1
s12	x			x		MAN	3F2/13R1	604	1
s13	x		x	x		TEF	3F2/13R1	585	1
s14	x		x	x		NIG	P1B/RNBS-D	613	1
s15	x			x		LM2T	3F2/13R1	652	1
s16	x			x		LM2T	3F2/13R1	653	1

A combinação de *primers* P1B/RNBS-D gerou 245 seqüências de alta qualidade, dentre as quais 150 apresentaram homologia com R genes ou RGAs. A combinação 3F2/13R1 gerou 241 seqüências de alta qualidade e 215 com homologia a R genes ou RGAs. A combinação de *primers* 3F2/13R1 foi mais eficiente na amplificação de RGAs NBS-LRR e as seqüências geradas a partir desta combinação tiveram maior participação na montagem dos *contigs* (Figura 14).

**Figura 14:** Eficiência das combinações de *primers* P1B/RNBS-D e 3F2/13R1 na amplificação de RGAs.



Os *contigs* e *singletons* foram submetidos à análises por Blastx para identificação de domínios de interesse (NBS-LRR) e similaridades entre os RGAs caracterizados neste trabalho e seqüências de nucleotídeos não redundantes (nr) do Genbank. A Tabela 11 mostra os resultados obtidos pelo Blastx identificando *hits* com outras espécies, como por exemplo, arroz, banana, mamona, aveia, cevada, entre outros. O amplo espectro de diferentes seqüências com *hits* apresentando alta similaridade pelo Blastx (28 seqüências protéicas de genes de resistência ou RGAs de plantas monocotiledôneas e 25 de plantas dicotiledôneas) confirma a conservação esperada dos domínios protéicos relacionados à resistência.

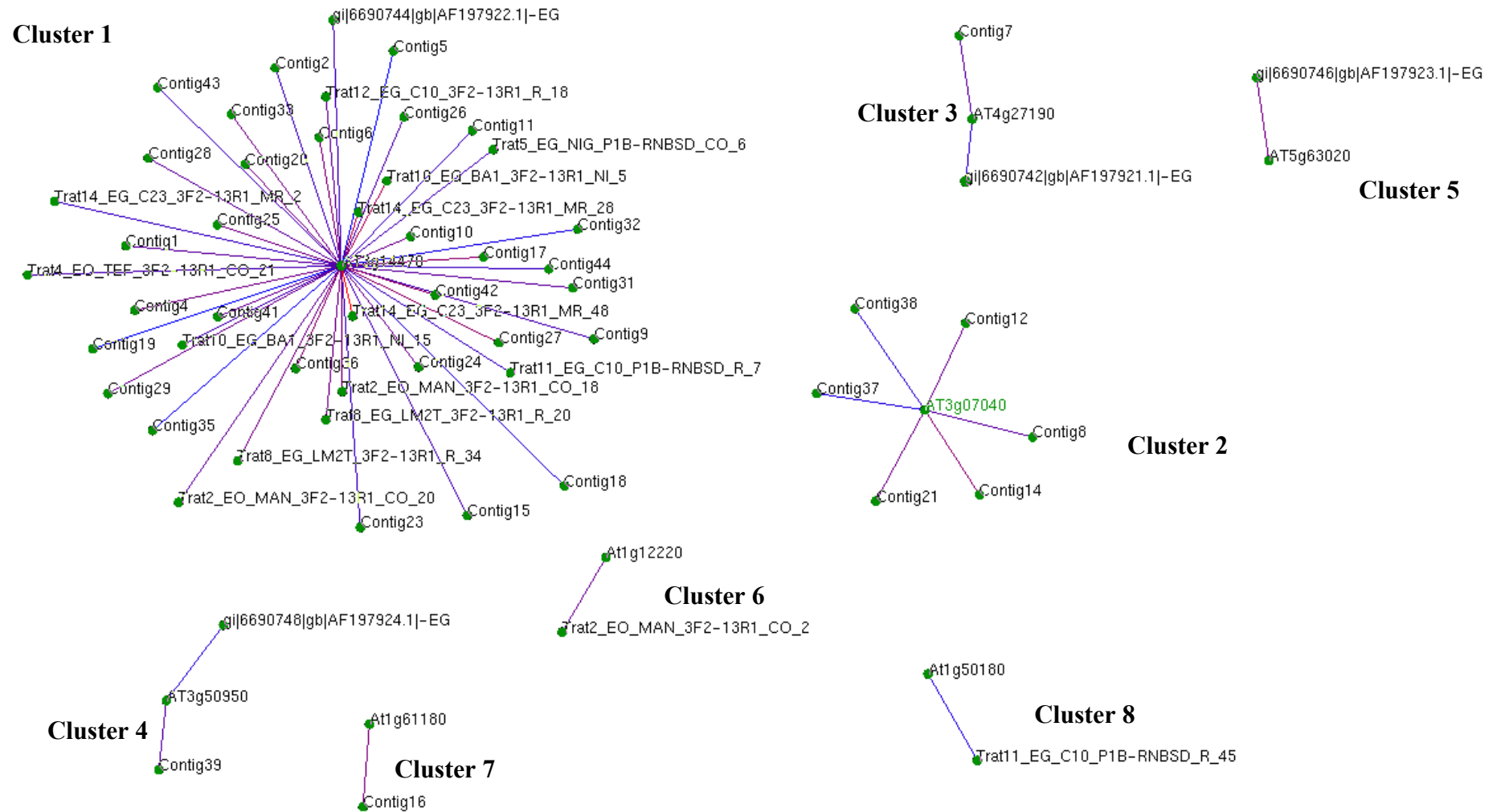
O resultado da análise de similaridade entre as seqüências protéicas de RGAs da família NBS-LRR de *E. guineensis* e *E. oleifera* com NB-ARC R-genes e RGAs de *A. thaliana* utilizando BLASTx e BioLayout Express<sup>3D</sup> foi uma imagem 2D, onde as seqüências foram agrupadas em 8 *clusters*. No centro de cada um destes *clusters* há uma seqüência de *A. thaliana* (Figura 15).

O *cluster* 1 foi formado por 41 seqüências diversas, não sendo possível a observação de uma tendência no agrupamento, visto que os *contigs* e *singletons* usados para a montagem do grupo possuíam seqüências de ambas espécies, todos os acessos e combinações de *primers*, e também divergiam quanto a resistência ao patógeno. Apesar da variabilidade, as seqüências deste *cluster* apresentaram similaridade a uma seqüência de NBS-LRR RGA de *E. guineensis* depositada no Genbank (AF197922.1).

**Tabela 11:** Resultado das análises por Blastx para identificação de domínios de interesse (NBS-LRR) e similaridades entre os RGAs e sequências de nucleotídeos não redundantes (nr) do Genbank.

Seqüência	Número de Acesso	Proteína ou seqüência correspondente no Blastx	E-value
c1	ADX31660 gi 323135509 gb ADX31660.1	resistance protein [Musa ABB Group]	7E-68
c2	BAD67833 gi 55296114 dbj BAD67833.1	putative disease resistance protein [Oryza sativa Japonica Group]	9E-55
c3		no hit	
c4	XP_002525457 gi 255568974 ref XP_002525457.1	leucine-rich repeat containing protein, putative [Ricinus communis]	2E-78
c5	ADX31658 gi 323135505 gb ADX31658.1	resistance protein [Musa AAB Group]	9E-27
c6	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	3E-56
c7	ABS82596 gi 154467273 gb ABS82596.1	putative NBS-LRR protein [Dracaena marginata]	2E-42
c8	XP_002301573 gi 224064832 ref XP_002301573.1	CC-NBS-LRR resistance protein [Populus trichocarpa]	6E-54
c9	BAB44079 gi 13872974 dbj BAB44079.1	putative NBS-LRR type resistance protein [Oryza sativa Japonica Group]	2E-49
c10	ACF21695 gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [Musa acuminata subsp. malaccensis]	2E-50
c11	ADX31660 gi 323135509 gb ADX31660.1	resistance protein [Musa ABB Group]	3E-68
c12	AAL76181 gi 1862682 gb AAL76181.1 AC074283_4	Putative disease resistance protein RPR1 [Oryza sativa]	7E-55
c13		no hit	
c14	AAL76181 gi 18642682 gb AAL76181.1 AC074283_4	Putative disease resistance protein RPR1 [Oryza sativa]	3E-55
c15	ACO88904 gi 226860356 gb ACO88904.1	putative resistance protein [Avena strigosa]	73-54
c16	BAB92866 gi 20805198 dbj BAB92866.1	putative NBS-LRR type disease resistance protein [Oryza]	9E-50
c17	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	3E-36
c18		no hit	
c19	AAB96980 gi 2792210 gb AAB96980.1	NBS-LRR type resistance protein [Hordeum vulgare subsp. vulgare]	3E-33
c20	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	2E-62
c21	AAL76181 gi 18642682 gb AAL76181.1 AC074283_4	Putative disease resistance protein RPR1 [Oryza sativa]	1E-67
c22	AAA72988 gi 209229 gb AAA72988.1	luciferase/kanamycin resistance protein [synthetic construct]	4E-55
c23	ACO88904 gi 226860356 gb ACO88904.1	putative resistance protein [Avena strigosa]	4E-52
c24	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	2E-55
c25	XP_002525457 gi 255568974 ref XP_002525457.1	leucine-rich repeat containing protein, putative [Ricinus communis]	4E-59
c26	XP_002301713 gi 224065200 ref XP_002301713.1	NBS-LRR resistance protein [Populus trichocarpa]	2E-47
c27	AAN03742 gi 22652532 gb AAN03742.1 AF456247_1	NBS-LRR-like protein [Oryza sativa Japonica Group]	4E-42
c28	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	1E-63

Seqüência	Número de Acesso	Proteína ou seqüência correspondente no Blastx	E-value
c29	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	5E-49
c30		no hit	
c31	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	1E-52
c32	gi 13957628 gb AAK50583.1 AC084404_8	putative resistance protein [Oryza sativa Japonica Group]	9E-66
c33	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	3E-65
c34		no hit	
c35	AAB96985 gi 2792220 gb AAB96985.1	NBS-LRR type resistance protein [Oryza sativa Japonica Group]	2E-53
c36	gi 38636689 dbj BAD03110.1	putative resistance gene analog PIC27 [Oryza sativa Japonica Group]	9E-59
c37	CAD45036 gi 28555914 emb CAD45036.1	NBS-LRR disease resistance protein homologue [Hordeum vulgare]	1E-35
c38	CAD45036 gi 28555914 emb CAD45036.1	NBS-LRR disease resistance protein homologue [Hordeum vulgare]	2E-47
c39	CBW30227 gi 301154124 emb CBW30227.1	Disease resistance protein (CC-NBS-LRR) [Musa balbisiana]	3E-64
c40		no hit	
c41	BAD08985 gi 42407842 dbj BAD08985.1	putative NBS-LRR resistance protein RGH1 [Oryza Japonica Group]	5E-69
c42	XP_002525457 gi 255568974 ref XP_002525457.1	leucine-rich repeat containing protein, putative [Ricinus communis]	2E-72
c43	AAN03742 gi 226532 gb AAN03742.1 AF456247_1	NBS-LRR-like protein [Oryza sativa Japonica Group]	1E-49
c44	CBW30227 gi 301154124 emb CBW30227.1	Disease resistance protein (CC-NBS-LRR) [Musa balbisiana]	8E-58
s1	XP_002301713 gi 224065200 ref XP_002301713.1	NBS-LRR resistance protein [Populus trichocarpa]	4E-56
s2	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	1E-33
s3	XP_002304589 gi 224075299 ref XP_002304589.1	NBS-LRR resistance protein [Populus trichocarpa]	3E-16
s4	XP_002525457 gi 255568974 ref XP_002525457.1	leucine-rich repeat containing protein, putative [Ricinus communis]	8E-56
s5	XP_002513078 gi 255544031 ref XP_002513078.1	leucine-rich repeat-containing protein, putative [Ricinus communis]	1E-44
s6	CBW30227 gi 301154124 emb CBW30227.1	Disease resistance protein (CC-NBS-LRR) [Musa balbisiana]	2E-62
s7	CBW30200 gi 301154106 emb CBW30200.1	Disease resistance protein (CC-NBS-LRR) [Musa balbisiana]	1E-28
s8	CBI35132 gi 297742679 emb CBI35132.3	unnamed protein product [Vitis vinifera]	3E-21
s9	CAN71189 gi 147804651 emb CAN71189.1	hypothetical protein VITISV_005044 [Vitis vinifera]	4E-53
s10	XP_002525457 gi 255568974 ref XP_002525457.1	leucine-rich repeat containing protein, putative [Ricinus communis]	1E-67
s11	XP_002517530 gi 255552975 ref XP_002517530.1	Disease resistance protein RPS5, putative [Ricinus communis]	2E-27
s12	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	1E-48
s13	AAT69649 gi 50080315 gb AAT69649.1	putative NBS-LRR type resistance protein [Oryza sativa Japonica Group]	2E-45
s14	AAF24310 gi 6690745 gb AAF24310.1 AF197922_1	resistance protein [Elaeis guineensis]	1E-76
s15	AAP45181 gi 109289909 gb AAP45181.2	Disease resistant protein rga3, putative [Solanum bulbocastanum]	2E-55
s16	XP_002530393 gi 255579089 ref XP_002530393.1	leucine-rich repeat containing protein, putative [Ricinus communis]	6E-56



**Figura 15:** Imagem de BioLayout 2D mostrando as relações de similaridade derivadas de Blastx ( $E$ -value  $1E^{-05}$ ) entre RGAs da família NBS-LRR de *E. guineensis* e *E. oleifera* e R-genes ou RGAs de *A. thaliana*. Foram formados dois *clusters* principais, outros dois grupos menores e quatro proteínas *outliers*.

O *cluster* 2, formado por seis *contigs*, possui seqüências provenientes de *E. oleifera* e *E. guineensis* de diferentes acessos e que foram amplificadas por ambas combinações de *primers*. Uma característica comum entre todos os acessos que forneceram as seqüências que montaram estes *contigs* e este *cluster* é a resistência a *Fusarium oxysporum* f.sp. *elaeidis*. Os *contigs* deste *cluster* não apresentaram similaridade a nenhuma seqüência de RGA de *E. guineensis* depositada no Genbank.

Os *clusters* 3 e 4, formados por um *contig* cada, possuem seqüências que apresentaram similaridade a RGAs de *E. guineensis* depositados no Genbank. O *contig* 7, formador do *cluster* 3, possui todas as suas seqüências amplificadas pela combinação de *primers* P1B e RNBS-D e são provenientes de acessos de *E. guineensis* que apresentam resistência a *Fusarium*. Houve, neste *cluster*, similaridade a uma seqüência de RGA de *E. guineensis* depositada no GeneBank. O *cluster* 4 é composto por uma seqüência de RGA de *E. guineensis* depositada no GeneBank, uma seqüência de *A. thaliana* e pelo *contig* 39. Este *contig* foi formado a partir de tres seqüências provenientes do mesmo tratamento (*E. guineensis*, acesso Nigéria, combinação de *primers* 3F2 e 13R1, resistente).

O *cluster* 5 foi formado apenas por seqüências do GeneBank de *A. thaliana* e *E. guineensis*. Os *clusters* 6, 7 e 8 foram constituídos por proteínas *outliers*, ou seja, não formaram grupos robustos.

#### 4.4.2 Análise filogenética

A análise filogenética feita pelo método *Neighbor Joining* das seqüências de aminoácidos entre os motivos *kinase-2* e RNBS-D dos 60 RGAs NBS-LRR amplificados, incluindo aqueles com códon de terminação precoce está demonstrada na Figura 16. Embora somente as regiões contendo os motivos conservados tenham sido alinhadas e comparadas, uma divergência significativa foi observada na árvore, com um total de 5 *clusters* e numerosos subclados. Este resultado indica a presença de uma família diversa de genes que codificam proteínas com domínio NBS-LRR no gênero *Elaeis*.

Não foi possível verificar evidência na formação dos diferentes *clusters*. Os componentes de cada *cluster* divergiam grandemente em relação à espécie, acesso e combinação de *primers* que originaram as seqüências. Além disso, divergiam também quanto à característica resistência a *Fusarium*.



## 5. DISCUSSÃO

### 5.1 Importância de *Elaeis*

O dendezeiro é considerado entre as oleaginosas a de maior produtividade no mundo, podendo atingir de 8 a 10 t de óleo/ha/ano em condições ideais. O óleo é de amplo uso na indústria de alimentos, farmacêutica e química, e hoje é considerado um dos melhores óleos para produção de biodiesel. A produção brasileira, que representa em torno de 0,4% da produção mundial (Oil World, 2005), tem evoluído lentamente nas últimas décadas, e não atende a demanda nacional. No ano de 2008, o Brasil importou 262,2 toneladas de óleo de palma, o que representa mais da metade do consumo interno, significando um dispêndio de US\$297,4 milhões (SRI/MAPA e SECEX/MDIC).

A cultura é indicada como uma das principais alternativas para agricultura sustentável e recuperação de áreas degradadas na Amazônia. Contudo, a expansão segura, competitiva e sustentável depende da solução de problemas limitantes à cultura no Brasil e no continente americano, como pragas e doenças, uso inadequado de fertilizantes, restrita base genética das cultivares e necessidade do desenvolvimento e avaliação de cultivares com adaptabilidade aos ecossistemas de expansão da cultura no país.

### 5.2 Necessidade de melhoramento de *Elaeis*

Os programas de melhoramento genético do dendezeiro tem apresentado excelentes resultados em termos de ganho anual. O programa de melhoramento do Cirad, por exemplo, do qual são oriundas as cultivares produzidas pela Embrapa, tem apresentado um ganho anual de 1,2% até o terceiro ciclo de seleção. Contudo, os ciclos de melhoramento são longos, o que aumenta o prazo necessário para lançamentos de novas linhagens (tipicamente entre 7 a 10 anos). Adicionalmente, os materiais atualmente cultivados em várias partes do mundo têm base genética estreita, o que restringe a introdução de novas características (Parveez *et al.*, 2000).

Hoje, um caminho reconhecido para contornar as principais ameaças como o amarelecimento fatal e murcha vascular é o de explorar a resistência demonstrada pela espécie americana, *E. oleifera*, conhecida no Brasil como Caiuá (Hardon & Tan, 1969; Hardon, 1969; Meunier & Hardon, 1976; Meunier, 1976; Hartley, 1988) que é transmitida para os híbridos interespecíficos obtidos do cruzamento com o dendê. O caiuá também apresenta



resistência a outras pragas e doenças e características de interesse para o melhoramento do dendezeiro, como reduzida taxa de crescimento do tronco e alta taxa de ácidos graxos insaturados (Meunier, 1975; Barcelos *et al.*, 2001). Por isso, o melhoramento interespecífico entre o caiaué e o dendezeiro é uma prioridade em diversos programas de melhoramento do dendezeiro. Nos últimos cinco anos, vários experimentos foram instalados na Embrapa para avaliação de híbridos interespecíficos utilizando linhagens elite de dendezeiro e plantas de caiaué de diferentes origens da Amazônia brasileira. Para a diversificação da base genética das variedades comerciais de dendê, acessos de diferentes regiões da África, centro de origem de *E. guineensis*, estão em avaliação.

Tipicamente, os experimentos de melhoramento genético do dendezeiro tem elevado custo, ocupando grandes áreas devido a baixa densidade de plantio, com as plantas passando por uma fase de formação de três anos, que é improdutiva, um período longo de avaliação, do quarto ao décimo ano após o plantio, e uma produção contínua ao longo do ano (sem safra ou período definido de colheita). Para redução dos custos e aumento na eficiência da avaliação e seleção dos cruzamentos ou plantas superiores, estratégias novas de seleção assistida são necessárias, e poderão ser desenvolvidas a partir de informações geradas por estudos não convencionais de melhoramento, como a análise de dados genômicos e de mapeamento de genes utilizando marcadores moleculares.

### **5.3 Objetivos do isolamento de RGAs**

Embora o dendê seja, atualmente, a oleaginosa de maior rendimento em escala mundial, o progresso no melhoramento genético nesta cultura perene tem sido relativamente lento em comparação com outras culturas oleaginosas anuais, como soja e colza. O melhoramento genético visando resistência a doenças ainda está em fase de amadurecimento, focando, principalmente, a identificação de novas fontes de resistência às doenças de importância para a cultura como a podridão causada por *Ganoderma* (Gasselin *et al.*, 2005) e a murcha de *Fusarium* (Sekou *et al.*, 2010), e o teste de métodos para a transferência efetiva de resistência em variedades comerciais.

Neste sentido, a caracterização de RGAs é um importante e eficiente caminho para a identificação de *loci* putativos de resistência em espécie de genomas ainda não seqüenciados. Porém, a caracterização RGAs em *Elaeis* é, até o momento, limitada, com apenas quatro seqüências de possíveis proteínas de resistência a doenças da classe NBS-LRR depositados no Genbank (acessado em dezembro de 2011). Neste contexto, os objetivos deste estudo foram

isolar NBS-LRR RGAs e explorar a sua diversidade em acessos de *E. guineensis* e *E. oleifera* atualmente utilizados em programas de melhoramento locais, avaliando materiais que representam o centro de origem de cada espécie e que contrastam em resistência à murcha de *Fusarium*. Utilizando *primers* degenerados previamente desenhados para amplificar NBS-LRR RGAs em genomas de monocotiledôneas (Miller *et al.*, 2008) e *primers* degenerados desenhados para amplificar NBS-LRR RGAs em genomas de dicotiledôneas, mas efetivos na amplificação em ambas classes de plantas (Peñuela, 2008; Miller *et al.*, 2008), relatamos a primeira análise em grande escala de NBS-LRR RGAs no gênero *Elaeis*. Similaridades entre as seqüências foram determinadas e considerável polimorfismo foi observado. Trata-se de um importante estudo e grande potencial pode ser explorado, visto que os RGAs aqui descritos perfazem, provavelmente, parte do potencial destes análogos no genoma de *Elaies*.

#### 5.4 Caracterização de RGAs

A metodologia baseada em PCR para a descoberta de RGAs no gênero foi eficaz, aproveitando a relativa conservação do domínio NBS em *Elaeis*. Um total de 54 NBS-LRR RGAs foram isoladas, com divergência considerável observada ao nível de DNA e proteínas.

Embora 19 seqüências contivessem códons de parada precoces, sendo, provavelmente, pseudogenes, que co-amplificaram como resultado da degeneração dos *primers*, 39 seqüências não redundantes continham ORFs contíguas. Todas estas seqüências tiveram resultado positivo para o domínio NB-ARC, utilizando o Genewise, e codificavam proteínas com os motivos esperados. A presença de ORFs ininterruptas em diversos *contigs* de RGAs é relevante, pois, provavelmente, pelo menos alguns destes RGAs derivam de genes funcionais. Este trabalho é uma contribuição considerável para o gênero, embora o número de seqüências NBS-LRR caracterizadas provavelmente sejam ainda em número muito menor do que há na realidade, uma vez que genes NBS-LRR compõem uma parte significativa de genomas de plantas. No genoma de *A. thaliana*, por exemplo, com 172Mb (Bennett e Leitch, 2005) 149 genes NBS-LRR foram descritos, o que representa aproximadamente 1% do genoma (Meyers *et al.*, 2005). Da mesma forma, *O. sativa* (402 Mb) (International Rice Genome Sequencing Project, 2005) contém cerca de 600 seqüências NBS-LRR, que por sua vez representa cerca de 1% do genoma (Goff *et al.*, 2002; Meyers *et al.*, 2003; Monosi *et al.*, 2004). Dado que o genoma haplóide de *E. guineensis* é de aproximadamente 1,8 bilhões de pares de base, quatro vezes maior do que o de arroz, e a probabilidade de *E. oleifera* ter um tamanho similar de genoma, dado que ambos, *E. guineensis* e *E. oleifera*, têm o mesmo número de cromossomos

( $2n = 2x = 32$ ), a proporção de NBS-LRR RGAs no gênero *Elaeis*, fazendo um paralelo com o que se observa em arroz e outras monocotiledôneas, provavelmente representa apenas uma pequena parte do seu potencial. O baixo nível de redundância encontrado nos NBS-LRR RGAs também é um indicativo de que estas sequências representam um pequeno subconjunto de RGAs nos dois genomas.

## 5.5 Pseudogenes

Genes de resistência em plantas costumam ocorrer em *clusters* de genes funcionais, pseudogenes, fragmentos de genes e elementos transponíveis (Lescot *et al.*, 2004; Wei *et al.*, 2002). A presença de pseudogenes provavelmente resulta da recombinação inter-alélica e da conversão gênica, que mudam as combinações e orientações das sequências, especialmente no domínio LRR. Mutações pontuais, inserções e deleções também podem ter um papel importante na recombinação alélica de genes de resistência e pseudogenes, de modo que todos estes mecanismos, provavelmente, estão relacionados com a evolução divergente dos genes e com a co-evolução existente entre genes de resistência das plantas e os fatores *Avr* dos patógenos (Michelmore & Meyers, 1998). A co-amplificação de pseudogenes parece ser comum quando se usa *primers* degenerados na busca por RGAs. Martinez Zamora *et al.* (2004), por exemplo, identificaram 51 RGAs em morango, dos quais 23 possuíam códons de terminação precoces ou *frameshifts*. De forma similar, Liu e Ekramoddoullah (2003) encontraram 18 pseudogenes em um total de 46 RGAs em *Pinus*.

## 5.6 Conjuntos de *primers* e a eficiência apresentada

No geral, foi observada uma diversidade considerável entre os NBS-LRR RGAs isolados em ambas espécies. Esta diversidade provavelmente contribui para o *fitness* no gênero *Elaeis*, dada a constante pressão evolutiva sobre genes de resistência e os correspondentes genes de avirulência do patógeno. Duas combinações foram testadas: a combinação dos *primers* P1B-*forward* (Bertioli *et al.*, 2003), que foi originalmente desenhado para o motivo P-loop em sequências de dicotiledôneas, e RNBSD-*reverse* (Peñuela *et al.*, 2003), que anela no motivo RNBS-D não-TIR, e a combinação dos *primers* 3F2-*forward* e 13R1-*reverse*, que foram previamente desenhados por Miller *et al.* (2008), levando em consideração o número de degenerações, o tamanho do *primer*, a composição dos nucleotídeos, a posição da degeneração dentro de cada *primer* e a prevalência de alvos

putativos em seqüências NB-ARC em monocotiledôneas. Em estudos feitos com *Musa acuminata* estas combinações foram efetivas na amplificação de RGAs NBS-LRR não-TIR. As combinações P1A/P3A (Bertioli *et al.*, 2003) e 1F/P3B (Miller *et al.*, 2008), compostas por *primers* desenhados para outras dicotiledôneas, não foram capazes de amplificar produtos específicos ou RGAs, o que sugere que os motivos aos quais se dirigiam estão ausentes ou não conservados em nosso material de *Elaeis*.

A combinação P1B/RNBS-D foi menos eficiente, com 41% das seqüências apresentando homologia a RGAs NBS-LRR. Isto pode ter ocorrido devido a inespecificidade do *primer* P1B, desenhado para o motivo P-loop de dicotiledôneas. Em contrapartida, 59% das seqüências amplificadas pela combinação 3F2/13R1 mostrou homologia a RGAs NBS-LRR. Como esta combinação de *primers* gera *amplicons* que se estendem do motivo *kinase-2* do domínio NBS até motivos conservados dentro do domínio LRR, a eficiência na amplificação de seqüências de genes relacionados à resistência a doenças é potencialmente maior, pois, como a literatura relativa ao assunto mostra, o domínio NBS está presente em genes de resistência em plantas e também em genes que quinases ou enzimas ATP/GTP dependentes, mas proteínas contendo ambos domínios, NBS e LRR, apenas estão descritas como genes de resistência a doenças e análogos destes genes.

## 5.7 Origem das seqüências de RGAs NBS-LRR

Embora um número considerável de RGAs NBS-LRR não-TIR tenha sido caracterizado neste estudo, uma análise mais profunda do gênero requer o uso de combinações adicionais de *primers*, a caracterização de germoplasma e seqüenciamento de mais *amplicons*.

Em muitos casos, seqüências idênticas dos membros dos *contigs* de RGAs foram amplificadas a partir de diferentes acessos e mesmo das duas diferentes espécies. A amplificação de RGAs idênticos a partir de diferentes materiais sugere que as seqüências são originárias de regiões ortólogas. Da mesma forma, a redundância na amostragem do mesmo *locus* dentro de um único acesso também foi comum, com numerosas seqüências provenientes de clones múltiplos do mesmo acesso, amplificados com o mesmo par de *primers*. Por fim, também foi observado o isolamento de RGAs distintos provindos de um único acesso e da amplificação de uma única combinação de *primers*, o que provavelmente surgiu a partir de amostragem de diferentes *loci* de RGAs ou diferentes alelos do mesmo *locus* de RGA.

Poucas seqüências de RGAs NBS-LRR foram isoladas do acesso C25 de *E. guineensis*, material caracterizado como suscetível, o que pode indicar diferenças na abundância e/ou conservação dos motivos conservados de NBS-LRR nestas plantas.

## 5.8 Motivos conservados

Todas as seqüências que foram traduzidas visando a busca de motivos funcionais de aminoácidos característicos das regiões NBS (Figura 13) pertenciam a subfamília não-TIR. Este resultado está de acordo com estudos anteriores que relataram a ausência ou a raridade desta subfamília em monocotiledôneas (Miller *et al.*, 2008; Tarr e Alexander, 2009). O resíduo final do motivo *kinase-2* é especialmente útil para a classificação de uma seqüência como TIR ou não-TIR, com o motivo *consensu* LLVLDDVW esperado em seqüências não-TIR (Meyers *et al.*, 1999). Apenas duas seqüências mostraram um resíduo diferente nesta região, embora este fato não possa classificá-las como pertencentes à subclasse TIR, pois não contém os aminoácidos típicos para esta subclasse. As seqüências de aminoácidos também mostraram outros motivos característicos de regiões NBS, ou seja, P-loop, *kinase-2*, RNBS-C, GLPL, RNBS-D, com divergência considerável sendo observada. Estas variações indicam a grande variabilidade no gênero, pois mesmo dentro dos motivos conservados houve tal variação.

Apesar de apenas seqüências de RGAs NBS-LRR não-TIR terem sido isoladas neste estudo, de acordo com a hipótese de que a subfamília TIR seja restrita a dicotiledôneas (Pan *et al.*, 2000), a existência do motivo TIR já foi relatado no genoma de arroz, embora em número reduzido (Bai *et al.*, 2002; Meyers *et al.*, 2002). A ausência, até o momento, da subfamília TIR em *E. guineensis* e *E. oleifera* pode refletir as limitações na amplificação via PCR ou mesmo a sua inexistência. Porém, análises mais extensas devem ser feitas para tal confirmação.

## 5.9 Análise de similaridade

### 5.9.1 Biolayout

Relações de similaridade (*E*-value  $1E^{-05}$ ) entre as seqüências protéicas previstas para RGAs NBS-LRR isolados de *E. guineensis* e *E. oleifera* e genes de resistência ou RGAs de *A. thaliana* foram determinadas com base em conjuntos de dados de Blastx que foram

transformados em escores ponderados de semelhança na forma de *E-values* de BLAST. Com a utilização do programa BioLayout Express3D, dois *clusters* principais foram formados, bem como um *clusters* menores e proteínas *outliers*.

No *cluster* 1, que abrigou 41 seqüências, não foi possível verificar em seus membros a influência de características como resistência a *Fusarium*, *primers* utilizados e proveniência dos acessos, embora as seqüências apresentassem similaridade a uma seqüência de NBS-LRR RGA de *E. guineensis* depositada no Genbank (AF197922.1). O *cluster* 2 somente foi formado por acessos resistentes e possuíam em comum o motivo GLPL. O agrupamento deste *cluster*, que evidencia a característica resistência a *Fusarium*, sugere maiores estudos e caracterizações destas seqüências e acessos utilizados, visando a busca, por exemplo, de marcadores moleculares baseados nestes RGAs, o que seria de grande utilidade no melhoramento genético do gênero.

### 5.9.2 Neighbor joining

O agrupamento baseado em Neighbor joining das seqüências protéicas de RGAs NBS-LRR de *E. guineensis* e *E. oleifera* revelou 5 *clusters* principais. Embora esta análise tenha revelado grande diversidade, não mostrou sentido óbvio em relação ao estresse biótico, visto que os agrupamentos não seguiram um padrão visível de conservação de motivos, *primers* ou resistência a *Fusarium*. Os limites estabelecidos para a análise podem ter influenciado este não agrupamento, o que diverge com o que foi observado na análise feita por Biolayout. Para a confirmação da grande diversidade existente nas seqüências, nova análise foi feita levando em consideração, desta vez, as regiões entre os motivos conservados. O resultado foi ainda maior divergência entre as seqüências e a diminuição do número de clados. Um próximo passo poderia ser uma análise filogenética com *bootstrapping*, o que aumentaria a confiabilidade da análise.

## 6. CONCLUSÃO

No contexto da limitada diversidade genética existente em cultivares comerciais de *E. guineensis*, a ameaça contínua de patógenos em evolução, juntamente com o lento progresso na descoberta de genes e confirmação de suas funções em *Elaeis*, a abordagem baseada em PCR e utilizada neste estudo para a descoberta de RGAs NBS-LRR pode contribuir potencialmente para o melhoramento genético através de seleção assistida por marcadores e engenharia genética. Este estudo representa o primeiro grande isolamento e caracterização de RGAs NBS-LRR no gênero *Elaeis*, com amostragem de ambos *E. guineensis* e *E. oleifera*, a partir de materiais comerciais e de germoplasma do centro de origem de cada espécie, as fontes que provavelmente oferecem resistência a diversos estresses bióticos.

Um total de 365 seqüências apresentaram similaridade significativa para genes de resistência NBS-LRR e seqüências de RGAs conhecidos. A montagem de grupos de seqüências de alta qualidade gerou um total de 60 seqüências contíguas não redundantes, entre *clusters* e *singletons*, das quais 54 foram identificadas como RGAs NBS-LRR, apresentando o domínio NB-ARC e codificando proteínas com o os motivos de aminoácidos esperados para o domínio NBS. Este resultado, apesar de animador, não cobre a grande possibilidade de caracterização de RGAs no gênero *Elaeis*, visto que perfazem uma parte do provável total de RGAs em dendê.

A maioria dos genes de resistência já identificados codificam proteínas pertencentes à família NBS-LRR, que possui motivos conservados ao longo de sua seqüência. O alinhamento das sequências geradas, com base nestes motivos, confirmou a conservação dos motivos de NBS. Apesar da grande conservação do domínio NBS, o alinhamento múltiplo mostrou grande diversidade de sequências. Tal fato pode ser explicado com base nas regiões das sequências que estão situadas entre os motivos que não considerados críticos para funções bioquímicas, sendo com isso, alvo de grandes mutações. A abordagem, porém, foi eficiente.

Um total de 39 sequências continham ORFs contíguas. Estas seqüências podem ser consideradas candidatos a genes de resistência funcionais, conferindo resistência a diversos patógenos. Além disso, o ensaio de Biolayout mostrou a formação de um *cluster* com potenciais seqüências candidatas a genes de resistência, pois eram provenientes de acessos resistentes a *Fusarium oxysporum* f.sp. *elaeidis*. Estas seqüências poderão, potencialmente, ser utilizadas em programas de seleção assistida por marcadores moleculares.

São necessários mais ensaios em busca de RGAs no gênero *Elaeis*. A potencial aplicação de dados como os obtidos neste estudo na identificação de marcadores moleculares,

mapeamento físico e construção de bibliotecas BAC, análise de expressão, transformação de plantas e, por fim, no melhoramento genético de *Elaeis*, exige o esgotamento exaustivo da caracterização de RGAs.



## 7. PERSPECTIVAS

### 7.1 Construção de bibliotecas BACs

*Clusters* de genes de resistência e RGAs NBS-LRR como resultado de duplicações em tandem de seqüências parálogas parecem ser comuns em genomas de plantas (Ramalingam *et al.*, 2003; Hulbert *et al.*, 2001; Meyers *et al.*, 2003). Dentro destes *clusters* de RGAs, numerosos genes de resistência podem estar presentes e conferindo resistência a diferentes cepas de um patógeno ou a diversos patógenos (van der Vossen *et al.*, 2000). A formação de *clusters* de genes de resistência e de RGAs também pode facilitar a variação genética para a evolução de novos genes de resistência com nova especificidade. Dado que 54 RGAs NBS-LRR foram identificados neste estudo, a identificação de clones BAC que possuam seqüências NBS-LRR proporcionaria uma rota para a identificação e caracterização de *loci* putativos de resistência no genoma de *Elaeis*. Este estudo não fornece uma informação clara de quantos *loci* não redundantes foram amostrados entre os acessos. Pesquisas que promovam o mapeamento físico e a caracterização de seqüências poderiam, sem dúvida, aumentar nossa compreensão sobre a organização e a evolução de genes de resistência NBS-LRR nas duas espécies. Dado os inúmeros projetos de seqüenciamento em *Elaeis*, seja do genoma completo, seja de genes de interesse, após a liberação dos dados para o domínio público, o mapeamento físico dos RGAs caracterizados no presente estudo e a identificação de seqüências NBS-LRR ortólogas serão, então, possíveis.

Em relação à clonagem de genes de resistência, embora o isolamento de clones BAC seja facilitado pela presença de múltiplas RGAs em um locus de gene de resistência, é complexo fazer distinção entre um candidato a gene de resistência e seqüências não funcionais, com implicações no desenvolvimento de marcadores genéticos específicos e mapeamento de genes de resistência (Xiao *et al.*, 2007).

### 7.2 Análise de expressão gênica

Proteínas NBS-LRR atuam, hipoteticamente, como receptores pré-formados de proteínas em plantas resistentes e reconhecem efetores específicos do patógeno em uma correlação gene-a-gene (Van Der Biezen e Jones, 1998). Uma série de pesquisas confirmaram que genes de resistência NBS-LRR são expressos constitutivamente. Esta hipótese pode ser confirmada por técnicas como *Northern blot* e RT-PCR, e várias pesquisas mostram os baixos

e constantes níveis de transcrição, em plantas não infectadas, de genes de resistência como RPM1 (Grant *et al.*, 1995), PRF (Salmeron *et al.*, 1996), RPP5 (Parker *et al.*, 1997), Mi (Milligan *et al.*, 1998), I2 (Mes *et al.*, 2000) e transcritos de RGAs NBS-LRR em *Musa acuminata* (Peraza-Echeverria *et al.*, 2008). Este último trabalho tem também relatou a ausência de expressão gênica em acessos suscetíveis à raça 4 de *Fusarium oxysporum* f. sp. *cubense*, o que sugere, possivelmente, o papel de RGAs na resistência de plantas à doenças. Observações semelhantes também foram descritas para a resistência a *Fusarium* em tomate (Mes *et al.*, 2000). A expressão diferencial de genes em relação à presença do patógeno também tem sido observada (Radwan *et al.*, 2004; Gao *et al.*, 2010), bem como a indução de resistência por meio da produção de ácido salicílico e flagelina (Tan *et al.*, 2007).

Diante desta grande diversidade de RGAs NBS-LRR caracterizados no presente estudo, mais pesquisas com ênfase na expressão de genes de resistência durante a interação com patógenos devem ser realizadas. Dado que foram formados *clusters* de RGAs exclusivos de acessos de *Elaeis* resistentes a *F. oxysporum* f.sp. *elaeidis*, estudos posteriores podem investigar se estas numerosas seqüências candidatas estão envolvidas na resposta de resistência.

### 7.3 Marcadores genéticos e co-segregação com QTLs

A primeira aplicação de marcadores moleculares e mapeamento genético no melhoramento de dendê foi descrita por Mayes *et al.* (1997) utilizando marcadores RFLP. Mapas subseqüentes foram gerados usando marcadores AFLP (Moretzsohn *et al.*, 2000), SSRs (Billotte *et al.*, 2005), e marcadores RFLP baseados em cDNA (Singh *et al.*, 2008). Poucas análises de QTLs foram realizadas até o momento. Rance *et al.* (2001), por exemplo, indentificou QTLs para a produção de frutos e crescimento vegetativo e publicaram o primeiro mapa genético na espécie. Singh *et al.* (2009) relatou QTLs para a qualidade de óleo em um novo mapa genético de referência.

Sabendo que genes que conferem resistência quantitativa podem apresentar homologia a genes de resistência (Lefebvre e Chevre, 1995), uma série de RGAs têm sido mapeados com êxito gerando QTLs associados com resistência a doenças (Kanazin *et al.*, 1996; Ramalingam *et al.*, 2003). A análise de marcadores RGAs NBS-LRR (incluindo os de RGAs expressos) em mapas genéticos derivados de populações de *Elaeis* segregantes para resistência a estresses bióticos possibilitaria estabelecer uma ligação entre RGAs e *loci* de genes de resistência.

Marcadores co-segregantes poderiam ser instrumentos valiosos em programas de seleção assistida por marcadores na busca por resistência a doenças.

#### **7.4 Confirmação de função *in planta***

A transformação genética de plantas é um componente importante na descoberta e caracterização funcional de genes. A escala de dados genômicos que agora está se tornando disponível no gênero *Elaeis*, tais como seqüências genômicas, conjuntos de dados EST e famílias de genes específicos, como os RGAs apresentados neste estudo, fazem com que aplicação de sistemas de transformação eficientes para o gênero sejam uma prioridade para a determinação da função de alguns genes. Os métodos de transformação por biobalística e *Agrobacterium* foram testados nos últimos anos, buscando o desenvolvimento de resistência à doenças, como a podridão causada por *Ganoderma*, e pragas, como *Metisa plana* (Ismail *et al.*, 2010).

#### **7.5 Avanços recentes no conhecimento da imunidade em plantas**

O desenvolvimento de novas técnicas de proteção de culturas exige avanços contínuos na compreensão de como os mecanismos moleculares controlam a imunidade da planta. Ferramentas moleculares e genômicas possibilitaram avanços neste campo e duas linhas de ação fundamentais na imunidade de plantas a patógenos são conhecidas atualmente. Em um deles, de defesa inata da planta (*PAMP-triggered immunity* ou PTI), há o reconhecimento de padrões moleculares conservados associados a patógenos (PAMPs). No entanto, patógenos bem sucedidos podem suprimir a sinalização PTI por meio de determinadas proteínas efetoras. As plantas, em resposta, por meio da co-evolução, desenvolveram receptores de proteínas citoplasmáticos que reconhecem efetores do patógeno por meio de um mecanismo conhecido como *effector-triggered immunity* (ETI), a segunda linha de ação do sistema imune das plantas. No entanto, o desenvolvimento de resistência específica e durável depende de uma compreensão mais completa da imunidade das plantas. A relação entre PTI e ETI, por exemplo, continua obscura e o papel dos componentes de defesa das plantas carece de maior elucidação. Além da necessidade de uma caracterização mais completa de genes de resistência em plantas, para o qual este estudo faz uma contribuição específica na forma de caracterização de RGAs no gênero *Elaeis* no contexto de resistência à murcha de *Fusarium*, se requer também uma compreensão mais detalhada das especificidades dos motivos de genes

de resistência e sua interação com os domínios das proteínas efetoras do patógeno. Isto, provavelmente, possibilitaria avanços na engenharia do reconhecimento de patógenos, elucidando, por exemplo, por meio de reorganização de motivos LRR (Lucas, 2011).

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

- Aarts, M.G.M.; Hekkert, B.T.L.; Holub, E.B.; Beynon, J.L.; Stiekema, W.J. and Pereira, A. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interaction* 11(4): 251-258.
- Adam, H.; Jouannic, S.; Escoute, J.; Duval, Y.; Verdeil, J-L; Tregear, J.W. (2005) Reproductive developmental complexity in the African oil palm (*Elaeis guineensis*, Arecaceae). *American Journal of Botany* 92(11): 1836-1852.
- Aist, J. R. Papillae and related wound plugs of plant cells. (1976) *Annual Review Phytopathology* St. Paul 14: 145-163.
- Akhunov, E. D.; Akhunova, A. R.; Dvorak, J. (2005) BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *A. tauschii*, the diploid ancestors of polyploidy wheat. *Theoretical and Applied Genetics* 111:1617-1622.
- Alptekin, E.; Canakci, M. (2008). Determination of the density and viscosities of biodiesel – diesel fuel blends. *Renewable Energy* 33: 2623-2630.
- Altenbach, D.; Robatzek, S. (2007) Pattern Recognition Receptors: from the Cell Surface to Intracellular Dynamics. *Molecular Plant Microbe Interaction* 20: 1031-1039.
- Amblard, P.; Noiret, J. M.; Kouamé, B.; Potier, F.; Adon, B. (1995) Performances comparées des hybrides interspécifiques et du matériel commercial *E. guineensis*. *Oléagineux Corps gras Lipides* 2: 335–340.
- Ameline-Torregrosa, C.; Wang, B. B.; O’Bleness, M. S.; Deshpande, S.; Zhu, H.; Roe, B.; Young, N. D.; Cannon, S. B. (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol* 146:5–21.
- Ammiraju, J. S.; Luo, M.; Goicoechea, J.L.; Wang, W.; Kudrna, D.; Mueller, C.; Talag, J.; Kim, H.; Sisneros, N. B.; Blackmon, B.; Fang, E.; Tomkins, J.B.; Brar, D.; MacKill, D.; McCouch, S.; Kurata, N.; Lambert, G.; Galbraith, D.W.; Arumuganathan, K.; Rao, K.; Walling, J. G.; Gill, N.; Yu, Y.; SanMiguel, P.; Soderlund, C.; Jackson, S. e Wing, R. A. (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Research* 16(1):140-147.
- Bai, J.; Pennil, L. A.; Ning, J.; Lee, S. W.; Ramalingam, J.; Webb, C. A.; Zhao, B.; Sun, Q.; Nelson, J. C.; Leach, J. E.; Hulbert, S. H. (2002) Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Research* 12(12): 1871-1884.
- Barcelos, E. (1986) Características genético-ecológicas de populações naturais de caiaué (*Elaeis oleifera* (H.B.K.) Cortés) na Amazônia brasileira. Instituto Nacional de Pesquisas da Amazônia, Fundação Universidade do Amazonas. Dissertação de mestrado.
- Barcelos, E.; Amblard, P. (1992) Oil palm breeding program at Embrapa / Brasil; Embrapa –

CPAA.

Barcelos, E.; Nunes, C. D. M.; Cunha, R. N. V. (2000) Melhoramento genético e produção de sementes comerciais de dendezeiro. In: Viégas, I. J. M.; Müller, A. A. (Ed.) (2000) A cultura do dendezeiro na Amazônia Brasileira. Embrapa Amazônia Oriental/Manaus: 145-174.

Bell, A. A. (1981) Biochemical mechanisms of disease resistance. *Ann Rev Plant Pathol* 32: 21-81.

Bennett, M. D.; Leitch, I. J. (2005) Nuclear DNA amounts in angiosperms: Progress, problems and prospects. *Ann. Bot.* 95: 45-90.

Bent, A. F.; Kunkel, B. N.; Dahlbeck, D.; Brown, K. L.; Schmidt, R.; Giraud, J.; Leung, J.; Stakawicz, B. J. (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* 265: 1856-1860.

Bent, A.F.; Mackey, D. (2007) Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology* 45: 399-436.

Bergamin Filho, A.; Amorim, L.; Camargo, L. E. A.; Rexende, J. A. M. (1997) Manual de Fitopatologia. Doenças de plantas cultivadas. *Agronômica Ceres*: 184-200.

Bertioli, D. J.; Leal-Bertioli, S. C. M.; Lion, M. B.; Santos, V. L.; Pappas, G.; Cannon, S. B. (2003) A large scale analysis of resistance gene homologues in *Arachis*. *Mol. Genet. Genomics* 270(1): 34-45.

Billotte, N.; Marseillac, N.; Risterucci, A. M.; Adon, B.; Brottier, P.; Baurens, F.C.; Singh, R.; Herra, A.; Billot, C.; Amblard, P.; Durand-Gasselin, T.; Courtois, B.; Asmono, D.; Cheah, S.C.; Rohde, W.; Ritter, E.; Charrier, A. (2005) Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet* 110: 754–765

Boller, T.; Felix, G. (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60: 379-406.

Büsches, R.; Hollricher, K.; Panstruga, R.; Guus, S.; Wolter, M.; Frijters, A. (1997) The barley Mlo gene: a novel control element of plant pathogen resistance. *Cell* 88: 695-705.

Cai, D. G.; Kleine, M.; Kifle, S.; Harloff, H. J.; Sandal, N. N.; Marcker, K. A.; KleinLankhorst, R. M.; Salentijn, E. M. J.; Lange, W.; Stiekema, W. J.; Wyss, U.; Grundler, F. M. W.; Jung, C. (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science* 275: 832-834.

Camarero, P. (1981) Cultura do dendê: produção, industrialização e aplicações. [S. l.: s. n.], 1981.

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(4): 548-562.

- Chia, G. S. (2009) Repetibilidade da produção de cachos de híbridos interespecíficos entre o caiaué e o dendezeiro. *Acta Amazonica* 39: 249-254.
- Chisholm, S. T.; Coaker, G.; Day, B.; Staskawicz, B. J. (2006) Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* 124: 803-814.
- Clark, S. E.; Williams, R. W.; Meyerowitz, E. M. (1997) The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* 89: 575-585.
- 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. *Mol. Plant Microbe Inter.* 11: 968-978.
- Conceição, H. E. O.; Müller, A. A. (2000) Botânica e morfologia do dendezeiro. In: Viégas, I. de J. M.; Müller, A. A. (Ed.) *A cultura do dendezeiro na Amazônia Brasileira*. Belém: Embrapa Amazônia Oriental/Manaus: Embrapa Amazônia Ocidental: 31-45.
- Corley, R. H. V.; Tinker, P. B. (2003) *The oil palm*. Oxford: Blackwell Science, 562 p.
- Crute, I. R.; Pink, D.A.C. (1996) Genetics and Utilization of Pathogen Resistance in Plants. *Plant Cell* 8(10): 1747-1755
- Danesh, D.; Peñuela, S.; Mudge, J.; Denny, R. L.; Nordstrom, H.; Martinez, J. P.; Young, N. D. (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. *Theoretical and Applied Genetics* 96: 196-202.
- Dangl, J. L.; Jones, J. D. G. (2001) Plant pathogens and integrated defence responses to infection. *Nature* 411: 826-833.
- De Franqueville, H.; Renard, J. L. (1990) Bilan de l'amélioration du niveau de la tolérance du palmier à huile à la fusariose. Evolution de la maladie sur la plantation. *Oléagineux* 45(10): 399-405.
- Deng, Z.; Huang, S.; Ling, P.; Chen, C.; Yu, C.; Weber, C.A.; Moore, G.A.; Gmitter, F.G. (2000) Cloning and characterization of NBS-LRR class resistance-gene candidate sequences in citrus. *Theoretical and Applied Genetics* 101(5): 814-822.
- Devos, K. M.; Ma, J.; Pontaroli, A. C.; Pratt, L. H.; Bennetzen, J. L. (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. *Proceedings of the National Academy of Sciences of the United States of America* 102(52): 19243-19248.
- Donald, T. M.; Pellerone, F.; Adam-Blondon, A. F.; Bouquet, A.; Thomas, M. R.; Dry, I. B. (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. *Theoretical and Applied Genetics* 104(4): 610-618.
- Dransfield, J.; Uhl, N. W.; Asmussen, C. B.; Baker, W. J.; Harley, M. M.; Lewis, C. E. (2005) An outline of a new phylogenetic classification of the palm family, *Arecaceae*. *Kew Bulletin*.

TSO Publication, Norwich, Uk

Ellis, J.; Jones, D. (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. *Curr Opin Plant Biol* 1: 288-293.

Emediato, F. L. (2009) Caracterização de genes análogos de resistência (RGAS) em cultivares de musa acuminata contrastantes em resistência à estresses bióticos. Dissertação de Mestrado.

Ferwerda, J. D. (1975) Ecophysiology of the african oil palm. In: Alvim, P. de. (Coord.). *Ecophysiology of tropical crops*. Ceplac 2:1-49.

Flood, J. (2006) A Review of Fusarium Wilt of Oil Palm Caused by *Fusarium oxysporum* f. sp. *Elaeidis*. *Phytopathology* 96(6): 660-662.

Freeman, B.C.; Beattie, G.A. (2008) An Overview of Plant Defenses against Pathogens and Herbivores. *The Plant Health Instructor* 226(1).

Gao, Y.; Xu, Z.; Jiao, F.; Yu, H.; Xiao, B.; Li, Y.; Lu, X. (2010) Cloning, structural features, and expression analysis of resistance gene analogs in tobacco. *Mol Biol Rep.* 37: 345-54.

Gentzbittel, L.; Mouzeyar, S.; Badaqui, S.; Mestries, E.; Vear, F.; DE Labrouhe, D.T.; Nicolas, P. (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. *Theoretical and Applied Genetics* 96(3-4): 519-525.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 79-92.

Göhre, V.; Robatzek, S. (2008) Breaking the barriers: Microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.* 46: 189-215.

Gomez-Gomez, L.; Boller, T. (2000) FLS2: An LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol. Cell* 5: 1003-1011.

Gonçalves, A. C. R. Dendezeiro (*Elaeis guineensis* Jacq.). In: Castro, P. R. C; Kluge, R. A. (Coord.) (2001) *Ecofisiologia de culturas extrativas: cana-de-açúcar, seringueira, coqueiro, dendezeiro e oliveira*. Cosmópolis: Stoller do Brasil: 95-112.

Gowda, B. S.; Miller, J. L.; Rubin, S. S.; Sharma, D. R.; Timko, M. P. (2002) Isolation, sequence analysis, and linkage mapping of resistance-gene analogs in cowpea (*Vigna unguiculata* L. Walp). *Euphytica* 126(3): 365-377.



Grant, M.; Godard, L.; Straube, E.; Ashfield, T.; Leward, J.; Sattler, A.; Innes, R.; Dangl, J. (1995) Structure of the Arabidopsis RPM1 enabling dual specificity disease resistance. *Science* 269: 843-846.

Grant, M. R.; Godiard, L.; Straube, E.; Ashfield, T.; Lewald, J.; Sattler, A.; Innes, R.; Dangl, J. L. (1995) Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* 269 (5225): 843-846.

Guimarães, P. M.; Proite, K.; Leal-Bertioli, S. C. M.; Bertioli, D. (2005) Análise in silico da expressão gênica diferencial de *Arachis stenosperma* inoculado com *Meloidogyne arenaria*. *Boletim de Pesquisa e Desenvolvimento/ EMBRAPA Recursos Genéticos e Biotecnologia* 85.

Hammond-Kosack, K. E.; Jones, J. D. G. (1996) Inducible plant defence mechanisms and resistance gene function. *The Plant Cell* 8: 1773-1791.

Hardon, J. J.; Rao, V.; Rajanaidu, N. (1985) A review of oil-palm breeding. In: G.E. Russell (Ed.), *Progress in Plant Breeding*,: 139-163.

Hartley, C. S. W. (2000) The oil palm. *A Cultura do Dendzeiro na Amazônia Brasileira*. Embrapa. p.213.

Hayden, H. L.; Carlier, J.; Aitken, E. A. B. (2003) Population differentiation in the banana leaf spot pathogen *Mycosphaerella musicola*, examined at a global scale. *Plant Pathology* 52(6): 713-719.

He, C. Y.; Tian, A. G.; Zhang, J. S.; Zhang, Z. Y.; Gai, J. Y.; Chen, S. Y. (2003) Isolation and characterization of a full-length resistance gene homolog from soybean. *Theoretical and Applied Genetics* 106(5): 786-793.

Hinchliffe, D. J.; Lu, Y.; Potenza, C.; Gopalan, C. S.; Cantrell, R. G.; Zhang, J. (2005) Resistance gene analogue markers are mapped to homeologous chromosomes in cultivated tetraploid cotton. *Theoretical and Applied Genetics* 110(6): 1074-1085.

Ho, C. L.; Kwan, Y. Y.; Choi, M. C.; Tee, S. S.; Ng, W. H.; Lee, K. A.; Lee, Y. P.; Ooi, S. E.; Lee, W. W.; Tee, J. M.; Tan, S. H.; Kulaveerasingam, H.; Syed, S. S. R.; Ong, M. (2007) Analysis and functional annotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics* 8: 381.

Holub, E. B (2001) The arms race is ancient history in Arabidopsis, the wildflower. *Nat Rev Genet* 2: 516-527.

Huettel, B.; Santra, D.; Muehlbauer, F.; Kahl, G. (2002) Resistance gene analogues of chickpea (*Cicer arietinum* L.): isolation, genetic mapping and association with a *Fusarium* resistance gene cluster. *Theoretical and Applied Genetics* 105(2-3): 479-490.

Hulbert, S. H.; Webb, C. A.; Smith, S. M.; Sun, Q. (2001) Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* 39: 285–312.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800.

- Ismail, I.; Iskandar, N. F.; Chee, G. M.; Abdullah, R. (2010) Genetic transformation and molecular analysis of polyhydroxybutyrate biosynthetic gene expression in oil palm (*Elaeis guineensis* Jacq. var *Tenera*) tissues. *Plant Omics Journal* 3: 18-27.
- Jeong, S.; Trotochaud, A. E.; Clark, S. E. (1999) The *Arabidopsis* CLAVATA2 gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell*. 11(10): 1925-34.
- Jin, H.; Cominelli, E.; Bailey, P.; Parr, A.; Mehrtens, F.; Jones, J.; Tonelli, C.; Weisshaar, B.; Martin, C. (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J.* 19: 6150-6161.
- Jones, D.; Jones, J. D. G. (1997) The role of leucine-rich repeat proteins in plant defenses. *Adv. Bot. Res. Adv. Plant Pathol.* 24: 89-167.
- Jones, J. D. G.; Dangl, J. L. (2006) The Plant Immune System. *Nature* 444: 323-329.
- Joyeux, A.; Fortin, M. G.; Mayerhofer, R.; Good, A. G. (1999) Genetic mapping of plant disease resistance gene homologues using a minimal *Brassica napus* L. population. *Genome* 42(4): 735-743.
- Kanazin, V.; Marek, L. F.; Shoemaker, R. C. (1996) Resistance gene analogs are conserved and clustered in soybean. *Proc Natl Acad Sci U S A* 93(21): 11746-11750.
- Proite, K. (2007) Busca de Genes Envolvidos na Resistência de Amendoim Silvestre ao Nematóide das Galhas (*Meloidogyne arenaria*).
- Keen, N. T. (1990) Gene-for-gene complementarity in plant-pathogen interactions. *Annu. Rev. Genet.* 24: 447-463.
- Kobe, B.; Kajava, A. V. (2001) The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* 11(6): 725-732.
- Kobe, B.; Deisenhofer, J. (1994) The leucine-rich repeat: A versatile binding motif. *Trends Bio. Sci.* 19: 415-421.
- Lacock, L.; Van Niekerk, C.; Loots, S.; Du Preez, F.; Botha, A. M. (2003) Functional and comparative analysis of expressed sequences from *Diuraphis noxia* infested wheat obtained utilizing the conserved Nucleotide Binding Site. *African Journal of Biotechnology* 2(4): 75-81.
- Lanaud, C.; Risterucci, A. M.; Pieretti, I.; N'Goran J. A. K; Fargeas, D. (2004) Characterisation and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.). *Molecular Breeding* 13(3): 211-227.
- 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. *Plant Cell* 7(8): 1195-1206.

Le Guen, V.; Amblard, P.; Omore, A.; Koutou, A.; Meunier, J. (1991) Le programme hybride interspécifique *Elaeis oleifera* x *Elaeis guineensis* de l'IRHO. *Oléagineux* 46: 479-487.

Lee, S. Y.; Seo, J. S.; Rodriguez-Lanetty, M.; Lee, D. H. (2003) Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. *Molecular Genetics and Genomics* 269 (1): 101-108.

Lefebvre, V.; Chevre, A. M. (1995) Tools for marking plant disease and pest resistance genes: a review. *Agronomie* 15: 3-19.

Leister, D.; Ballvora, A.; Salamini, F.; Gebhardt, C. (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nat. Genet.* 14: 421-429.

Lescot, M. S.; Rombauts, J.; Zhang, S.; Aubourg, C.; Mathé, S.; Jansson, P.; Rouz W.; Boerjan. (2004) Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements. *Theor. Appl. Genet.* 109: 10–22.

Li, J. M.; Chory, J. (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90: 929-938.

Liu, X.; Lin, F.; Wang, L.; Pan, Q. (2007) The in silico map-based cloning of Pi36, a rice coiled-coil-nucleotide-binding site-leucine-rich repeat gene that confers race-specific resistance to the blast fungus. *Genetics* 176: 2541-2549.

Liu, J.; Ekramoddoullah, J. A. K. M. (2003) Isolation, genetic variation and expression of TIR-NBS-LRR resistance gene analogs from western white pine (*Pinus monticola* Dougl. ex D. Don). *Mol. Genet. Genomics* 270: 432-441.

Lopez, C. E.; Acosta, I. F.; Jara, C.; Pedraza, F.; Gaitán-Solís, E.; Gallego, G.; Beebe, S.; Tohme, J. (2003) Identifying resistance gene analogs associated with resistances to different pathogens in common bean. *Phytopathology* 93(1): 88-95.

Lucas, J. A. (2011) Advances in plant disease and pest management. *The Journal of Agricultural Sciences* 149: 91-114.

Mago, R.; Nair, S.; Mohan, M. (1999) Resistance gene analogues from rice: cloning, sequencing and mapping. *Theoretical and Applied Genetics* 99(1-2): 50-57.

Marra, M. A.; Kucaba, T. A.; Green, E. D.; Brownstein, B.; Wilson, R. K.; McDonald, K. M.; Hillier, L. W.; McPherson, J. D. e Waterston, R. H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Research* 7: 1072-1084.

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

Mas, J. G.; Van Leeuwen, H.; Monfort, A.; De Vicente, M. C.; Puigdomènech, P.; Arús, P. (2001) Cloning and mapping of resistance gene homologues in melon. *Plant Science* 161(1):

165-172.

Matiello, R. R.; Ronzelli, P.; Puríssimo, C. (1999) Mecanismos de resistência: fatores biológicos, agrônômicos e genéticos. In: Curso de Manejo da Resistência de Plantas aos Herbicidas: 27-40.

Mayes, S.; Jack, P. L.; Marshall, D. F.; Corley, R. H. V. (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* 40: 116-122.

McDowell, J. M.; Woffenden, B. W. (2003) Plant disease resistance genes: Recent insights and potential applications. *Trends in Biotechnology* 21: 178-183.

Mepsted, R. (1992) Studies on Fusarium wilt of oil palm. Ph.D. thesis. University of Bath, UK.

Mepsted, R.; Flood, J.; Paul, T.; Airede, C.; Cooper, R. M. A. (1995) A model system for rapid selection for resistance and investigation of resistance mechanisms in Fusarium wilt of oil palm. *Plant Pathology* 44: 749-755.

Mes, J.; Van Doorn, A.; Wijbrandi, J.; Simons, G.; Cornelissen, B.; Haring, M. (2000) Expression of the Fusarium resistance gene I-2 colocalizes with the site of fungal containment. *Plant J* 23: 183-193.

Messing, J.; Bharti, A. K.; Karlowski, W. M.; Gundlach, H.; Kim, H. R.; Yu, Y.; Wei, F.; Fuks, G.; Soderlund, C. A.; Mayer, K. F.; Wing, R. A. (2004) Sequence composition and genome organization of maize. *Proceedings of the National Academy of Sciences of the United States of America* 101(40): 14349-54.

Métraux, J. P. (2001) Systemic acquired resistance and salicylic acid: current state of knowledge. *European Journal of Plant Pathology* 107: 13-18.

Meyers, B. C.; Kozik, A.; Griego, A.; Kuang, H.; Michelmore, R. W. (2003) Genome-wide analysis of NBS-LRR encoding genes in *Arabidopsis*. *Plant Cell* 15: 809-834.

Meyers, B. C.; Kaushik, S.; Nandety, R. S. (2005) Evolving disease resistance genes. *Curr. Opin. Plant Biol.* 8: 129-134.

Milach, S. C. K. (1998) Principais Tipos de Marcadores e suas Características. In: Milach, S. Marcadores Moleculares em Plantas. Porto Alegre: S. C. K. Milach: 17-28.

Miller, R. N.G.; Bertioli, D. J.; Baurens, F. C.; Santos, C. M. R.; Alves, P. C.; Martins, N. F. (2008) Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: isolation, RFLP marker development, and physical mapping. *BMC Plant Biol.* 8: 15.

Miller, R. N. G.; Bertioli, D. J.; Baurens, F.C.; Quirino, B. F.; Ciampi, Ana Y.; Santos, C. M. R.; Martins, N. F.; Souza Júnior, M. T.; Pappas Júnior, G. J. (2009) Understanding plant responses to biotic stress: ongoing research in musa. *Acta Horticulturae* 828: 255-272.

- Milligan, S.; Bodeau, J.; Yaghoobi, J.; Kaloshian, I.; Zabel, P.; Williamson, V. (1998) The root knot resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding site, leucine-rich repeat family of plant genes. *Plant Cell* 10: 1307–1319.
- Mindrinos, M.; Katagiri, F.; Yu, G. L.; Ausubel, F. M. (1994) The *Arabidopsis thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78(6): 1089-1099.
- Mohan, M.; Nair, S.; Bhagwat, A.; Krishna, T.G.; Yano, M.; Bhatia, C.R.; Sasaki, T. (1997) Genome mapping, molecular markers and markers-assisted selection in crop plants. *Molecular Breeding* 3: 87-103.
- Monosi, B.; Wissler, R. J.; Pennill, L.; Hulbert, S. H. (2004) Full genome analysis of resistance gene homologues in rice. *Theor. Appl. Genet.* 109: 1434–1447.
- Moretzsohn, M. C.; Nunes, C. D. M.; Ferreira, M. E.; Grattapaglia, D. (2000) RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet* 100: 63-70.
- Moretzsohn, M. C.; Ferreira, M. A.; Amaral, Z. P. S.; Coelho, P. J. A.; Grattapaglia, D.; Ferreira, M. E. (2002) Genetic diversity of Brazilian oil palm (*Elaeis oleifera* H.B.K.) germplasm collected in the Amazon Forest. *Euphytica* 124: 35-45.
- Nadeau, J. A.; Sack, F. D. (2002) Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* 296: 1697-1700.
- Noir, S.; Combes, M. C.; Anthony, F.; Lashermes, P. (2001) Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). *Molecular Genetics and Genomics* 265(4): 654-662.
- Nürnbergger, T.; Kemmerling, B. (2009) Pathogen-Associated Molecular Patterns (PAMP) and PAMP-Triggered Immunity, in *Annual Plant Reviews* 34.
- Pappas, G.; Miranda, R.; Martins, N.; Togawa, R.; Costa, M. (2008) SisGen: A CORBA-Based Data Management Program for DNA Sequencing Projects. In: *Data Integration in the Life Sciences*: 116-123.
- Parker, J.; Coleman, M.; Szabo, V.; Frost, V.; Schmidt, R.; Van der Biezen, E.; Moores, T.; Dean, C.; Daniels, M.; Jones, J. (1997) The *Arabidopsis* downy mildew resistance gene RPP5 shares similarity to the toll interleukin-1 receptors with N and L6. *Plant Cell* 9: 879-894.
- Parveez, G. K. A. (2000) Production of transgenic oil palm (*Elaeis guineensis* Jacq.) using biolistic techniques. In: Jain, S. M.; Minocha, S. C. (Eds). *Molecular Biology of Woody Plants*. Kluwer Academic Publishers 2: 327-335.
- Paul, T. (1995) Fusarium wilt of oil palm. Studies on resistance and pathogenicity. Ph.D. thesis. University of Bath, UK.
- Peñuela, S.; Danesh, D.; Young, N. D. (2002) Targeted isolation, sequence analysis, and physical mapping of non-TIR NBS-LRR genes in soybean. *Theoretical and Applied Genetics*

104(2-3): 261-272.

Peraza-Echeverria, S.; Dale, J. L.; Harding, R. M.; Smith, M.; Collet, C. C. (2008) Characterization of disease resistance gene candidates of the nucleotide binding site (NBS) type from banana and correlation of a transcriptional polymorphism with resistance to *Fusarium oxysporum* f.sp. *cubense* race 4. *Molecular Breeding* 22(4): 565-579.

Plocik, A.; Layden, J.; Kesseli, R. (2004) Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce and chicory. *Molecular Phylogenetics and Evolution* 31(1): 153-163.

Prendergast, A. G. (1957). Observations on the epidemiology of vascularwilt disease of the oil palm (*Elaeis guineensis* Jacq.). *J. W. Afr. Inst. Oil Palm Res.* 2: 148-175.

Radwan, O.; Mouzeyar, S.; Nicolas, P.; Bouzidi, M.F. (2004) Induction of a sunflower CC-NBS-LRR resistance gene analogue during incompatible interaction with *Plasmopara halstedii*. *Journal of Experimental Botany* 56: 567-575.

Ramalingam, J.; Vera Cruz, C. M.; Kukreja, K.; Chittoor, J. M.; Wu, J. L.; Lee, S. W.; Baraoidan, M.; George, M. L.; Cohen, M. B.; Hulbert, S. H.; Leach, J. E.; Leung, H. (2003) Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice. *Mol Plant Microbe Interact* 16: 14-24.

Reed, B. M.; Richardson, P. E.; Russell, C. C. (1979) Stems nematode infection of resistant and susceptible cultivars of alfalfa. *Phytopathology* 69: 993-996.

Renard, J. L.; de Franqueville, H. (1989). Oil palm vascular wilt. *Oleagineux* 44: 342-347.

Renard, J. L.; de Franqueville, H. (1991) Effectiveness of crop techniques in the integrated control of oil palm vascular wilt disease. *Oleagineux* 46: 255-265.

Renard, J. L.; Noiret, J. M.; Meunier, J. (1980). Sources and ranges of resistance to *Fusarium* wilt in the oil palms *Elaeis guineensis* and *Elaeis melanococca*. *Oleagineux* 35: 387-392.

Rivkin, M. I.; Vallejos, C. E.; McClean, P. E. (1999) Disease-resistance related sequences in common bean. *Genome* 42(1): 41-47.

Rossi, M.; Araujo, P. G.; Paulet, F.; Garsmeur, O.; Dias, V. M.; Chen, H. (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Molecular Genetics and Genomics* 269: 406-419.

Salmeron, J.; Oldroyd, E.; Rommens, C.; Scofield, S.; Kim, H.; Lavelle, D.; Dahlbeck, D.; Staskawicz, B. (1996) Tomato Prf is a member of a leucine-rich repeat class of plant disease resistance gene and lies embedded within the Pto kinase gene cluster. *Cell* 86: 123-133.

Shen, K. A.; Meyers, B. C.; Islam-Faridi, M. N.; Chin, D.; Stelly, D. M.; Michelmore, R. W. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. (1998) *Molecular Plant-Microbe Interaction* 11(8): 815-823.

- Shiu, S. H.; Bleecker, A. B. (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci USA* 98: 10763-10768.
- Shiu, S. H.; Karlowski, W. M.; Pan, R.; Tzeng, Y. H.; Mayer, K. F.; Li, W. H. (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* 16: 1220-1234.
- Shizuya, H.; Birren, B.; Kim, U. J.; Mancino, V.; Slepak, T.; Tachiiri, Y.; Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the United States of America* 89: 8794-8797.
- Singh, R.; Tan, S. G.; Panandam, J. M.; Rahman, R. A.; Cheah, S. C. (2008) Identification of cDNA-RFLP markers and their use for molecular mapping in oil palm (*Elaeis guineensis*). *AsPac J Mol Biol Biotechnol* 16(3): 53-63.
- Singh, R.; Tan, S. G.; Panandam, J. M.; Rahman, R. A.; Ooi, L. C. L.; Low, E. T. L.; Sharma, M.; Jansen, J.; Cheah, S. C. (2009) Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biol* 9: 114.
- Singh, R.; Zaki, N. M.; Ting, N. C.; Rosli, R.; Tan, S. G.; Low, E. T. L.; Ithnin, M.; Cheah, S. C. (2008) Exploiting an oil palm EST the development of gene-derived and their exploitation for assessment of genetic diversity. *Biologia* 63: 1-9.
- Somssich, I. E.; Hahlbrock, K. (1998) Pathogen defence in plants - a paradigm of biological complexity. *Trends Plant Sci* 3: 86-90.
- 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. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804-1806.
- Surre, C.; Ziller, R. (1969) *La palmera de aceite*. Barcelona: Ed. Blume: 11-28.
- Taguchi-Shiobara, F.; Yuan, Z.; Hake, S.; Jackson, D. (2001) The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. *Genes & Dev.* 15: 2755-2766.
- Tan, X.; Meyers, B. C.; Kozik, A.; West, M. A.; Morgante, M.; St Clair, D. A.; Bent, A. F.; Michelmore, R. W. (2007) Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. *BMC Plant Biol.* 23: 56.
- Tang, X.; Frederick, R. D.; Zhou, J.; Halterman, D. A.; Jia, Y.; Martin, G. B. (1996) Initiation of Plant Disease Resistance by Physical Interaction of AvrPto and Pto Kinase. *Science* 274(5295): 2060-2063.
- Timmerman-Vaughan, G. M.; Frew, T. J.; Weeden, N. F. (2000) Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). *Theoretical and Applied Genetics* 101(1-2): 241-247.

- Tomkins, J. P.; Yu, Y.; Miller-Smith, H.; Frisch, D. A.; Woo, S. S.; Wing, R. A. (1999) A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics* 99: 419-424.
- Totad, A. S.; Fakrudin, B.; Kuruvinashetti, M. S. (2005) Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). *Euphytica* 143(1-2): 179-188.
- Van der Biezen, E.; Jones, J. D. G. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences* 23: 454-456.
- Van der Hoorn, R. A. L.; Kamoun, S. (2008) From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell* 20: 2009-2017.
- Van der Hoorn, R. A. L.; De Wit, P. J.; Joosten, M. H. (2002) Balancing selection favors guarding resistance proteins. *Trends Plant Sci.* 7: 67-71.
- Van der Vossen, E. A.; van der Voort, J. N.; Kanyuka, K.; Bendahmane, A.; Sandbrink, H.; Baulcombe, D. C.; Bakker, J.; Stiekema, W. J.; Klein-Lankhorst, R. M. (2000) Homologues of a single resistance gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *Plant J.* 23: 567-576.
- Viégas, I. J. M.; Müller, A. A. A. (2000) *Cultura do Dendzeiro na Amazônia Brasileira*. Embrapa. p.213.
- Vorwerk, S.; Somerville, S.; Somerville, C. (2004) The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci* 9: 203-209.
- Walbot, V. (1985). On the life strategies of plants and animals. *Trends Genet.* 1: 165-169.
- Wang, W.; Wen, Y.; Berkey, R.; Xiao, S. (2009) Specific Targeting of the Arabidopsis Resistance Protein RPW8.2 to the Interfacial Membrane Encasing the Fungal Haustorium Renders Broad-Spectrum Resistance to Powdery Mildew. *Plant Cell* 11.
- Wang, S. C.; Pinckard, J. A. (1973) Cotton boll cuticle, a potential factor in boll rot resistance. *Phytopathology* 63: 315-319.
- Wardlaw, C. W. (1946) A wilt disease of oil palm. *Nature*: 158-256.
- Wei, F.; Wing, R. A.; Wise, R. P. (2002) Genome dynamics and evolution of the Mla (powdery mildew) resistance locus in barley. *Plant Cell* 14: 1903-1917.
- Wenkai, X.; Mingliang, X.; Jiuren, Z. H.; Fengge, W.; Jiansheng, L.; Jingrui, D. (2006) Genome-wide isolation of resistance gene analogs in maize (*Zea mays* L.). *Theor Appl Genet.* 113: 63- 72.
- Whitham, S. A.; Anderberg, R. J.; Chisholm, S. T.; Carrington, J. C. (2000) Arabidopsis RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. *Plant Cell* 12: 569-582.



- Whitham, S.; Dinesh-Kumar, S. P.; Choi, D.; Hehl, R.; Corr, C.; Baker, B. (1994) The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* 78(6): 1101-1115.
- Wulff, B. B. H.; Chakrabarti, A.; Jones, D. A. (2009) Recognition specificity and evolution in the tomato - *Cladosporium fulvum* pathosystem. *Molecular Plant-Microbe Interactions* 22: 1191-1202.
- Xiao, S.; Gao, W.; Chen, Q. F.; Ramalingam, S.; Chye, M.L. (2008) Overexpression of membrane-associated acyl-CoA-binding protein ACBP1 enhances lead tolerance in *Arabidopsis*. *Plant J* 54: 141-151.
- Xiao, S. Y.; Brown, S.; Patrick, E.; Brearley, C.; Turner, J. G. (2003) Enhanced transcription of the *Arabidopsis* disease resistance genes RPW8.1 and RPW8.2 via a salicylic acid-dependent amplification circuit is required for hypersensitive cell death. *Plant Cell* 15: 33-45.
- Xiao, W.; Zhao, J.; Fan, S.; Li, L.; Dai, J.; Xu, M. (2007) Mapping of genome-wide resistance gene analogs (RGAs) in maize (*Zea mays* L.). *Theor. Appl. Genet.* 115: 501-508.
- Yang, M.; Sack, F. D. (1995) The too many mouths and four lips mutations affect stomatal production in *Arabidopsis*. *Plant Cell* 7: 2227-2239.
- Yoshimura, S.; Yamanouchi, U.; Katayose, Y.; Toki, S.; Wang, Z. X.; Kono, I.; Kurata, N.; Yano, M.; Iwata, N.; Sasaki, T. (1998) Expression of Xa1 a bacterial blight-resistance gene in rice is induced by bacterial inoculation. *Proceedings of the National Academy of Sciences of the United States of America* 95(4): 1663-1668.
- Young, N. D. (2000) The genetic architecture of resistance. *Curr. Opin. Plant Biol.* 3: 285-290.
- Yu, Y. G.; Buss, G. R.; Saghai Maroof, M. A. (1996) Isolation of a super family of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proceedings of the National Academy of Sciences of the United States of America* 93(21): 11751-11756.
- Yüksel, B.; Estill, J. C.; Schulze, S. R.; Paterson, A. H. (2005) Organization and evolution of resistance gene analogs in peanut. *Molecular Genetics and Genomics* 274(3): 248-263.
- Zeven, A. C. (1964) On the origin of the oil palm. *Grana palynologica* 5: 121-123.
- Zhang, H. B.; Choi, S.; Woo, S. S.; Li, Z.; Wing, R. A. (1996) Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. *Molecular Breeding* 2: 11-24.
- Zhang, L. P.; Khan, A.; Niño-Liu, D.; Foolad, M. R. (2002) A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum* x *Lycopersicon hirsutum* cross. *Genome* 45(1): 133-146.
- Zhou, T.; Wang, Y.; Chen, J.; Araki, H.; Jing, Z.; Jiang, K.; Shen, J.; Tian, D. (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics* 271: 402-415.

## 9. APÊNDICES

### 9.1 Apêndice A: Lista de sequências de nucleotídios de *contigs* e *singletons*

>Contig1 (n=659)

CGGCCAAGTCGTGCATAATGTCATGCACCCGAAAATTTGCCTGCAGAAGATTTTCGCATAGTCAACTCTCTGTAGTACTCTTTT  
GCTGTCTCTTCCATTTGTCGAACCTCCCCTCAGCTTTCACAAAATCCCCTCAGCAATCCAAAAGTTTAGTAATATAATCCATTTCAATT  
TGATGGTCTTCGGGAAATAATGAGCAGTATACAAAACATTGTTTCAAATAGCTTGGTAGGTCTTCGTAACCTCAAATACAATGC  
TCCCATAACCTCCTCAGGAAGCCCAGATGTGGACCATGCGGTGCTTTCAGAAGCTTATTCCAAGCTCGCCTGGTCTTCTCCTT  
GGCGCATAGCACCCCTGCAACCACTTTAATGGCTAGAGGAAGGCCATCACATTTCTTAACAATCTCCATCCCTATGTCTTTTA  
GATGCTGGATCTCTCCCTCCTCACCTGTCAAGACCACCTTCTTGCAAAGCAATGACCAACCATCTCCTAAAGATAGCTTGTCC  
ACGTGATGGGTGTTCACTGCCATCATCTGCATCGCGATCTGTTTATTCTTGTAGTGACAAGGACCCTACTGCCAGCAGCACACA  
ACCATATAATGTGTTACGTAAACAAGTCGCACCAAACGTTTGGCGTCCAAAACATCGTCCAGCACCAGGAAGTACCTC

>Contig2 (n=716)

TTCAGTGTGATTCGGCCAAGTCGTGCATCAGGTCGTGCATGTAGTATAATTCATCATTAGAGTGTTCACCTCGTATAAAGAA  
AGACCTCTGCAGCAAATCATTGAAGTAGCTTCCCCTATATCCTCCAATTTTCGGGTGTTTATCCAGGTTTGAATGAACCCATG  
TGCCATCCACATATGGACCAACCTTTCTTTGATGAACTTGTAACCTTTGGGAAATATACTACAGTAAGCAAAGCAGCGCTGCA  
GATGTGCTGGTAGCCGCGATAGCTCAATCTCAAAGCTGGCATGATGTCATTAGGCCCTGTTCTAGTTTCCATAAATCACTC  
GTCAAAATGCTTGTCCAATAACTGTGATCAATGTTAGCACTTAGCAGTTTACCCACGGTCTTTGCTGCAAGAGGGTTGCCTTTC  
AGCTTCTTAGCAATCTCCATGCCAACAAGTTTAAATTCGGGCAACGAAATAAAGAAAACCTCTCGTTACCGAATGCACATGC  
TTCGAAAAATAACCAAAAATCCCTGTATGCTAAACCTCCAAGAAGATTTTCTCACGGTTCGACCCATCACATCTACACTCT  
TTTCTCTCTAGCCGTGATCAATATCTTGCTACCCTTTTCAGCATATGTTAGATAATCGCATAGTGCCTCCCATTTGGTATCCCA  
CACATCATCCAGCACCAGGAAGTACCTCNNNNNNNNNNNNNNNNNNNNNN

>Contig3 (n=411)

GAGGTAATCTCTGGTGTGGACGACGTTTGGTGGCTCTCTGATCAAAAGGCTGCAGGGATTTATCGAGAAGCTGCTGCTTCAG  
CTTCAGAAGATCTTACAGATCAGATCTAGATCTGAATTTTGGGAGCAATATTCATGAGGAGAAGACGATTCAGATCCTGCTTG  
AGTGGATACCGATAGAGACTAGGCAATTGTGTGGCTATTTTAGAACCTCGGATGTTGCTGCGATCATCTACAGGGTAATCTAG  
TACTCTAGAGGTATTTCTAATCTTAAAGTTTAAATTTAATTTTATAGATAAATAAGAATCAGATTTAATAG  
ACCTTAGATCAAAATAAAGTAGGATAATTTTTTTTTTAAATATTTACCCAGATCTTATTAGATCTGGAAGATCCTAC

>Contig4 (n=698)

GAGGTAATCTCTGGTGTGGATGATGTGTGGAATGAGGAGGAAGCTGAATGGGAAAGACTAAAACCTTTGCTGAGGGGTGGC  
AAGAAGGGAAGTAAGATCATAAGTACGACCCGCAAGTAAAGAGTTGCTAGCATAATGGGTAGCTCTGCACTACTCAGATTGC  
CAGTGTACCGACTGATGACTGCTGGACTTTGTTTCAGGCAGAGGGCGTTTCGGGTGGGAAGAGCTGAAGAGACGCCAAGCTT  
GGTAAAAATAGGGAAGAAGATTGTCCAGAAGTGTGGGGTCTGCCGTTAGCCGCAAAGGCTCTTGGAAAGCCTAATGAGCTCC  
AAGAGGGGGGAGGTGGAGTGGTTAGCTGTCAAAAATAGCGAAATCTGGAAACTACCTGCGAAGGAGACTGGAATATTACCT  
GCCTTAAGATTGAGTTATGATCATTGCCATCTCATCTGAAGCAGTGTTCCTACTGTTTATTGTTCCAAAAGATTATTCA  
ATTGAAAGAGAAAGATTGATTCAGTTGTGGATTGCTGAAGGGTTTATTAATCATCCATCAGATAATAATATGGATTTAGAGAA  
TATCGGCAATCAGTTTTTTAATAATTTGCTATGGAGGTCTTTTTTTCAGGACGCAGAAAAGGATTCTGATGGTAACGTGACGG  
TGTGCAAGATGCACGACCTGATGCACGACTTGGCCG

>Contig5 (n=426)

TCATAACTCAAAAATAAAGCTCCTGGGAGCTCTTTGTCAAGTTGGTTCATAGACCATGCATCACCTTTGAGAACCTTGCTCCA  
CTCCATGTTGGTTTTCTCCTTGAACCTCAAACCCCTGCGATAGTCTTGATCGCAAGAGGAAGACCGTCACATTTTTCAATAAT  
TTAACTCCAACCTCTTTTAAATCAGAGATCACCTCCTCAACAACATCATCTCAAAAATCTTTTACGTAGCATTCCCAGCC  
ACTCTCATTATCCATTTTTTAAACATGATGGATATCTGCTTCCGTATTTCCATCTCCTCTCATACTTCCAACCACATTTCGTATGA

CGAGTGGTAACCAATCTGCCACTAGCAGTTGCACGTTCAAAGAATTTCTAAGCAAATCTCCCATAAATCTGCGTTCA  
TACATCA

>Contig6 (n=630)

CGGCCAAGTCGTGACAAGGTCGTGCATCTTGCAAGTTGTTATGTAATAAACTTTACTGCTGCGTACAAAAATCCGTCGAACC  
CGGCATGTCCTCAACTTCTTGATGTCTGAAAGAAAGATCTCGACGCCAACTCATTAAAAATTCATGCCCCCTGTCTCCA  
ACTCCTTTCTTCCACCAGATGGAATGAATCCATTGGCCATCCATAGTTGAATCAACAAATCCTTTTCCATCTCATAATCCTTTG  
GAAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATGACTGTAGCTCAATCGTAGCGCTGGTAAAAATCGC  
GTCTTCCACCAACCTGCATATCCCAAATCTCACTATCCCTCACGAACAACCAATTCCCTTCTGGCTCTTGAATACAATAAGCC  
CCCCATTGTCTTCACTGCTAAAGGCAGTCCACGACATTTCTTGACAATCTCCTTGCCAATATTTACGAGATTCTGAGGCTCTTC  
TCTCCCTTTCAAATGCTCTCTGCCTAAACAATGTCCAGGAATCATCTTCAGTCAAATATGGTAGCTGATAGGTGGCGAGTG  
TACCCATCATCGACGATACCCGCTCACTTCTTGTGTCACAATGAT

>Contig7 (n=639)

GGAAGAGGCAACAATAGGGAAAAACAACACTTTCATCTCGTCCGTTTTCAACTGTTTGTAGCTCAGTTCTAGCGACTTGTACAA  
CTGTTCTCCACACCTTCTAGGTTTGCAGGAGTGGATTGTTGGAGCTGGAGCAGTGCATCATCCAGACACTTCTTCTTAAA  
TCTTCGCAATGCCCTTCCGACTACAACAATTGCCAGCGGCAAACCACCACACTCCTTTGCAACTTTTTTTCGACCGGTTCTAC  
ATCAGAGGTTTCAACAACCTTTCAGCCTTACTTTTAAAAAGGTCCCACGAATCTTCTTCTGACAACACATCTACTAGGACCG  
AAACATCAGTTTCCATCAAATCACATACTCAGATTTCCGAGAAAGTTAGTAAAAATTTTGCAACCTTGTGATCATCTGAAA  
GGAATGCCAATGTCTTTAACTCTAACCTAGCCATATGTATCCAGCATTATCGCAATTTCTTCTCCTTACCCAACCTAGCA  
GCCAGTGCCTAGCTGCAGATTCACTCCCCAGAGTGCATACCCAACCTCTCAGCAATGTCAGTCTTATCTGCTCTATG  
TTTGATTCTGTGATACGACCACTTTGACAACACCACCAAACCTTCTTATC

>Contig8 (n=740)

GGAAAGAGGCAACAGTAGAGAAAGCAGTTCGGTAAATGATACGATAGATCTCCAAAAGTACTGCTTCAATATATTTCTTACCC  
TTTCGGACATTGGCATATTAGTCTCTTCCCAATCGAGACCCTCACAGAATTTCTCCATTCCAACCTTCTCCCTCTCTCGGTATG  
ACAGCGAGCTTCTGCGGCCACAATAGCTAGAGGCAAGCGTGGCACCTCTGCAAGATTTTCTAGCCCAGTACTCTAGATCT  
GAAGGACAACCTTCTGTCTTATATTTCCAGAATGCTTTATTACGGAAGAGAATCCATGCGTCACCCTCTTGCAGAGGCTGCAG  
CTCCAGAACACGGCTTTCAACCGCTAGGTAAGCTACATCATGATTCCGAGTTGTATAACAATTTCTACTTCTACATTTACTATC  
TACCAGCACTGGACTTAATTCATCAAACCAACCAATTTATATGCCACATATCATCTAGGACGATCAGATACCTTTTGTGCA  
AATAAGCATGAAGTACCTCCGCCAAAGTGTGCAAGGCATTGTGTCAATGTTGTCTGGAAGAATATCACGATTTTCATTGTAG  
AGGTCTTCGATGAGTCTTCTATTAGATCTTCGGTTGTATTTCTGGGAAACGCGCACCCATGCTCTGCGATCAAATACCCG  
TTGACTTCTGGCTTCTGTAGATATCATTGACGAGAGTCGTCTTCCCCCCCCCCCCATCCC

>Contig9 (n=684)

CGGCCAAGTCGTGCATCATGTCGTGCATCACGTACTTGTGTTTCAGAGTTGGTAGAAAGAAGTACCTGCTGGCTAATTCATCA  
AATATCTGACCCCGATGTCTCCATTCTCATTCTTACTCGATTCTTGTCTGGATAAAAACCTTGAGCCATCCACATCTGGACC  
AACCGATCTTTTCGAAAAAATAGCTCCTGGGAAATATGGAACAATAAGCGAAGCATTGCTTGAGATTTGCGTTCAAGTGTT  
GTAACCTAAAGCAAGAGATGGAAGGATGTTATCTAAAACATATTATGTTCCACCACACTCACTTCTAGGATCATCTCCAGT  
AATCCTCATCTAGTTTGGACTTCAATAAACTTCCAGTACCTTTCCAGCCAGGGGTAAACCATGTAGTTTTTCAGATATTTCC  
TTCCTATTGTTTGAATTTTTCTTCTGCTCTTCCCTCGAGAGTGGCTTACCAAATGCACAATGTTTCGAAAAGTCTCCAGTGGT  
CATCTTCTCTAGGCTTTTCAATCTATCGGTTCCATCGTGCCATGACTTCTGCTACTACTGGACTTTGAGTTGTACCAAAAC  
CATGCTTCTTTCGCTCCGGATGTTAAGACACCACGGAGTCTTTCCACTTGTCTCCTGTCTCATCCACACATCATCCAGCAC  
CAGGAAGTACCTC

>Contig10 (n=688)

NNNNNNNNNNNNNNNNNNCGGCCAAGTCGTGCATCATGTCGTGCATCATGTAATATGTATGACCATCTTGCTCGAGTGTTG  
AAAGAAAGACCTTGACAATAAGTCATTGAAGTAGACTTTCCTATATCCTCTATCCTCACACTGTTCCAGCTTTGGGCTGAA  
TGAAACCCTGTGCCATCCACATGTAGACCAACCTATCTGGTTCAAATTTCCAGTCTTGGGAAATATACTGCAATATGCAAAG  
CAGCGCTGCAAATGTGCCGGAAGGTTTTGATAACTCAATCTCAAAACAGGCATAATGCCATCCCATATGTCACTTTTCAAAAC  
ATTCTCCATTTTCAAACATCCAATTCAGCATTACGATTCTCCACCACCTTTGCAGCCAAAGGTGAGCCATTTAGCCTCTC

TGCTATTTGCGTACCAATTCGTAATAACTCTAGATGATTATCAGGGTTCTCACAAACAAAGGCATGCTTTTGAAAAAGCAGCC  
AATAATCATCACCTCGCAACCCCTTCAAATAGATCGAATTCACAGCACCAACATCGTCACCACCATTTCCTTCTAGTTGTA  
GCCAAGATCTTGCTTCCCTTGCCCCAGATCTCAAAGGGGCGAGGAGCTTCCTCCATCTATCCCTGTTCTCCAGTCCTCCACG  
TTCTCAGCATTCCACACATCATC

>Contig11 (n=643)

CGGCCAAGTCGTGCACCAGGTCGTGCATCCGGAATATCGCTGCAGAAGATTCGCATAATCAACTCTCTGTAGTACTCTTCA  
GCTGTCTTCCATCGTCAAACCTCCCCTGATCAGCTTTCACAAATCCCTCAGCAATCCAAAGTTCAGTAATTTTCATGCATGCTG  
AATGTATAGTCTTCAGGAAATAATGAGCAATATAAAAAAGCATTGCTTCAGATAGCCTGGTAGGTCTTCATAAATCAAATACAG  
TGCTCCCTGAACATCCTCGGGAAGCCAGATGTGGACCATGCGGTGCTTCCAAGAAGCTCGATTCCAAGCACTCCTGGTCATCT  
CCTTCGTGCATAGCACCCCTGCAATCGCTTAAATGGCTAGAGGAAGGCCATCACACTTCTTAACAATCTCCATCCCTATGCCT  
CCAGATGCTGGATCTCTTCTCCTCCCCTGTCAAGACCACCTTCTTGCAAAGCAATAACCAGCTATCTTTGAAGACAGTTTGT  
TGATGTGATGGGTGTTCACTGCCCTCATCTGCATCGCGATCTGTTCTGTTTCTGTAGTGACAAGGATCTACTGCCAACAGCAC  
AACTTTTAAACGTATTGCATAACAAGTCACACCAGACCTTTGCCGTCCACACATCA

>Contig12 (n=721)

GAGGTAATCTCTGGTCTGGATGATGTGTGGGATACCAAATGGGAGGCACTATGCGATTATCTAACATATGCTGAAAAGGGT  
AGCAAGATATTGATCACGGCTAGAGAGGAAAAGAGTGTAGATGTGATGGGTCGAACCGTGAGAAAAATCTTCTGGAGGGT  
TAGCATAACAGGATTTTTGGTTATTTTCGAAGCATGTGATGATGTGTGGGATCCAGAAGCTTTAATGATATTGACAAGCC  
TTTGTAAACAATAAAGGAAGCAGAATAATCATTACAACCTCGGAGTACTGAGGTAGCTTACTAGCACATGATAGTCGGAAGC  
TTGAGCTGAAGGGTTGCAAAGCACTCAGGCATGGGATCTATTCTGTAAGAGGGCATTCTGAAAAGAGAAAAGTGAGAGAATG  
TCCTAAAGAGCTGGAGGAGTTGGGGGAGAAAATTGTATCAAAATGTCAAGGGTTGCCACTAGCTATCGTCTCTTTAGGCAGC  
CTGCTGTCCCTGAGAGAGAAAACCAAGTCAGAATGGGGGAAAGTTATGATCGGCTGAGCTGGGAGTTGAATCACAACCCAA  
ATTTGGACAATGTGAAGCATGTTTTGAATCTCAGCTACAACCTATCTACCGAGATATCTTAAGAAGTCTTCTTGCATTGCAGC  
ATGTTCCCGGAGGATCATTGCTTCAGAGAAAAGAGGCTTATGAGACTGTGGGTCCGAGAAGGA

>Contig13 (n=633)

NNNNNNNNNNNNNNNCGGCCAAGTCGTGCATAAGGTCGTGCATCAGCTTTAGCTTCTCAGAGAACCAATTTTGAGTTTTGT  
GCTTACTACATGCCAATCAAGAGAATGTTATATGATGGGCTTGACAGTCAATAACCTACAGGCAAATCTTGAAAAATGA  
AAATTATAATTGAAGAAAAAATTCAGTTCCATGCCCTATAGGATTTGTTGTCTTATAGCCCATCCAAAACAGTAGTTATA  
ACCAACCAATCTAATATCCACGTGATGATGAAATATTAGTCGACCATAAGACCTGAAACCTCCATAAAGCATGCAAGAGCA  
AGTCAATCCAAACATGCAATTGAGTAAAACATGTATGAGCTTGCTTGTCTTCTGGTTCCAAACGCCGTAAATGACTTAGCCA  
CAAAGCCTGTCATAGAAAATAAGGTACAAGAGAAACAGAGGCAAGAAAATGAAGAAGGTGCATGGGACTTATTAATACAT  
ACAATAACAAGTTACTTTAATTGTGAAGAAAATTAATGTCACCCATCAGACTTCAAGCTGTTTGTATGCTCGACATATGAAGT  
AGCCTGTCTAAAATAATCGCATTTCCACACATCATCCAGCACAGGAAGTACCTC

>Contig14 (n=476)

TCATTCTTCTTACAACCTGAAGCATGCATCGATGGATGAGCTCCTTGAGATAGTCTCTGCCACTTCTTCCATTGTGCTCGAT  
CCCCTTCTCCATAATTCCTCTGCGACCCACAGCCTCATAAGCCTCTTCTCTGAAGCACATGATCCTCCGGGAACATGCTG  
CAATGCAAGAAGCAGTTCTTAAGATATCTCGGTAGATAGTTGTAGCTGAGATTCAAAACATGCTTCACATTGTCCAAATTTGG  
GTTGTTATFCAACTCCAGCTCAGCCGATCATAAATTTCTGCCATTCTGACTTGGTTCTCTCTCAGGGACAGAAGGCTGCC  
TAAAGAGACGATAGCTAGTGGCAACCCTTGACATTTTGAACAATATCCTGCCCCAGCTTCTCAGCTCTCAGGACATTTCT  
ATCTTCTTCTTCCGGAATGCCTTCTTACAGAATAGATTCCATGCCTCTGTGCTTTC

>Contig15 (n=703)

NNNNNNNNNNNNNNNCGGCCAAGTCGTGCACCATGTCGTGCATCACGTACCTGTTATTTCAGAGTTGGTAGAAAAGAAGTAC  
CTGCTAGTTAATTCGTCAAATATCTGACTTCCAATGTCCTCCATTCTCATTCTTATCCGATTCTTGTCTGGATAAAACCTTGAG  
CCATCCACATCTGGACTAACTGATCTTGCTCGAAAACATAGCTCTTGGGAAATATGGAGCAGTAAGCGAAGCATTGCTTTAGA  
TTTGCCTTCAAGTGTTCGTAACCTAAAGCCAGAGATGGAAGGATATTATCTAAAACATATTGATGTTCCACCACACTCTTCT  
AAGATCATGTTCCAGGACTTCTCGTTTGTGGACTTCAATAAACTTCCCACTACCTTCCAGCTAGAGGTAATCCATGCAGT  
TTTTAGATATTTCTTCTTATTGTTGCAATTTTCTTCTGCTTCTCCTTAAAGATTGTTTCCACAAATACACAATGTTGAA

AAGTCTCCAGAAGTCATCTTCTGCAGACTTTCAATTCTACCGGGTCCATCGTGCCCATGACTTCTGCCACTATTGGACTTTG  
AGTTGTACCAGAACCATGCTTCTTTTGTCCAGATGTCAAGACACCACGGAGTTCTTTCCACTTGCTACCAGTCTCATCCCA  
AACATCGTCCAGCACCAGGAAGTACCTC

>Contig16 (n=658)

GTCGTGCATCAGGTCGTGCATTTTACCACATCGTCTTCGTTCAACCCAGGCTCTAGCAAACAAGCGGATTTTAGTTTTCTAT  
GATAACATGCCACTGTCATAGGCCTCGTTGATGTCTTGAAGCCTCCGACATCTATCAGCCCATGGCCCATCCAGCACTCGA  
TGAGATCAATTTTCTTTATAGAAAAATCTTCCGGCCACAATGCGCACAGCAAGAAACACTTTCTTACAGCATCATCTGGCAA  
GAATCATAGCTAAGCTTCAGTCTGGGAAGCATGTCTCGTCTCGTCTTGTCTGGGATCTCAGGAATCTTTGAGCTCTGCAGT  
AGTGTGATGGCATTGTTGCCACTCTTGCGAGTCTTCTCGTTGACATGGCCTTCCCGATGATTGTGAGAGCCAAGGGCAATCC  
GCGGCACTCCTCGACCCTGCTTGGCGAGCGACGTTCGGTATTCTTGTGAGTGTCTTGGCCGACCATATCTTGAACAAC  
TCCAGGAGTCTCTGGCTCCAAGCAGTTTATCTTGATCTTCTCGTGGGCTCCCATGGAGCCACACACCTGCTCCGTCCGTGTGG  
TAAACACCACCTTATGCTTGTGCTGGCCGGTTCGTTCTTGGATGGAATGGGAACCCCAACGGCCCTCCAAATC

>Contig17 (n=436)

TCGTGGCGAATCTGTGAATTATCAAGTCTTGGAAAGAAGGACCTTGACAATAGCTCTTGGAAAAATCTCATGGCCTTTTCAATTC  
CATGTCCTGCTCCGGATCAGATGGAATGAATCCATTAGCCATCCACAACCTGGATCAGCATCTCCTTGTCTATTTCATAATCCTT  
CGGAAATATGGCGCAATAAGCAAAGCATTGCTTACATGACAAGGCAGGTGATTATAGCTCAACCTCAATGAAGAGAGGATC  
TTATCATCATTTAGTTTCCAAATTCACCTTCCATCACTGATCTCCATTCCTCACCTCCATCTTCGTGCGCATTAAATCTCCCA  
GCATCTTCGCCGCCAGGGGCGAGGCTCCGCACTTATCCGCAATGGCTTTTCCAATGTGTACCAGATTTGCGGACTTCTCCGCA  
CCACCCTCAGAAAATGCGT

>Contig18 (n=152)

GGAAAGAGGCAGCAGTAGGGGAAACAGCGCTGCAAGTGCAGTGGTAGGTGTTGATAACTAAATATCAAAACTGGCATAATG  
TCATCCCATACTTCTCTCTAAATAGTCTTCAATTCTTTTATGTCCTACTTATGATTCAATCTCCCTCC

>Contig19 (n=612)

GGAAAGAGGCAACAGTAGAGGAAGCACCCTGCAAGTGCAGGAGGTGTTGATAACTGAAGTCCAGAACTGGCATGATA  
TCATCCCATATGTCATTGGCTTTCCAACACTTTGATCCATTTCATCAGCATCCAGCTTATTCTTTAACAGCCCTGCAACTGCCT  
TTGCTGTACAGGTAAGCCTTTCAGCTTGTAGCTATTTTTCTCCGATATTCTGCAATTTGGGGTCTTGTGAGGATGCACACC  
ATCAAATGTGCATTCTTGTATCAAGGACCAGCAATCATGATCTCTCAAACCAGCCAAATTTATCGAGTCCATTGTACCCACCA  
TCTCTGCCACTGTTGCCAACCTTGTGTATCAAGATCTTGTACCTTCTTCCAGCCTTCAAAGGAGACAAAAATCTCTCC  
ACCTCTCCTCCTCCACCAACCTACTGCTCCTCTCATCTTCCACACATCATCTAAAACGAGTAGGAACCTCTTCAACATTAGCT  
TCTCCATGACTGTCATTGAAGACTGTTGAAATTTGTAATACCCTTGTATTACGCTTGTACCAGAAACATATTCTAACATCT  
CTCTTGAAGCTTAGTCTCATC

>Contig20 (n=675)

AGTCGTGCATAAGGTCATGCATCTTGCATGTTGTTATGCAATAAAGTTCATGCCTATCTGTGCTATCATCGTCTCCTCAACTT  
CCTTGATATCCTGAAAGAAAGATCTCGAAGCCAATCATTAAAAATCTCGTCCCTTTGTCTCCAGCTCCTTTCTCCATCAG  
ATGGAATGAACCCATTAGCCATCCATAGTTGAATCAACAGATCCTTTCCATCTCATAATCCTTTGGAAATATGGCACAAAA  
GCAAAGCATTGTTTTAAATGGGAAGGCAATGACTGTAGCTCAATCTTAGCGCTGGTAAAAATCGCGTCTTACCAACCTGCAT  
ATCCCAAATCTCACTATCCCTCACGAACAACCTTCCCTTCTGGCTCTTGGAAATACAATAAGCCCCCATTGTCTTCACTGC  
TAAAGGCAGTCCACCACATTTCTTGAACAATCTCTTGGCAATATTTACGAGATTCTGAGGCTCTTCTCTCCCTCTCAAATGC  
TCTCTCCTAAACAATGTCCAGGAATCATCTTCACTCAAATATGGTAGCTGATAGGTGGCGAGTGTACCCATCATCGACGATA  
CCCGTCACTTCTTGTGTGACAATGATCCTACTCCCTACTCCACCAGTTCTTAGTAAAGTTTTTAACTCATCCCACTT

>Contig21 (n=624)

GCCAAGTCGTGCATCATGTCGTGCATTGACAGTGTTTTATTCTACCAAATTCATTCTCCTTACAACCTGAAGCATGCATCGA  
TGGATGAGCTCCTTGAGATAGTCTCTGCCACTTCTCCATTGTGCTCGATCCCTTTCTCCACAAATCCCTCTGCGACCCAC  
AGCCTCATAAGCCTCTTCTCTGAAGCACATGATCCTCCGGGAACATGCTGCAATGCAAGAAGCAGTTCTTAAGATATCTCGG  
TAGATAGTTGTAGCTGAGATTCAAAACACGCTTACATTCTCCAGATTGGGGTCTTATCCAACTGCCAGCTCAGCTGATCAT

AAACTTTCTGCCATTCTGACTTGGTTTTCTCTCTCAGGGACAGAAGGCTGCCTATAGAGACAATAGCTAGTGGCAATCCTTGA  
CATTTTAATAACAATTTCTTCCCAACTCCTCCAGTTCTTCAGGACATTTTCTCACTTTTTCTTCCGGAATGCCCTCTACAGA  
ATAGATCCCATGCCTCAGTGCTTTCCAAAACCTGGAGCTCAAGCTTCCGACTATCATGTGCTAGTGAAGCTACATCAGCACTC  
CGAGTTGTA AAAAATTATTCTGCTTCCTTTGTTGTCATCA

>Contig22 (n=426)

CGAAGACGAACACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTAATAACAAGGATATCAGGTGGCCCCGCTGAATTGG  
AATCGATATTGTTACAACACCCCAACATCTAATTAGCGGAGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACCTCCCGCCG  
CCGTTGTTGTTTTGGAGCACGGAAGACGATGACGGAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAA  
AAAGTTGCGCGGAGGAGTTGTGTTGTGGACGAAGTACCGAAAAGTCTTACCGAAAACCTCGACGCAAGAAAAATCAGAGA  
GATCCTCATAAAGGCCAAGAAGGGCGGAAAGTCCAAATTGTA AAAATGTA ACTGTATT CAGCGATGACGAAATTC TAGCTAT  
TGTAATCCTCCGAGG

>Contig23 (n=663)

CGGCCAAGTCGTGCATCAGGTCGTGCATCACGTA CTTATTATTCAGAGTTGGTAGAAAAGAGTACCTGCTGGTTAATTCCTCA  
AATATCTGACTTCCGATGTCCTCCATTCTCATTCTTACCCATTCTTGTCTGAATAAAAACCTTGAGCCATCCACATCTGGACT  
AACCGATCTTTTTCAAAAACATAGCTCTTGGGAAACATGGAGCAGTATGCGAAGCATTGCTTTAGATTGCATTCAAGTGTT  
ATAACTTAAAGCCAGAGATGGAAGGATATTATCTAAAACATCACTATTTTGCCACCCTCACTTTCTAAGATCATGTTCCAGG  
AATTCTCATTTAGTTGGACTTCAATAAACTTCCCACTACCTTTCCAGCTAGAGGTAATCCATGCAGTTTTTCAGATATTTTCT  
TCCTATGTTTGCAATTTTTCTTCTCGCTTTCCTTAAGATTGTTTACCAAATGCACAATGTTGCAAAAAGTCTCCAGAAGTCA  
TCTTCTGCAGACTCTTCAATTCTACCGGTCCATCGTGCCATGACTTCTGCCACTATTGGACTTTGAGTTGTCACCAGAACC  
ATGCTTCTTTTTGCTCCAGATGTCAAGACACCACGGAGTTCTTTCCACTTGCTGCCAGTCTCATCCACACATCAT

>Contig24 (n=619)

GCATGAGGTCGTGCATCTTGCATGTTGTTATGCAGTAACCTTCAAACCTGCCCTTGGAAACCATCGCCCCCGTTGCTTCCTTGA  
TATCTGAAAAGAGGACCTGCAAGCCAACCTCTTAAAAATCTCATGCCCTTTGTCGTCCAACCGATCCTTTCTTCCATCAGAA  
GGAATGAATCCATTGGCCATCCATAGTTGAATCAGCAGATCCTTTCCATCTCATCATCTTCGAAAATATGGCACAAAAAGC  
AAAGCATTGTTTTAAATGGGAAGGCAATGACTGTAGCTCAGCCTTAGTGCTGGCAAGATCCCATCTTACCAACCTGCATCT  
CCCAAATCTCGCTATTCTCACCAACAACCATTCCTTTTCTCGCTCTTGGAAATGCATTAAGCTCCCAAGCCTTCACTGCTA  
GAGGCAGCCCACCACATTTCTTGACAATCTGCTTGCCGATTTTTCCAGAGTTGGAGGCACTTCTGCTCCCTCTTCAAATGCC  
TCTTCTAAACAATGTCCAAGAATCATCCTCGGTCAAACCTTTTAAAAGATACGCATCGAGTGTGCCATTATCGATGATACC  
TGCTCATTTCTGTCGGTCACAATGATCCTGCTCCC

>Contig25 (n=632)

CGGCCAAGTCGTGCACCAGGTCGTGCATCTTAACCTGCATTACCACGCCGTCTGCATCTTCTGACCCTCTGGAATAGCGAT  
CTCGATAACAACCTCATCAACATACAGCAATCCACATCTTCCAAGTATGTACCTCCATTAGATGATTGCACAAAGCCTTCTGC  
AATCCATAGCTGGACCAATTGCTTGTCTTATTCTGACCTTTAGGAAAATATGGAACAATATGCAAAGCATTGTTTCTGATG  
AGATGGCAGATGATTATAGCTCAGCATCAAAGTGGGTAAGATTTTCTTCTCATCATCCGGCAAGTTCCAAATTTCACTCTCCT  
GATGGCCAACCACTGACTCTCTCTCTCCCGAAAACGCATCAAACCTTCTAAAGCCTTTGCTGCTAATGGAACACGTCACATT  
TTTTTACCATCTCCTTCCCAATGGCTGCCAAGTTTGGGGTCTTCTCGCAACTCCACAATCCAAATGCCCTCTTTCAAATAATA  
GCCAGCAATCGTAATCTGATAGACCCGATAAAAAGGTGTTGGTGCAGCTGTCCCATGATTGAAGCAACTCGTTCGCTACGCGTA  
GTCACTATGACTTTACTTCTCTGGCAATACCAATCAACACAGTTTTT

>Contig26 (n=532)

NNNNNNNNNNNNNGTTCATGCCTATCTGTGCTATCATCGTCTCTCAACTTCTTGATATCTGAAAAGAGGATCTCGAGG  
CCAATCATTAAAAATCTCGTTCCCTTTGTCCTCCAGCTCCTTTCTTCCATCAGATGGAATGAACCCATTAGCCATCCATAGTT  
GAATCAGCAGATCCTTTTCCATCTCATAATTCTTCAAAAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAA  
TGACTGTAGCTCAGCCTTAGTGCCGGCAAGATCCCATCTTACCAACCTGCATCTCCCAAATCTCGCTATTCTCACCAACAAC  
CATTCTTTTCTCGCTCTTGGAAATGCATTAAGCTCCCAAGTGTCTTACCAGCTAGAGGCAACCCACCACATTTCTTGACAATC  
TCCCTGCCAATGCTTTCCAGATGTTGTGGCGTTTTGCTCTCTTANACAATTTCCAAGAATCATGGTCAATCAAACGGTCC  
AAACGATAAGCATCGAGTGGACCCATTATC

>Contig27 (n=632)

GAGGTACTTCTGGTGGTGGATGACGTTTGGGATGAGACTGGTAGCGCATGGGAAAACTCCGTAGTGCCTAACATCTGGAGCAAAGGAAGCATGGTTCTGGTGACAACCTCAAAGTCTAAGAGTGGCAGAAGTCATGGGCACGATGGACCCGATAGAATTGAAAAAGCCTAAAGGAAGATGACCACTGGAGACTTTTCGAACATTGTGCATTTGGTGAAGCCACTCTCGAGGAAGGGCAAAGGAAAAATTGCAAACAATAGGAAGGAAAATATCTGAAAACTACATGGTTTACCCTAGCTGGAAAGGTACTGGGAAGTTTATTGAAGTCCAACTAGATGAGGATTACTGGAAGGTGATCCTAGAAAAGTGAAGTGGTGGGAACATGAATATGTTTTAGATAACA TCCTTCCATCTCTGCTTTAAGTTACGAACACTTGAACGCAAATCTAAAGCAATGCTTCGCTTACTGCTCCATATTTCCCAAGA GCTATGTTTTCGAAAAAGATCGATTAGTCCAGATGTGGATGGCTCAAGGTTTTATCCAGAACAAGAATGGGTAAAGAATGAG AATGGAGGACATCGAAAGTCAGATATTTGACGAATTAAGCAGCAGGTACTTCTTTCT

>Contig28 (n=662)

CTTTACTGCTGCGTACAAAAATCCGTAGAACCCCGCATCTCCCTCAACTTCTTGATATCCTGAAAGAAAGATCTCGAAGCC AACTCATTAAAAATCTCATGCCCTTGTCTCCAACCTCTTCTTCCACCAGATGGAATGAATCCATTGGCCATCCATAGTTGA ATCAACAAATCCTTTTCCATTTCCATAATCCTTCGGAAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATG ACTGTAGCTCAGCCTTAGTGCTGGCAAGATCTCATCTTCCCAACCTGCATCTCCCAAATCTCGCTGTCCCTCACAAACAACCA TTCCTTTTCTCGCTCTTGAATGCATTAAGCTCCCCAGTGTCTTACCCTAGAGGCAACCCACCACATTTCTTGACAATCTC CCTGCCAATGCTTTCCAGATGTTGTGGCGGTTTTGTCTCTTCTTAAACAATCCCAAGAATCATGGTCATTCAAACAGTCCAA ACGATAAGCATCGAGTGGACCCATTATCCGGGATACCCCTCACTTCGTGCAGTCACAATGATCCTACTCCCTTCTCCACCAG TTCCAAGTAAAGCTTTCAACTCATCCACTTTACCTGATCCTCGTCCAAACATCATCCAGCACCAGGAAGTACCTC

>Contig29 (n=607)

GAGTGCGGGAACCTCGAAATTCCTTGAACCTCTGGAAGAAGGATCTTGACAAAAGCTCGTTGAAAATCTCCTGGCCTTTCAAT TCCATGTCTCTCTCCGATCGGATGGAACAAATCCATTAGCCATCCACAACCTGGATCAGCATCTCCTTGTCTATTTCCATAATCC TTGGGGAATATGGCACAATAAGCAAAGCATTGTTTACATGACAAGGCAGGTGATTATAGCTCAACCTCAATGAAGGGAGGA TCTCATCATCACGTAGTTTCAAATTTCACTTTCCATCACGAACCTCCATTCCCTCTCCTCCCTCTTCGTGCGCAATAATCTCCC GAGCATCTTCGCCGCCAGGGGCAGGCTCCGCACTTGTCCACAATTGCTTTTCCAATTCGTACCAGTTTGCAGACTTCTCCGC ACCACCCTCAGAAAATGCGTTTCTTTCAAACAAGCTCCAGCATTATCATCGGACAGAGGTTTCAGTTCATAAGCATCGAGCG TGCCGATAATAGATGAAACAACAGTTGTACTGCGGGTTGTGCAATAACGCAACTTCCATCAGCTCCAGTTGTTAACAATTCT TTTAATTCTTCCCATGGA

>Contig30 (n=121)

GGAAAGAGGCAACAGTAGGGGAAACTTTTCGAGCTTCTATTAGATTGCTGAGACAGAGAGCTATTTTCGCTCCACCAGACGTCC CAACCGAGATTCAGTCGTCTTCCCCCCCCCCCCATCCC

>Contig31 (n=654)

GAGTGCGGGAACCTCGAAATTCCTTTCATGTCTTGAAGAAGGATCTTGACAACAGCTCGTAGAAAATCTCCTGGCCTTTCAAT TCCATGTCTCTCTAGATCAGATGGAATAAATCCATTAGCCTTCCACAACCTGGATCAGCATCTCCTTGTCTATTTCCATAATCC TTTGGGAATATGGCGCAATAAGCAAAGCATTGTTTACATGACAAGGCAGGTGATTATAGCTCAACCTCAACGAAGGGAGGA TCTTATCATCATTTAGCTTCAAATTTTACTTTCCATCACTGACCTCCATTCCCTCACCTCCCTCTTCGTGCGCATTAATCTCCC CAGCATCTTCGCCGCCAGGGGCAGGCTCCGCACTTCTCCACAATGGCTTTTCCAATATGTATCAGATCTGCGGACTTCTCCGC ATCACCTTAGAAAATGCGTTTCTTTCAAACAAGCTCCAGCATTATCATCGGACAGTGGTTTCAGTTCATAGGCATCGAGCG TGCCGATGATAGATGAAACAACGGTTGTACTGCGGGTTGTGCAATAACGCAACTTCCATCAGCTCCAGTTGTTAACAATTCT TTTAATTCTTCCCATGGATTCAAACATCATCCAGCACCAGGAAGTACCCTCCATCCATTAGGAATTC

>Contig32 (n=874)

CGGCCAAGTCGTGCACCAGGTCATGCATCAAATCATGTATCACATAATATGTGCGGTAGCCCTGCTTCAGTGTGTTGAAAGAAT GACCTCGACAATAAATCATCGAAGTAGTCTCTCCCTACATCTTCCATTCTGCGCTCCATCGTATCGCTGGGCCGAATGAAATCT TGTGCAATCCACAATTGACCAACTTCTTAGGCTCAAACCTTCCAATGCTCGGGAAAAGAAGCTACAATAGCGGAAGCAGCGCT GCAAGTGCAGTGGTAGGTTTTGATAACTAGATATCAAAAAGTGGCATAATGTATCCATACTTCTCTCTCCAAAATAGTCTTCC ATTCTTTTATGTCCAAGTTATGTTTCAATCTCCCTCCACAGCCTTGGCAGCTAAAGGAGAACCATTCAACATCGCAGCTATTT

CTCTGCCAATCACCTGTAATTCTGGATATCTACTGAATCTTCGCTACCAAATGCACATTTCTTGAACAGCAACCAGTACTCAT  
CTCTTGCTAATCCCTTTAGGAAGACAGGGCTCATCGAGTCCAACATTTCTGCCACCATTTCATCCGGGTGGTCACTAGAATTC  
GGCTTCCACTCTCCCAAATTGTAGAGGTGCAAGCAACTTCTCCATCTTCTTTGTTCTCCCATTTCTCTTTTCTCATCATT  
CCTCACATCATCCAACACAAGAAGAAACCTTTTGGAGAGCAGCATGCTTTAAGGATCCTCTGAAGCTTATCAAAATTCACAA  
CTTCATCATGTTTCTCATCAGTTGCAGATTGCAAGATCTCTTGTAAATCGAAATCTCATCGAACCTATCGGAGACACAAACCC  
ACATCTCACCTCAAAATATTTCTCAACTCTTCTG

>Contig33 (n=710)

CGGCCAAGTCGTGCACAAGGTCGTGCATCTTGCAAGCTGTGCGATAAAGTTACCCCGTGGATCCATAAAAATCCGTAGTACCCC  
GCATTTCCCGCAACTTCTTGATATCTGAAAGAAAGATCTCGAAGCCAACCTCATTAAAAATCTCATGCCCTTGTCTCCAA  
CTCCTTTCTCCACCAGATGGAATGAATCCATTGGCCATCCATAGTTGAATCAACAGATCCTTTTCCATTTATAATCCTTCGG  
AAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATGACTGTAGCTCAACCTTAGTGCTGGTAAGATCCCA  
TCTTACCAACCTGCATATCCCAAATCTCGCTGTCCCTCACAGACAACCATTCCTTTCTGACTCTTGGAAATGCATTAAGCTC  
CCCAATGTCTTCACTGCTAGAGGCAACCCACCACATTTCTTGACAATCTCCTTGCCAATTTTGGCAGAGTTGGAGGCACCTCT  
GCTCCCTTTTCAAATGCTCTCTTCTAAACAATGTCCAAGAATCATCCTCAGTCAAACCTTTTAAAAGATACGCATCGAGTGTG  
CCCATTATCGATGATACCTGCTCATTTCGTGCGGTCACAATGATCCTGCTCCCTTCTCCACCGGTTCTAGTAAAGTTTTAAAC  
TCATCCCACTTTGCCTGATCCTCGTTCCAACATCATCCAG

>Contig34 (n=121)

GGAAAGAGGCTACAATAGCGGAAGCTTTCGAGCTTCTATTAGATTGCTGAGACAGAGAGCTATTCGCTCCACCAGACGTCC  
CAACCGAGATTCAGTCGTCTTCCCCCCCCCCCCATCCC

>Contig35 (n=798)

GGCCAAGTCGTGCATCATGTCTGTCATCATGAAACCACCTCTCTGTGAACTTTGAAAAAATGATCTGCTCAGCAAAGCCTCGA  
AGTAATCACTCCCAATATCTCCATCTTTCTGTTCCCTGGACTTTGATAAAAACCAAGAGCCATCCAGATTTTAAATCAGCCTAG  
TTTTATCGAAGACATAGCCTTTGGGAAAGACAGCACAGTAGGGAAAGCACTGCTTAAGGGGAGCATCCAGTTGCTCATAGCT  
CAGTCTTAAACAGAAATAACGTTTACTCCATCTCTTCCCGCAGTTGTGATATTTAGTCTCCGACAAATGCTCCAGCGGCG  
TCGGTTTGATAATCCAAAAGACCTTGCAGTGTCTCCAGAGCCTTTGCTGCCAGTGGCATGCCCTGAGCCTTATTGGCTATCTC  
CTTGCCGATCTTCTCCAGGGATTGCAATCGCTGGGCGAGGCCCCAGAAAAGCAATCTTTCGGAAGATAGACCAGCACTCCT  
TGTCAGACATGGGTTCCAATATGATCTGGGTAGTGGGGACATTTCCGATAAAGCACTCCGGTTTCTGTGGTTCATGATAACT  
TTGCTTCTCGCAGTCCACAATTCAGTGGCTGGAAGAGGACTTGCCACAGGCGTGTGCTCATCCATAACATCATCCAACAC  
GATCAAGGATCTTTGTTGTTGAGCTCCTGTGCAAGTCTGCTCTGCAGTGAATCCAAGGTTTTAGGTATGGTCTAAAGCCAG  
CCACGGACTCTATCATGTCTCCAAGAAGCCTGATCTCATCCAATCTATG

>Contig36 (n=676)

CGGCCAAGTCGTGCATGAGGTCGTGCATCTTGCAAGTTCATGTTTCGAAGTAATTAGCACGTTGCCGAA  
GGTCTGAAAGAAGGATCTCCAAACCAGTTCATCAAAGATCTCTTACCCAAAACCTCCAACCTCTTCTCTCCAGATGGA  
ATGAATCCATTAGCGATCCACAGTTGGATCAGCATATCTTTTCCATCTCATAATCTTTGGGGAATATGGAGCAGAAGGCAAA  
ACATTGTTTCAAATGTGGAGGCAAGTGATTGTAGCTCAACCTTAAAATAGCTAAGATTCTTGGTTCACCAAACCTGGATATCCC  
AAATCTCATTTTCTGATACAGCCAACCACTCCTTCACTTCTTTTTTGGTGCCCATCACGGTCCCTAGGGTCTTATTGCTAACGG  
TAGTCCCCACACTTTTTTACAATTGCCTTGCCAATTTAATCAGCTCTGAACGTTCTTACGCTCCTGCTGCGAATGCCCTTTTG  
CTAAACAAAGTCCAGGAATCATCTCAGATAACAGTGGTAAATCATAAGAGGTTGATGCGCCCATGATCCGCGAGACCCGGC  
CACTGCGGGTGGTTACAAGAATTGCACTACCTTTCTCCTCCAGTGTCTAATAAAGTTTTAATGCATCCCACCTCTCTGGATC

>Contig37 (n=599)

TCGATGCTGCGGTTTAAAGTTTGTAGAGATCTTGAGGAGATCCAGGTTGCTCTCTTCCCCATGCTGGACCATCAAGAATGCT  
CTTCCATGCCATCTTGTCTCTTCGGTCCGTAAGATGTGCCTTCTATAGACACTAAAGCCAATGGCAAGCCATCACACTTCTC  
CACTATTCTTCTAGCCAAACCTCCAGTCTGGAGGACAGCTTCCATCGTTATCTCTCCAAATGCTTCTTGGCGAACAGTGT  
CCATGCATCGTATTCTTGCAATGGCTCAAGTTTCATAACCCGGATATCAACCGCCGCTGAAGCAGCTTCTGTTCCGTGTGGT  
AATGACGACTCTGCTCCACAACCAATTATCGAACAACGCACTACCAACATCATCCCAAACCTTCTTATGCCACACGTCATCCA  
AGACAATTAGGTACCTCCTAGGATGCACAAGTGAATTGAGTATCTCCACGAGTCCCCGGAAGGAATAGAATCCAGCCTGTG



CGGATGATTGGCCTGCCCTCTCTCCGGGTGCAGTGCCAATATGATTCTTTTGGAGCAGGTCATGGGCGTCGTAACCTTTGAGATA  
CAGAAACCCATGCACG

>Contig38 (n=695)

NNNNNNNNNNNNNNNNNNNGGAAAGAAGCTGCAATAGGGAAAGCAGATCCGTAGACGATAGGGGAGATCTCTGAAACTGA  
GCTTCAATATATTTTTCACACTATCGGACATTGGCATACTAGTCTCTTCCCAATCGAGGCCCTCACAGAATTCCCTCCATTGCA  
ATTCTCCTTCTCCCGATATGACAGCGCGCTTCTCGGCCACAATAGCAAGAGGTAAGCCGTGGCACCTCTTCAAGATTCT  
GTAGCCCAGTGCTTTAGATCTGAAGGACAACCTCTGTCTTATATTTCCAGAATGCTTTATTACAGAAGAGAATCCATGAGTC  
ACTCTCTTGCAAAGGTTTCAGCTCCAGAACACGGCTTTCAGCTGCCAGGCAAGCTACATCATGAATCCAAGTTGTTATAACAA  
TTCTACTTCTACATCTACTATCTACTAGCACTGAACTGAATTCATCAAAACAACGATTTATGCGCCACATATCATCTAGGACGA  
TAAGATACCTCTTTTGTGCAAATAGCTACGAAGTGTCTCCGCCAAAAGTGTCTTGAGGCATTGTGTCAATGTTGCCTGGAAGA  
ATATCACGATTTTCGTGGTAAAGGTTCCATGAGTCTTCTCATTAGATCTTCCGGTGCATATTTCTGGGAGACACGTACCCAT  
GCTCTGCGATCGAAATGTGCGCTTGA

>Contig39 (n=621)

CGGCCAAGTCGTGCATCAGGTCGTGCATCTTGCATGACCATCTATTATATGGCTGTAGAAGATTTCGCATAATCAACTCTCTGT  
AATAATCTTCGGCTATCTCTTCCATCGTCAAACCTCCCTCAACTCTCACAAATCCCTCAGCAATCCAAAGTCGAGTAATATCAC  
CCATCCAAATTTGATGGTCTTCAGGACATAATGAGCAGTACATAAAACATTGCTTCAGATAGCTTGGTAGGTCTTCGTAACCTC  
AGATATAATGCGCCCCTAACCTCCTCGGGAAGCCCAGATGTGGACCATGCGGTACTCCCAAGAACTCTATTCCAAGCACTCCG  
GGTCACCTCCTTCGTGCATAGCAACCCTGCAATCGCTTAAATGGCTAGAGGAAGGCCATCACATTCGTAACAATCTCCATCC  
CTATGTCCTTGAGATGCTGGATCTCTCCCTCCTTCCGGTCAGGACCACCTTCTTGCAAAGCAATAACCAACTATCTCTAAAG  
ACAGCTTGTGACGTGATGGGTGTTCACTGCCCTCATCTGCTTCGCGATCTGTTCAATTTCTTGATGTGACAAGGATCCTACTGC  
CAACAGCACAACTTTTTAACGTATTGCATAACAAG

>Contig40 (n=419)

TTTCTATAGTAAGTGCATACAAAAGTCGACAGCAAGGAGGTGATCGAATGAAGAACGACTGAGTATACATTGTCATGGCACA  
GGCATTGATTGAGTGGGTGTTTTGGCGGATGCATAGCTACGTTATACCAAGTAATGTCATAGAATGAAAGGTTTATGAG  
CAAATTGTGTGCACACTGGCTTTCGTGATCCGTTTCAGCTTCGTGATACACCTCTGGGAGTTGAGACTGGATAAATACACTGA  
TCAGAGACGTGCTCAGTCTACCAACACCCCCTTTCTTTCGCCCAGACTATTAATCGTGCATTCTCCTCTCATCACCTCT  
TTCCATTTTTACTGCCTCAACCCCTCAACTCACGCACTCTTCCCTGCCTGCCTGTCGTCTTCCCCCCCCCATCCC

>Contig41 (n=700)

AGTCGTGCATCAGATCGTGCATCTTGCATGTTGCTCTGCTTTCATGATCGAATTTGTAGTTGTAAGCACGTTGCCGAAGGTCCT  
GAAAGAAGGATCTCCAAACCAGTTCATCAAAGATCTTTCACCCAAAACCTTCCAACCTCTTCTTCCCTCAGATGGAATGAAT  
CCATTAGCGATCCACAGTTGGATCAGCATATCTTTTCCATCTCATAATCTTTGGGGAATATGGAGCAGAAGGCAAAACATTG  
TTTCAAATGTGGAGGCAAGTATTGTAGCTCAAACCTTCAAACAGCAAGGATTCTTTGTTACCAAACTCGATATCCCAAATCT  
CACTTTTCATTATATCCACCAATCCTTTACTTCTTTTTGGTGCGCATCAAACCTCCCTAGAGCCTTTACTGCTAACGGTAGTCC  
CCCACACTTTTTACGATCCCTCGCCAATTTCAATCAGCTCCGGACGTTCTCCAGCTCCTGGTGAATGCTCTTTGCCTAAA  
CAAAGTCCAGGAATCATCATCAGATAAAAATGATAAACTATAAGGAGGAACTGTGCGCAAGATCCGTGAGACCTGCTCACTA  
CGGGTAGTTACAAGGATTGACTGCCTTCCCTGCTAGATCCTAACAGATCTTTGAATTCGTCACCTCACTGGATCTTCATTC  
CACACATCATCCAGCACCAGGAAGTACCTC

>Contig42 (n=671)

CGTGCATCTAACCTGCCTCACCACACCATCTAAATCTTCTGACCTCTCTGGAATAGCGATCTTGATAACAACCTCATCAACAT  
ACAGCAAACCCACATCTTCCAAGTATGTTCTCCATTAGATGATTGCACAAAGCCTTCTGCAATCCATAGCTGGACCAATTGT  
TCGCTTCTTATTTCTTACCTTGAGGAAATATGGAACAATATGCAAAGCATTGTTTCAGATGAGATGGCAGATGATTATAGCT  
CAGCATCAAAGCGGGTAAGATTTCAATCTCATCATCCGGCAAGTTCAAAATTTCACTCTCCTTGATGGCCAACCACTGACTCTC  
TCCTCTCTGAAACGCATCAAACCTTCTAAAGCCTTTGCGGCTAATGGAACACCTCCACATTTTTTCAAAATCTCCTTCCCAAT  
AGCTACCAAATATGGGGTCTTCTCACAACCCCCCAATCCAAATGCCCTTTTTTCAAATAACAACCAGCAATCCTCCTCTGATA  
GATCCGATAAAAAGTGTGGCGCGACTGTTCCCATGATAGAAGCCACTCGTGCCTACGTTAGTACTATGACTTTACTTCCC

CTAGCACTACCAATCAACACAGTTCTCAGCCTATCCCATTTTTCATAATTCTCATTCCAAACGTCGTCAGCACCAGGAAGTAC  
CTC

>Contig43 (n=685)

CGGCCAAGTCGTGCATCATGTGCGTGCATCACATATTTGTTATTCTGAGTCGGTAGAAAGAAGTACCTGCTGCTTAATTCGTCA  
AATATCTGACTTCCGATGTCTCCATTCTCATTCTTAACCCATTCTTGTCTGGATAAAACCTTGAGCCATCCACATCTGGACT  
AACCGATCTTTTTTCGAAAACATAGCTCTTGGGAAATATGGAGCAGTAAGCGAAGCATTGCTTTAGATTTGCATTCAAGTGTT  
CATAACTTAAAGCCAGAGATGGAAGGATATTATCTAAAACATATTCATGTTCCCACCACTCACTTTCCAAGATCATGTTCCAG  
GACTTCTCATTTAGTTGGACTTCAATAAACTTCCCCTACCTTTCCAGCTAGAGGTAACCATGCAGTTTTTCAGATATTTTC  
CTTCCTATTGTTTGCAATTTTTCCCTTCGCTCTTCCCTCAAGATTGTTTTCACCAAATGCACAATGTTTCGAAAAGTCTCCAGAAGT  
CATCTTCTGCAGACTCTCAATTCTATCGGGTCCATCGTGCCCATGACTTCTGCCACTCTTAGACTTTGAGTTGTCACCAGAA  
CCATGCTTCCTTTTGTCTCCAGATGTTAAGGCACTACGGAGTTTTTCCCATGCGCTGCCAGTCTCATCCCAAAACATCATCCAGCA  
CCAGGAAGTACCTC

>Contig44 (n=634)

CGGCCAAGTCGTGCACCAGGTCGTGCATGAGGTCGTGCATCTGGCATACTGATCATATGGCTGCAGAAGATTTTCGCATA  
ATCAACTCTCTGTAGTACTCATTGCGGTCTCTTCCATGGTTGAACTCCCCTCAGCTTTCACGAATCCCTCAGCAATCCAAAGT  
TCAGTAATAGTACCCATCAAAATTCGATGGTCTTCAGGAAATAATGAGCAATATATAAAACATTGCTTCAGATAGCTCGGTAG  
ATCTTCATAACTCAGATATAATGCTCCCTTGACCTCCTCAGGAAGCCCAGATGTGGTCCATGCGGTGCTTTCAAGAACTCCATT  
CCAAGCTCTCCCGTCTCTCTCTTGGTGCATAGCACCCCTGCAACCGCTTTAATGGCTAGAGGAAGGCCATCACATTTCTTA  
ACAATCTCCATCCCTATGTCCCTGAGATGCTGGATCTCTCCCTCCTCCCTGTCAAGACCACCTTCTTGCAAAGCAATAACCAA  
CTATCTCTAAAGACAGTTTGTGACGTGATGGGTGTTCACTGCCATCATCTGCATCGCGATCTGTTTCAATTTCTTGTAGTGACA  
AGGATCCTACTGCCAACAGCACAACTTTTTAACGTATTGCATAACAAG

>Singleton1: Trat10\_EG\_BA1\_3F2/13R1\_NI\_5 (n=435)

CGGCCAAGTCGTGCATAAGGTCGTGCATCTTAACCTGCCTCACCAACCCACATCTAAATCTTCTGACCTCTCTGGAATAGAATA  
ATTGATCATAAATAATCACTCATCTACATACACCAAACCCACATCTTCCAAGTATGTTCCCTCCATTAGATGATTGCACAAAGC  
CTTCTGCAATCCATAGCTGGACCAATTGTTGCTTCTTATTTCTTACCTTTAGGAAATATGGAACAATATGCAAAGCATTGTT  
TCAGATGAGATGGCAGATGATTATAGCTCAGCATCAAAGCGGGTAAGATTTCACTTCTCATCATCCGGCAAGTTCCAAATTTCA  
CTCTCTTGATGGCCAACCACTGACTCTCTCTCTCTGAAACGCATCAAACCTTCTTAAAGCCTTTGCTGCTAATGGAACACCT  
CCACATTTTTTCACAATC

>Singleton2: Trat10\_EG\_BA1\_3F2/13R1\_NI\_15 (n=666)

CGGCCAAGTCGTGCATAATGTCATGCATCTTGCATGTTGTTATGCAGTAACGTTCAAACCTGCCCTTGGAGCCATCGTCCTCCG  
TTGCTTCTTGATATCCTGAAAGAAGGACCTGCAAGCCAACCTCTTAAAAATCTCATTCCCTTTGTCGTCCAACCGATCCTTTC  
TTCCATCAGAAAGGAATGAATCCATTGGCCATCCATAGTTGAATCAGCAGATCCTTTTCCATCTCATAATTCCTCGAAAAATAG  
GCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATGACTGTAGCTCAGCCTTAGTGCTGGCAAGATCCCATCTTAC  
CAACCTGCATCTCCAAATCTCGCTATTCCTCACCAACAACCATTCCTTTTCTCGCTCTTGAATGCATTAAGCTCCCCAGTG  
TCTTACCAGCTAGAGGCAACCCACCACATTTCTTGACAATCTCCCTGCCAATGCTTTCCAGATGTTGTGGCGGTTTTGCTCTCT  
TCTTAAACAATTTCCAAGAATCATGGTCATTCAAACGGTCCAAACGATAAGCATCGAGTGGACCCATTATCCGGGATACCCCC  
TCATTTCTGTCAGTCACAATGATCCTACTCCCTTCTCCACCATTCTTAGTAAAGCTTTCAACTCATCCCACTTTGCTGAT

>Singleton3: Trat11\_EG\_C10\_P1B/RNBSD\_R\_45 (n=493)

AGAGGCTGCAATAGAGGAAGCAAGAGCAAAATAGAGAAAGCAAGGCTTCGGATGATGGGGTAAGTCATCATAGCTAAGAAG  
TATTTCTTGTGTTGGTCTTCCCATATGCCCCATTACGTTTCTTCGTAGACTTCCACCTAGAACAACGATTGCGAGAGGTAAC  
CACCACATTTTCAACAATTTCTTAACGAAGAGACTCCAAATGGACCCGATATTTAGCTGACGTAATGATTTCTTCAACAAC  
AAATCCCAACTCTCTCTGGACTATGAAGTATCAACTCACAGGGAGGGCTATCAGTATCCGGGTCTGCATCCATGGCAACGTT  
CGTTTAAAGAGTAGCAAGCATCACCTCTACTTCTGTAGATCCCTGCGGAAATGCTGCTTTATATCCATTCCAGGCTTCTATGT  
CCCATATGCATCATCTATCATCAACCAAGTATCTCTTATTACTCAAATATCCATGAAAGTTATCCTTCTATGTCTTCCA

>Singleton4: Trat11\_EG\_C10\_P1B/RNBSD\_R\_7 (n=564)

TCATAACTCAACCGTAGTGTGGTAATATAACCATTGTCGTCGGCAGGCAATCTCCAGAGTTCGCTATCTCTGACATGTAACCA  
CTCTGCCTCCCCTCTTCTAGAGCTCATCAGACTTCCCAGAGCCTTAGCAGCCAGGGGCAAGCCTCCACATTTTTCAACAATTG  
TTCCCAATTTCAACCAACCTTGGAGTCTCTTCTCTCTCCTGGCCAAATGCTCTCTGCTTGAACAGAGTCCAGCAGTCAGC  
GGCTGTTAGGCCTTGCAATTTGTGCGGAGCCACCGTCCCCATGATCCTGGCAACTACATCGCTTCGTGTTGATGTTACGATCTT  
GCTTCCTTTCTTGGCCACATCTCAGTAAAGTCTTCAATCTCTCCCACTTTGCTTCATCCTCGTTCAGACATCGTCCAGAACAAGT  
AAGAAACTCCTCCCACTCAGTTTCGTCTGCAGCTGAAGCTGTGCAGCTTCCAAGTTTGGGACTTCACACTGAATAATCTTATTT  
ATAATTTCTTTGATGCTAAAATCATCAGAAAACATACCCATGATTTAAGATCAAAGTGC

>Singleton5: Trat12\_EG\_C10\_3F2/13R1\_R\_18 (n=537)

GGAACGATTTTCTTGTTAATATAACAAGAAGTCTTTGAATTTTTTGACAAGAATTAAGAACGATCTTCTTTGTAAGTCATTGA  
AATATCACTGCCAATCTCTTCCATTTCTTCTACCTCTAACTGCGATGTAGCCTTGTGCCATCCACAATCTGACTAGAACAG  
CTTTGTCAAACCTGATAACCTTGGGAAACATGGAACAATAAACGAAACAAGGTTTTAGATGTGCTGGCATGCGAGTATAGCT  
TAATCTGAGAGCTGCAAAAGTCTCATTATTTCTGTCTACTTCCCATAGATCACTTTGCAAGACATCCTTCCAATATTTTCATC  
TGTCATGTATCTTAGAAGGCTTCTTATTGTCTTACAGCCAATGGTAAACCGGAACACTTCTGACAATCTGCTTGCAGCATC  
TGTCAACTGTGATTGTTCTTTCAGAGTCTCGACCATCAAATGCGTAGTGCTTGAACAATAACCAAGATTCGTCTTCAGATAAGT  
AGCCAGGATGATAAGGATGCACCGTCTGCATGATCT

>Singleton6: Trat14\_EG\_C23\_3F2/13R1\_MR\_2 (n=614)

CGGCCAAGTCGTGCACAAGGTCGTGCATCTGGCATAACATAAGTTTCATATGGCTGCAGAAGATTTGCGATGATTAACCTCTCG  
TAGTACTCTTTTGCTGTCTCTCCATAGTCGAACTCCCCTCAGCTTTCACGAATCCCTCGGCAATCCAAAGTTGAGTAATATAA  
TGCATATTAAGTGATGGTCTTCAGGAAATAATGAGCAGTATATAAAAACATTGCTTTAGATAGCTTGGTAGGTCTTCGTAGCT  
CAAATATAATGCTCCCTTAACTCCTCAGGAAGCCAGACGTAGACCATGCAGTGCTTTCAAGAAGCTTATTCCAAGCCCTCC  
CGGTCCTCTCCTTCATGCATAGCACCCCTGCAACCGCTTAAATGGCTAGAGGAAGGCCATCACATTTCTTAACAATCTCCATCC  
CTGTGTCCTTGAGATGCTGGATCTCTCCCTCCTCTCCTGTCAAGGCCACCTTCTTGCAAAGCAATATCCAATATCCCCTAAAG  
ATAGCTTGTGACATGATGGATTTTCACTGCCGACATCTGCATCGCGATCTGTTTATTCTTGTAGTAAACAAGGACCCTACTGC  
CAGCAGCACAACCATATAATGTGTTACGT

>Singleton7: Trat14\_EG\_C23\_3F2/13R1\_MR\_28 (n=620)

CGGCCAAGTCGTGCATCAGGTCGTGCATCACGTAAGTGTATTTCGGAGTTGGTAGAAAGAAGTGCCTGCTAGCTAATTCATCA  
AATATCTGACCCCGATGTCCTCCATTCTCATTCTTACTCGATTCTTGTCTGGATAAAAACCTTGAGCCATCCACGTCCTGGACC  
AACCGATCTTTTCGAAAATATAGCTCCTGGGAAATATGGAACAATAAGCGAAGCATTGCTTGGAGATTGCGTTCAAGTGTTC  
GTAACCTAAAGCAAGAGATGGAAGGATGTTATCTAAAACATATTATGTTCCCACTCACTTTCTAGGGCCATTAAGCGG  
TTGCAGGGTACTATGCATGAAGGAGAAGACCAGGCGAGCTTGAATGGAGCTCTTGAAGCACTTCATGGTCCACATCTGA  
GCTTCCCGAGGAGTTATGGGAACCTTATATTTGAGTTACAAAGACCTACCAAGCTATTTGAAGCAATGTTTTGTATACTGCT  
CATTAGCTCCTGAGGACCATTTCAATGGCATGGATGAAAGTACTCGTCTTTGGATTGCTGAGGGATTTGTGAAAGCTGAAGGG  
AGTTTGACGATGGAAGAGACAGCAAAAGAGTACTACAGA

>Singleton8: Trat14\_EG\_C23\_3F2/13R1\_MR\_48 (n=466)

CGGCCAAGTCGTGATGATGTCGTGCATCACGTAAGTGTATTTCAGAGTTGGTAGAAAGAAGTACCTGCTAGTTAATTCGTCAA  
ATATCTGACTTCCAATGTCCTCCATTCTCATTCTTATCCGATTCTTGTCTGGATAAAAACCTTGAGCCATCCACATCTGGACTA  
ACTGATCTTGTCTGAAAACATAGCTCTTGGGAAATATGGAGCAGTAAGCGAAGCATTGCTTTAGATTTGCGTTCAAGTGTTCG  
TAACTTAAAGCCAGAGATGGAAGGATAATTTATCTAAAACATATTATGTTCCCACTCACTTTCTAAGATCATTGTTCC  
AGGACTTCTCGTTTAGTTGGGACTTCAATAAAAACCTTCCCCTACTACCCTTCCCAGCTAAAGGTAATCCATGCAGTTTTTCAAAA  
ATTTTCCCTTCTATTGTTTGAATTTTTCCCCTTCCCTCCTTCCCTTA

>Singleton9: Trat1\_EO\_MAN\_P1B/RNBSD\_CO\_7 (n=467)

TTGGAATGGAGAACTCCAAAGCAATTGGCACACCCATGAGCCCATCATGTAAGCTTGACAAGGACGAAGAAGGTAAAAGTGT  
AGACTTAAAATTTTATAGAGGCATGATTGGTTTTCTATTATATTTAACTGCTAGTAGACCTGATATTATGTTTAGTGTTTATCTT  
TGTGCTTCTTCAATCAAATCCAAAAGAATCACACTGGAATGCAGTTAAAAGAATCCTTAGATATTTAAATGGTACACAAAC  
TCTAGGATTATGGTATTCTAAAGACTCATTAAATTGACTTAATAGGATATTCAGATGCTGATTTTGTGGATATAGATTAGATAA

AAAAAGTACTAGTGAAACTTATCAATTTCTTGGAGTTAACCTAATCTCTTGGTTTAGCAAGAAACAAAATTTGATAGCACTGT  
CTACGGTCGAGGCCGAATACATTGCAGCCGAAAGTTGCTGTGCTCAAATC

>Singleton10: Trat2\_EO\_MAN\_3F2/13R1\_CO\_18 (n=672)

CGGCCAAGTCGTGCATCAGGTCGTGCATCTTACATGCCATATGTTGTTATACTTATCCTTTCTCGCATCTGAAAAAAGGATC  
TCCGCAACAAGCTATTGAAGTTTTGGTTGCCAACTTCCTTTCATGCATATCACCGTCCGATGTTTGGATGAAACCTTCGGCAA  
TCCACAGCTGAATCAGTTGCTTCCTTTCAATTTGATAATCTTTTCGAAAATATCGAACAATAAGCAAAAACACTGCTTTAAATCA  
GAGGGCAAATGATCATAACTCAATTGTAGTTCCAGTAAGATACGATTCTCGTTTCTAGGCAATCTCCAAAAGTTTCGCTATCTCT  
GACATGCACCCACTCCGCGTCTCAGTTTAAACCTCATCAGACTTCCGAGAACCTTTGCAGCCAAAAGGCAAGCCCCCTGCATT  
TCTCAACGATCTCCTTTCCAATTTGACCGACCTTGAAGTCTCTTCTTCTTCTTCTGCCCCAACGCTCTCTGCTTGAACAAAGT  
CCAGCAGTCATCGGATGTCAATCCTTGCAATTTGTGTGGTGTGACCGCTCCCATGATCCTGGCAACAACATCGCTTCGTGTTGT  
AGTTACGATCTTGCTTCCTTGCTTGCCACCTCTCAATAAGGTCTTCAGTTTCTCCACTCTGCTTCATCTCATTCCACACATCA  
T

>Singleton11: Trat2\_EO\_MAN\_3F2/13R1\_CO\_2 (n=438)

TCGTGCATGAGGTCGTGCAAATTATCATAGCTGAGCTTTAATATGCGAAACATGTCTCTTTCTTCATCTCCCTCCGGTAGGATC  
TCCATTTCTGGATCTCAGCATGCTTGGAGTCCTTCAGCAACTTGATTACGTGCCTCCAGTGCTTGGGTTCCTCCGGTTGCGC  
ATGGCCCGTCCGATCGTCTTGAGAGCGAGCGGCAAGCCTCTGCACTCCTTCGCAACCTCTCCTGCAAGCTCCGGTATGTCAGG  
GTGAGAGCTGAGAGTCTGCTTACCCACATAAAAGTTGAACAGATCCAGCGCTCTCTGCCCCAAGCAGTTCATCTTTATCT  
CGGCATCGGCGCTCATTTTGCCGCAAACGCTCTGCGATCGGGTGGTGAGTATAACTTTGCGCTTATGCTGATTCCCTCCGGTGC  
AGGGACGTGGGATCCCCACA

>Singleton12: Trat2\_EO\_MAN\_3F2/13R1\_CO\_20 (n=604)

CGTGCATCAGGTCGTGCATGCTGCATGTCTGTTTCGTGATCATCATGGGGTAACTTTGAGTGTCTTGAAGTCTTGAAGAAG  
GATCTTGACAATAGCTCGTTGAAAATCTCATGGCCTTTCAATTCCATATCCTGCGGATCAGATGGGATAAATCCATTAGCCAT  
CCACAACGGATCAGCATCTCCTTGTCATTATATAATCCTTGGGGAATGTGGCGCAATAAGCAAAGCATTGTTTCACATGAC  
AAGGCAGATGATCATAGCTCAACTTCAATGAAGGGAGGATCTTATTATCATTTAGTTCCAAATTTCACTTCCATCACGAAC  
CTCCATTCCTCTCCTCCCTACTCTTCGTGCGCAATAATCTCCCTAGCATCTTCGCCGCCAGGGGCAGGCTCCGCACCTGTCC  
ACAATTGCTTTTCCAATATGTACCAGATTTGGGGACTTCTCCGCACCACCCTCAGAAAATGCGTTTCTTTCAAACAAGCTCCAG  
CATTATCATCGGACAGAGTTTCAGTTCATAAGCATCGAGTGTGCCGATAATAGATGAAAACAACAGTTGTACTGCGGGTGT  
CACAATAACGCAACTTCCATC

>Singleton13: Trat4\_EO\_TEF\_3F2/13R1\_CO\_21 (n=585)

GAGGTACTTCTGGTCTGGACGATGTTTGGGATGAGACTGGTAGCAAGTGGGAAAACTCCGTAGTGCCTTGACATCTGGA  
GCAAAGGGAAGCATGGTTCTGGTGACAACCTCAAAGTCCGGTTGTAGCAAAAAGTAATGGGCACAATGGACCCGATAGAGTTGG  
AGTGTCTACAGGAAGATGACTACTGGAGACTTTTTGAACATTGTGCATTTGGTGACCAAATTTCTGAGGAAGAGGGAAGGGA  
AAAATTACAAGCAATAGGAAGGAAAATATCTGAAAAACTGCACGTTTACCTCTAGCTGGAAGGGTACTGGGAAGTTTATTG  
AAGACCAAATTAATGAGGATCACTGGAAGATGATCTTAGAAAAGTGAGTGGTGGGAACATGAATATGTTTTAGATAATATCC  
TTCCACCTCTGGCTTTAAGTTAGGAACACTTGAATGCAAATCTAAAGCAATGCTTCGCTTACTGCTCCATATTTCCCAAGAGCT  
ATGGTTTCGAAAAAGATCGGTTAGTCCAGATGTGGATGGCTCAAGGTTTTATCCAGAACAAGAATGGGTTAAGAATGAGAAT  
GGAGGACAT

>Singleton14: Trat5\_EG\_NIG\_P1B/RNBSD\_CO\_6 (n=613)

GGAAAGAGGCTACAATAGAGGAAGCACTGCTTCAGATGTCCAGGTAGTTGCTGATAGCTCAATCTTAGAGCTGACAGTATAT  
TTTTCGCATTCTGCGGCAGTTCCCATATCTCGCTCTCCAAGATGATTCTCCAGCTGTCTTCATCTAATTTGGAATACAAGAGGC  
CTCCCAGTATCTTTGCGGCTAGAGGTAAGCCTCTAATCTTGCTAACTATTTCTTGGCAATTTCTCTAATTTTGGATGCTCACT  
GGAGTTCTATTTGCGAATGCGATTCTTTAAATATCAGCCAGCAATCTTGGTTTGATAAGCCCTCCAAACAATACGATGAAG  
CCATGTCCATGATATTCGCCACATTTTGTGTTGCGTTGTTATTAAGATCTTGCTGCCCCTTTCCCATAACATCAAGGGGAGGA  
ATATTGGCTCCCATGGTTGCGATCTCATTCCAGACATCATCCAGTACCAGCAAATACCGATTTCTGGTAGTTGCTCTTTGA  
GATTGTCTGAACCATATTTCAATTTACCAGTTCAGATTGCTCCCTTAGAAGCAGATTCCATGATCTCTTAGTGATCCTTC  
TAACACTAAAATCATCTGACACATGAA

>Singleton15: Trat8\_EG\_LM2T\_3F2/13R1\_R\_20 (n=652)

XXXXXXXXXXXXXXXXXXXXCGGCCAAGTCGTGACAAGGTCGTGCATCTTGCAAGTTGTTATGTAATAAACTTTACTGCTGCGT  
ACAAAAAATCCGTCGAACCCGGCATGTCCCTCAACTTCCTTGATGTCCTGAAAGAAAGATCTCGACGCCAACTCATTAAAAAT  
CTCATGCCCTTGTCTCCAACCTCTTTCTCCACCAGATGGAATGAATCCATTGGCCATCCATAGTTGAATCAACAAATCCTT  
TTCCATCTCATAATCCTTTGGAAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATGACTGTAGCTCAATC  
GTAGCGCTGGTAAAAATCGCGTCTTACCAACCTGCATATCCCAAATCTCACTATCCCTCACGAACAACCATTCCCTTTCCTGGC  
TCTTGGAATACAATAAGCCCCCATTGTCTTCACTGCTAAAGGCAGTCCACGACATTTCTTGACAATCTCCTTGCCAATATTTA  
CGAGATTCTGAGGCTCTTCTCCTCCCTCTTNCAATGCTCTCTGCCTAAACAATGTCCAGGAATCATCTTCAGTCAAATATGGTA  
GCTGATAGGTGGCGAGTGTACCCATCATCGACGATACCCGCTCACTTTCTGTTGTACAATGATCCTA

>Singleton16: Trat8\_EG\_LM2T\_3F2/13R1\_R\_34 (n=653)

TCGTGCATCATGTCGTGCATCTTGCATGTTGTTGTATAATAAAGTTCATGCCTGAATGTGCGATCACCAAGATAAAAAATGTTG  
CCCACGGTACGTCTCGGCTTCTCAACTAACTTGATATCCTGAAAGAAAGATCTCGAAGCCAACCTCATTAAAAATCTCATGCC  
CTTTTCTCCTCCAACCTCTTTCTTCCATCAGAAGGAATGAATCCATTGGCCATCCATAGTTGAATCAGCAGATCCTTTCCATCT  
CATCATCCTTCGGAAATATGGCACAAAAAGCAAAGCATTGTTTTAAATGGGAAGGCAAATGACTGTAGCTCAGCCTTAGTGC  
TGGAAGATCCCATCTTACCAACCTGCATATCCCAAATCTCGCTGTCCCTCACAGACAACCATTCTTTTCTGACTCTTGGA  
ATGCATTAAGCTCCCCAATGCCTTCACTGCTAGAGGCAACCCACCACATTTCTTGACAATCTCCTTGCCAATTTTTTCTAGATT  
TGGAGGCACATCTGCTCCCTTTTCAAATGCTCTCTTCTAAACAATGTCCAGGAATCATCCTCAGTCAAACATGATAGGGCAG  
AAGTGGGAAGTGTACCCATTATCGACGATACTTCTCACTTCGTGCGGTCACAATGATCCTGCTCCCTTC

## 9.2 Apêndice B: Lista de sequências protéicas de *contigs* e *singletons*

### >Contig 1

RYFLVDDVWVTANVWCDLLRNTLYGCAAGSRVLVTRNEQIAMQMMAVNTHHVDKLSLGDGWSLLCKKVVLTGEEGEIQHLK  
DIGMEIVKKCDGLPLAIKVVAGVLCACEKTRRAWNRVLESTAWSTSGLPPEVMGALYLSYEDLPSYLKQCFVYCSLFPEDHQIEM  
DYITKLWIAEGFVKAEGSSTMEETAKEYYRELTMRNLLQANFRVHDMHDLAX

### >Contig2

XXXXXXXXRYFLVDDVWDTKWEALCDYLTYAEKSGKILITAREEKSVDVMGRTVRKIFLEGLAYRDFWLFFEACAFGNERSVSLFR  
WPELKLVGMEIAKCLKGNPLAAKTGKLLSANIDHSYWTSILTSDLWKLEQGPNDIMPALRLSYRRLPAHLQRCFAYCSIFPKGYK  
FIKERLVHMWMAHGFIQTWDEHPKLEDIGESYFNDLLQRSFFIRGEHSNDELYMHDLMHDLAESLVX

### >Contig3

RYFLVDDVWVWLSQKAAGIYREAAAASASEDLTDQI\*I\*ILGAIFMRRRRFRSCLSGYR\*RLGNCVAILEPRMLLRSSSTG\*SSYRSGIS  
LIFKF\*I\*F\*IKY\*IIRIRFNRP\*IKIK\*DNFFLNISPQILLDLEDPT

### >Contig4

RYFLVDDVWNEEEAEWERLKPLLRRGGKKGSKIIVTRSERVASIMGSSALLRPLVPTDDCWTLFRQRAFGLGRAEETPSLVKIGK  
KIVQKCGGLPLAAKALGSLMSSKRGEVWLAVKNSEIWKLPACKETGILPALRLSYDHLPSHLKQCFAYCSLFPKDYSIERERLIQLW  
IAEGFINHPSDNNMDLENIGNQFFNLLWRSFFQDAEKDSDGNVTVCMMHDLMHDLAX

### >Contig5

DV\*NADLWEDLLRNSFERATASGRIVVTRHTNVVGSMRGDGNTADIIHHVKKMDNESGWEMLRKRIFGDDVVEEVIDLKEVG  
VKIIEKCDGLPLAIKTIAGVLSSEKTNMEWSKVLKGDASWMLQDKELPGALFLSYX

### >Contig6

IIVTRSERVSSMMGLTATYQLPYLTEDDSWTLFRQRAFEEGGEEPQNLVNIKEIVKKCRGLPLAVKTMGGLLYSKSQEREWLHV  
RDSEIWDQMVGEDAILPALRLSYSHLPSHLKQCFAFCAIFPKDYEMEKDILLIQLWMANGFIPSGGRKELEDKGHEIFNELASRSFFQ  
DIKEVEGHAGFDGFFVRSSKVVYITTCMMHDLVTTWP

### >Contig7

DKKFGGVVKKVVVSQNPNIQIRRDIAEELGMTLWGSGESAAALAAARLGKEKIAIMLDDIWARLELDIGIPFRDDHKGCKILLTS  
RKSEVCDLMDTVDVSVLVDVLSSEDSWDLFKSKAGKVVETSDVEPVAKKVAKCEGGLPLAIVVGRALRGFKERSVWDDALLQLQ  
QSTPANLEGVEEQLYKSLELSYKQLKTDEMKVLFSLLLPLP

### >Contig8

GMGGGGKTTLVNDIYRSQEVKRYFDRRAWVRVSQKYTTEDLMRRLIEDLYNENRDILPDNIDTMPCDTLAEVLHAYLQKRYLIV  
LDDMWHINGWFDELSPVLVDSKCRSRIVITTRNHDVAAYLAVESRVLELQPLQEGDAWILFRNKAFAWKYKDRSCPSDLEYWARKIL  
QRCHGLPLAIVAAGSSLSYRERELEWRKFCEGLDWEETNMPMSERVRNILKLSFGDLSYHLRNCFLYCLFP

### >Contig9

RYFLVDDVWVWDETGSKWKELRGVLTSGAKGSMVLVTTQSPVVAEVMGTMEPIELKSLEEDDHWLRFEHCAFGEATLEEERRKLL  
QTIGRKISEKLHGLPLAGKVLGSLKSKLDEDYWKMISESEWEHEYVLDNIPSLALSIEHLNANLKQCFAYCSIFPRSYFFEKDR  
LVQMWMAQGFIQNKNRVRMRMEDIGGQIFDELASRYFFLPTLNNKYVMHDMHDLAX

### >Contig10

DDVWNAENVEDWENRDRWRKLLAPLRSGARGSKILATTRREMVVMTLGAVNISYKGLRGDDYWLLFQKHAFGCENPDNHLEL  
LRIGTQIAERLNGSPLAAKVVGGMLNAELDVGKWKVNLKSDIWDGIMPVLRSLYQNLPAHLQRCFAYCSIFPKDWKFEPDRLVYM  
WMAQGFIQAQSWNSVRIEDIGRVYFNDLLSRFFQTLEQDGHYYMMHDMHDLAXXXXXXX

>Contig11

DVWTAKVWCDLLCNTLKSCAVGSRILVTTRNEQIAMQMRAVNTTHINKLSSKDSWLLCKKVVLTGEEEEIQHLEDIGMEIVKCC  
DGLPLAIKAIAGVLCTKEMTRSAWNRVLGSTAWSTSGLPEDVQGALYLSYEDLPGYLKQCFLYCSLFPEDYTFMHEITELWIAEG  
FVKADQGSMTMEETAEEYYRELIMRNLLQRYFRMHDLVHDLAX

>Contig12

RYFLVLDDVWDTKWEALCDYLTYAEKGSKILITAREEKSVDVMGRTVRKIFLEGLAYRDFWLFFACDDVWDPEAFNDRQAFVN  
NKGSRIIITRSTEVASLAHDSRKLKGLQSTQAWDLFCRAFWKEKVRCPKEEELGKIVSKCQGLPLAIVSLGSLSLREKTK  
SEWGKVYDRLSWELNHNPNLNDNVKHVNLNSYNYLPRYLKNCFLHCSMFPEHLLQRKRLMRLWVAEG

>Contig13

RYFLVLDDVWKCDFRTGYFICRASNSLKV\*WVTLIFFTIKVTICYCM\*FNKSHAPSSFSCLCSLVPYFL\*QALWLSHLRAFGTRRQA  
SSYMFYSIACLD\*LALACFMEVSGLMVD\*YFIITWNIRLVGYKLLFWMGYKTNKSYRAWN\*NFFFNYNFHFSTRICL\*VI\*LSKPII\*HS  
LDWHVVSTKLKIGSLRS\*S\*CTTLCCTWPXXXXXX

>Contig14

ESTEAWNLFCKKAFRKEEDRKCPEELKKGQDIVSKCQGLPLAIVSLGSLSLRERTKSEWQKVYDRLSWELNNPNLNDNVKHV  
NLNSYNYLPRYLKNCFLHCSMFPEHLLQRKRLMRLWVAEGIMEERGSSTMEEVAEDYLKELIHRMQLQVVRNXX

>Contig15

RYFLVLDDVWDETGSKWKELRGVLTSAGAKGSMVLVTTQSPIVAEMGTMDPVELKSLQEDDFWRLFHCVFGENNLKEERREKL  
QTIGRKISEKHLHGLPLAGKVVGSLLKSQLNEKSWNMILESEWWEHEYVLDNLPSLALSIEHLNANLKQCFAYCSIFPKSYVFEQDQ  
LVQMWMAQGFQKKNRIRMRMEDIGSQIFDELTSRYFFLPTLNNRYVMHDMVHDLAXXXXXXX

>Contig16

IWRAVGVPIPSRRPTGQHKHKVVFTRTEQVCGSMGAHEKIKINCLEPEDSWKLFQDMVGQDTLTRIPTSARQVVEECRGLPLAL  
TIIGKAMSTRKTRQEWQNAITLLQSSKIPEIPDKDEDEDMLPRLKLSYDSLPPDAVRKCFLLCALWPEDFSIKKIDLIECWMGHGLID  
VGGFKDINEAYDSGHVIGKLSACLLEPGLNEDDVVKMHDLMH

>Contig17

AFSEGAEKSANLVHIGKAIADKCGGLPLAAKMLGRLMRTKMEVREWRVMESEIWKLNDDKILSSLRLSYNHLPHVQCFAY  
CAIFPKDYEIDKEMLIQLWMANGFIPSDPEQDMELKRP\*DFPRAIVKVLPRLDNSQIRHX

>Contig18

REIES\*VGHKRIEDYLGERSMG\*HYASFDI\*LSTPTSALAALFPLLLPLS

>Contig19

DETKLTREMLEYVSGDKREYKGITNFNSLQVTVMEKMLKRFLVLDDVWVWDERSSRLVEERWERFLSPLKAGKKGSKILMTTR  
LATVAEMVGTMDSINLAGLRDHCWSLIKECTFDGVHPDKHPKLNIGRKIANLKLGLPVAAKAVAGLLKK\*AGC\*\*NGSKCWK  
ANDIWNIMPVLDVSYQHLPAHLQRCLFYCCLFP

>Contig20

KWDELKTLRLTGGVGSRIIVTTRSERVSSMMGTALATYQLPYLTEDDSWTLFRKRAFEEGGEEPQNLVNIKEIVKCKGGLPLAVKT  
MGGLLYSKSQEREWLFVRDSEIWDVQVGEDAILPALRLSYSHLPSHLKQCFAFCAIFPKDYEMEKDLIQLWMANGFIPSDGRKEL  
EDKGNEIFNELASRSFFQDIKEVEEDDDSTRHELYCITTCCKMHDLMHDX

>Contig21

DDNKGSRIIITRSTEVASLAHDSRKLKGLQVLESTEAWDLFCRAFRKEKVRKCPKEEELGKIVLKCQGLPLAIVSIGSLSLREK  
TKSEWQKVYDQLSWQLDKNPNLENVKNVNLNSYNYLPRYLKNCFLHCSMFPEHLLQRKRLMRLWVAEGFVEERGSSTMEEVA  
EDYLKELIHRMQLQVVRNNEFGRIKHCRMHDMHDLA

>Contig22

SEDYNS\*EFRHR\*QLHFTIWTFRPSWPL\*GSL\*FLLRRVFR\*DLSVLRPQTQLLRATFSRLLLDWRRNPRSLFPSSSFRAPKQRRREV  
HRRHRREDLPLR\*LDVGV\*QYRFQFSGGHLISFVFN\*RLQAVNYEEVFVFX

>Contig23

DVWDETGSWKELRGVLTSGAKGSMVLVTTQSPIVAEVMGTMDPVELKSLQEDDFWRLFHCAFGENNLKEERREKLQTIGRKIS  
EKLHGLPLAGKVVGSLLKSKLNENSWNMILESEWWQNSDVLNLPALSLSYEHLNANLKQCFAYCSMFPSYVFEKDRLVQMW  
MAQGFQNKNGVRRMRMEDIGSQIFEELTSRYFFLPTLNKYMHDLMHDLAX

>Contig24

GSRIIVTARNEQVSSIMGTLDAYLLKGLTEDDSWTLFRKRAFEEGAEVPTLEKIGKQIVKCCGGLPLAVKALGSLMHSKSEEKWL  
LVRNSEIWEWQVGEDGILPALRLSYSHLPSHLKQCFAFCAIFPKDDEMEKDLLIQLWMANGFIPSDGRKDRLDDKGHEIFKELACRS  
FFQDIKEATGGDGSKGRFEGYCITTCMHDLMX

>Contig25

KTVLIGIARRSKVIVTTRSERVASIMGTVAHLLSGLSDYDCWLLFEKRAFGLWSCEKTPNLAAIGKEMVKKCGRVPLAAKALGSL  
MRFRGESQWLAIKENEIWNLPDDENEILPTLMLSYNHLPASHQKQCFAYCSIFPKGAEIRSKQLVQLWIAEGFVQSSNGGTYLEDVG  
LLYVDELLSRSFLQSGQEDADGVVMQVKMHDLVHDLAX

>Contig26

IMGPLDAYRLDRLNDHDSWELXKKRAKPPQHLESIGREIVKCCGGLPLAVKTLGSLMHSKSEEKELLVRNSEIWEWQVGEDGIL  
PALRLSYSHLPSHLKQCFAFCAIFSKNYEMEKDLLIQLWMANGFIPSDGRKELEDKGNEIFNELASRSFFQDIKEVEEDDDSTRHEX  
XXXX

>Contig27

EVLPGAG\*RLG\*DW\*RMGKTP\*CLNIWSKRKHGSGDNSKSKSGRSHGHDPDRIEKSLKEDDHWRLFHCAFGEATLEEGQRKLL  
QTIGRKISEKLHGLPLAGKVLGSLKSKLDEDYWKVILESEWWEHEYVLDNLPSLALSIEHLNANLKQCFAYCSIFPKSYVFEKDR  
LVQMWMAQGFQNKNGLRMRMEDIGSQIFDELSSRYFFL

>Contig28

RYFLVLDVWVWNEQVQVWDELKALLGTGEGSRIIVTARSEGVSRIMGPLDAYRLDCLNDHDSWELFKKRAKPPQHLESIGREIVK  
KCGGLPLAVKTLGSLMHSKSEEKELFVRDSEIWEWQVGEDEILPALRLSYSHLPSHLKQCFAFCAIFPKDYEMEKDLLIQLWMAN  
GFIPSGGRKELEDKGHEIFNELASRSFFQDIKEVEGDAGFYGFFVRSKX

>Contig29

HGKN\*KELLTTGADGSCVIATTRSTTVVSSIIGTLDAYELKPLSDDECWLSLAFERNAFSEGGAEKSAKLVRIKAIVDKCGGLPLAAK  
MLGRLLRTRKREEREWRVMESEIWKLRDDEILPSLRLSYNHLPCHVKQCFAYCAIFPKDYIDKEMLIQLWMANGFVPSDREEDME  
LKGQEIFNELLSRSFFQEFKEFRSSRTX

>Contig30

GMGGGGKTTESRLGRLVERNSSLQSNRSSKVSPTVASFX

>Contig31

EFLMDGGYFLVLDVWVWNEQVQVWDELKALLGTGEGSRIIVTARSEGVSRIMGPLDAYRLDCLNDHDSWELFKKRAKPPQHLESIGREIVK  
IGKAIVEKCGGLPLAAKMLGRLMRTKREVREWRVMESEIWKLRDDEILPSLRLSYNHLPCHVKQCFAYCAIFPKDYIDKEMLIQ  
LWKANGFIPSDLEEDMELKQEIFYELLSRSFFQDMKEFRSSRTX

>Contig32

RRVEEYFEVRMWCVSDRFDEISITREILQSATDEKHDEVVNFDKLQRIKSMMLSKRFLVLDVVRNDEEKRWENKERWRKLL  
APLQFGESGRILVTRMEMVAEMLDSMPVFLKGLARDEYLLFKKCAFGESEFSRYPELQVIGREIAAMLNGSPLAAKAVGGRL  
KHNLDIKEWKILEREVWDDIMPVLISSYQNLVHLLQRCFRYCSFFPEHWKFEPKLVNLWIAQDFIRPSDTMERRMEDVGRDYFD



DLLSRFFQTLKQGYRTYYVIHDLMHDLVHDLAX

>Contig33

LDDVWNEDQAKWDELKTLGTTGGEGSRIVTARNEQVSSIMGTLDAYLLKGLTEDDSWTLFRKRAFEGAEVPPTLAKIGKEIVKK  
CGGLPLAVKTLGSLMHSKSQEKEWLSVRDSEIWDQVGEDGILPALRLSYSHLPSHLKQCFAFCAIFPKDYEMEKDMLIQLWMAN  
GFIPSGGRKELEDKGHEIFNELASRSFFQDIKEVAGNAGYYGFYGSTGELYRTACKMHDLVHDLAX

>Contig34

GMGGGGKTTESRLGRLVERNSSLQSNRSSKASAIVASFX

>Contig35

HLRDEIRLLGDMIESVAGFRPYLKTLDLSQRSQAQELNNKRSLIVLDDVWDEQQRLWQVLFQPLNCGLRGSKVIMTTRNRSALSEN  
VPTTQIILEPMSDKECWSIFRKIAFAGASPSDCNPLEKIGKEIANKAQGMPLAAKALGDTARSFGLSNRRRWSNLSETEISQLREENG  
VNVISVLRLSYEQLDAPLKQCFPYCAVFPKGYVFDKTRLIKIWMALGFIKVGQTERMEDIGSDYFEALLSRFFQSSQRRGGFMMHD  
MMHDLA

>Contig36

IQRGGMH\*KLY\*TLEEKGSAILVTTRSGRVSRLMGASTSYDLPLLEDSDSWTLFSKRAFAAGAEERSELKIGKAIIVKKCGGLPLAIKT  
LGTVMGTKEVKEWLAVSENEIWDIQFGEPRILAILRLSYNHLPPHLKQCFAFCSIFPKDYEMEKDMLIQLWIANGFIPSGGRKELEV  
LGEEIFDELVWRSFFQDLRQRANYFEHGNQSRVTCKMHDLMHDLAX

>Contig37

RAWVSVSQSYDAHDLLKRIILALHPERGQANHPHRLDSIPSRGLVEILKSLVHPRRYLIVLDDVWHKEVWDDVGSALFDNGCGSRV  
VITTRNQEAAASAVDIRVMKLEPLQEYDAWTLFRKEAFGRDNDGSCPELEGLARRIVEKCDGLPLALVSIGRHILRTEETKMAWK  
SILDGPAWGKRANLDDLKISTNLNRSIX

>Contig38

KRHFDRRAVWRVSQKYAAEDLMRRLMEDLYHENRDILPGNIDTMPQDTLAETLRSYLQKRYLIVLDDMWRINRCDEFSSVLV  
DSRCSRIVITTWIHDVACLAASRVLLEKPLQESDSWILFCNKAFWKYKDRSCPSDLKHWATEILKRCHGLPLAIVAAGSALSURE  
KEELQWREFCEGLDWEETSMPMSDSVKNILKLSFRDLPYRLRICFPYCSFFPXXXXXX

>Contig39

LLCNTLKSACAVGSRLVTTRNEQIAQMRAVNTHHVNKLSEDSWLLCKKVVLTGEEGEIQHLKDIGMEIVTKCDGLPLAIAIAG  
LLCTKEVTRSAWNRVLGSAWSTSGLPEEVGALYLSYEDLPSYLKQCFMYCSLCPEDHQIWMGDITRLWIAEGFVRVEGSLTME  
EIAEDYYRELIMRNLLQPYNRWSCKMHDLMHDLAX

>Contig40

DGGGGKTTTGKAGKSA\*VEGVEAVKMGKR\*\*ERRMHDLIVAGERKWGVGRLSTSLISVFIQSQLPEVYHEAERITKASVTHNLLIN  
LSFYDITWV\*RELCIRQNNPLNQMPVP\*QCILSRSSFDHLLAVDFCMHLL\*K

>Contig41

RYFLVLLDDVWNEDPVRWDEFKDLLGSSREGSTILVTTRSEQVSRILRTVPPYSLSFLSDDDSWTLFRQRAFSPGAGERPELIEIGEGIV  
KKCGGLPLAVKALGSLMRKKEVKDWVDIMKSEIWDIEFGEQRILAVLKLKSYNHLPPHLKQCFAFCSIFPKDYEMEKDMLIQLWIA  
NGFIPSEGRKELEVLGEEIFDELVWRSFFQDLRQRAYNYKFDHESRATCKMHDLMHDX

>Contig42

RYFLVLLDDVWNENYEKWDRLRTVLIGSARGSKVIVTTRSRVASIMGTVAPHFLSDLSEEDCWLLFEKRAFGLGGCEKTPYLVAIG  
KEIVKCKGGVPLAAKALGSLMRFRGESQWLAIKESEIWNLPDDENEILPALMLSYNHLPSHLKQCFAYCSIFPQGEIRSEQLVQL  
WIAEGFVQSSNGGTYLEVDVGLLYVDELLSRSLFQRGQEDLDGVVRQVKMHX

>Contig43

RYFLVLDDVWDETGSAWEKLRSAITSGAKGSMVLVTTQSLRVAEVMGTMDPIELKSLQEDDFWRLFHCAFGENNLEERREKL  
QTIGRKISEKLHGLPLAGKVVGSLKSKLNEKSWNMILESEWWEHEYVLDNLPSLALSIEHLNANLKQCFAYCSIFPKSYVFEKRS  
VSPDVGSRFYPEQEWVKNENGGHRKSDI\*RIKQQVLLSTDSE\*QICDARHDARLGR

>Contig44

LLCNTLKSCAVGSRLVTTTRNEQIAMQMMAVNTHHVNKLSLEDSWLLLCKKVVLGTGEEGIEIHLRDIGMEIVKKCDGLPLAIKAV  
AGVLCTKERGPGLGMEFLKAPHGPHLGLRRSREHYI\*VMKIYRAI\*SNVLYIAHYFLKTIEF\*WVLLNFGLLRDS\*KLRGVQPW  
KRPQMSTTES\*LCEIFCSHMISMYARCTTSCTTWCTTWP

>Singleton1: Trat10\_EG\_BA1\_3F2/13R1\_NI\_5

IVKKCGGVPLAAKALGSLMRFRREGESQWLAIKESEIWNLPDDENEILPALMLSYNHLPShLKQCFAYCSIFPKGEEIRSEQLVQLWI  
AEGFVQSSNGGTYLEDVGLVYVDE\*LFMINYSIPERSGRFRWCGEAG\*DARPYARLGR

>Singleton2: Trat10\_EG\_BA1\_3F2/13R1\_NI\_15

QAKWDELKALLGNGGEGSRIIVTARNEGVSRIIMGPLDAYRLDRLNDHDSWELFKKRAKPPQHLESIGREIVKKCGGLPLAVKTLGS  
LMHKSSEEKEWLLVRNSEIWEMQVGEDGILPALRLSYSHLPShLKQCFAFCAIFSRNYEMEKDLIQLWMANGFIPSDGRKDRLDD  
KGNEIFKELACRSFFQDIKEATEDDGSKGRFERYCITTCCKMHDIMHDLAX

>Singleton3: Trat11\_EG\_C10\_P1B/RNBSD\_R\_45

WKT\*RITFMDI\*VIRDTW\*\*\*MMTYGT\*KPGMDIKQHFRDLQEVVMLATLKANVAMDADPDTDSPPCELILHSPEESWDL\*RK  
SFTSAKYRVHLESLR\*EIVEKCGGLPLAIVVLGGLRRNVGMHGRPNKNTS\*L\*\*LTPSSEALLSLFCSCFLYCSLX

>Singleton4: Trat11\_EG\_C10\_P1B/RNBSD\_R\_7

HFDLKSWVCVSDDFSIKEIINKIICEVNPLEAAQLQLQTKLSGRSFLVLDVWNEDEAKWERLKTLLRCGKKGSKIVTTTRSDVV  
ARIMGTVAPHKLQGLTAADCWTLFKQRAFPGREEETPRLVEIGKQIVEKCGGLPLAAKALGSLMSSRRGEAWLHVRSSELWRL  
PADDNGILPALRLSYX

>Singleton5: Trat12\_EG\_C10\_3F2/13R1\_R\_18

IMQTVHPYHPGYLSEDESWLLFKHYAFDGRDSEEQSQLTDAGKQIVRKCSGLPLAVKTIGSLRLRYMTDENSWKDVLQSDLWEVDK  
NNETFALRLSYTRMPAHLKPCFVYCSMFPGYQFDKAVLVRWMAQGYIAVRGRRRMEEIGSEYFNDLQRRSFFNSCQKIQRSS  
CILTRKSFY

>Singleton6: Trat14\_EG\_C23\_3F2/13R1\_MR\_2

RNTLYGCAAGSRVLVTTTRNEQIAMQMSAVKIHVNKLSLGDWILLCKKVALTGEEGIEIHLKDTGMEIVKKCDGLPLAIKAVAG  
VLCMKERTGRAWRVLESTAWSTSGLPPEVKGALYLSYEDLPSYLYKQCFIYCSLFPEDHTLNMHYITQLWIAEGFVKAEGSSTME  
TAKEYYRELIMRNLLQPYETYVCQMHDLVHDLAX

>Singleton7: Trat14\_EG\_C23\_3F2/13R1\_MR\_28

SVVLFCLFHRQTPFSFHKSLSNPKTSTFIHANEMVLR\*\*AVYKTLQIAW\*VFVTQI\*GSHNLLGKLRCGP\*SAFKSSIPSSPGLLLH  
A\*YPCNRFNGPRK\*VVG\*ICFR\*HPSISCFKLRTLERKSQAMLRLLFHISQELYFRKRSVGPVVDGSRFYPEQESSKNENGGHRGSDI  
\*\*IS\*QALLSTNSE\*QVRDARPDARLGR

>Singleton8: Trat14\_EG\_C23\_3F2/13R1\_MR\_48

REGGKGNCKQ\*EGKIFEKLHGLPLAGK\*WGSFIEVPTKREVLEQ\*S\*KVSGWEHEYVLDKLSFHLWL\*VTNT\*TQI\*SNASLTAP  
YFPRAMFSSKIS\*SRCGWLVLSRTRIG\*E\*EWRTLEVRYLTN\*LAGTSFYQL\*ITGT\*CTTSSRLGR

>Singleton9: Trat1\_EO\_MAN\_P1B/RNBSD\_CO\_7

GMENSKAIGTPMSPCKLDKDEEGKSVDLKFYRGMIGFLLYLTA SRPDMFVSVYLCACFQSNPKESHWNVAVKRILRYLNGTQTLGL  
WYSKDSLIDLIGYSADAFAGYRLDKKSTSETYQFLGVNLISWFSKKNLIALSTVEAEYIAAESCCAQI

>Singleton10: Trat2\_EO\_MAN\_3F2/13R1\_CO\_18

DVWNEDEAEWEKLTLLRGGKQGSKIVTTTRSDVVARIMGAVTPHKLQGLTSDDCWTLFKQRAFGPGREEETSRSVEIGKEIVEK  
CRGLPLAAKVLGSLMRFKTEDAEWVHVRDSELWRLPRNENRILLELQLSYDHLPSDLKQCFAYCSIFPKDYQIERKQLIQLWIAEGF  
IQTSDGDMHEEEVGNQNFNSLLRRSFFQDARKDKYNNIWACKMHDLMHDLAX

>Singleton11: Trat2\_EO\_MAN\_3F2/13R1\_CO\_2

VGIPRPGTGGNQHKRKKVILTRSDVCGKMSADAEIKMNCLGQREALDLNFYVVGKQTLSSHPDIPELAGEVAKECRGLPLALKTI  
GRAMRNRREPKHWRHVIKLLKDSKHAEIQEMEILPEGDEERDMFRILKLSYDNLHDLMHX

>Singleton12: Trat2\_EO\_MAN\_3F2/13R1\_CO\_20

DGSCVIVTRSTTVVSSIIIGTLDAYELKPLSDDECWLSLERNFSEGGAEKSPNLVHIGKAIVDKCGGLPLAAKMLGRLLRKTSREER  
EWRFMSEIWKLNNDKILPSLKLSYDHLPCVHKQCFAYCATFPKDYIMDKEMLIQLWMANGFIPSDPQDMELKGHEIFNELLSRS  
FFQDFKNTQSYPMMITKQTCMHDLMHX

>Singleton13: Trat4\_EO\_TEF\_3F2/13R1\_CO\_21

RYFLVDDVWDETGSKWEKLSALTSAGAKGSMVLVTTQSPVAVVMGTMDPIELECLQEDDYWRLFHCAFQDQILEEEGREKL  
QAIGRKISEKLHGLPLAGRVLGSLKTKLNEDHWKMILESEWWEHEYVLDNLPLALS\*EHLNANLKQCFAYCSIFPKSYGFEKDR  
LVQMWMAQGFIQNKNGLRMRMEDX

>Singleton14: Trat5\_EG\_NIG\_P1B/RNBSD\_CO\_6

HVSDDFSVRRITREIMESASKGRQSELVNWVMVQDNLKGQLPGNRYLLVDDVWVWEDRNQWEIFLPLMYGERGSKILITTQKN  
VANIMDMASSYCLEGLSNQDCWLIFKRIAFANKNSSEHPKLEEIGKEIVSKIRGLPLAAKILGGLLYSKLDEDSWRIIESEIWELPQN  
AKNILSALRSYQQLPGHLKQCFLYCSLFP

>Singleton15: Trat8\_EG\_LM2T\_3F2/13R1\_R\_20

RIIVTTESERVSSMGTALATYQLPYLTEDDSWTLFRQRALXEGGEEPQNLVNIGKEIVKCCRGLPLAVKTMGGLLYSKSQEREWLF  
VRDSEIWDMMQVGEDAILPALRSYSHLPSHLKQCFAFCAIFPKDYEMEKDILLIQLWMANGFIPSGRKELEDKGHEIFNELASRSF  
QDIKEVEGHAGFDGFFVRSSKVVYITTCMHDLVTTWPXXXXXX

>Singleton16: Trat8\_EG\_LM2T\_3F2/13R1\_R\_34

EGSRHVTARSEKVVSSIMGTLPCTRLSCLTEDDSWTLFRKRAFEEKGADVPPNLEKIGKEIVKCCGGLPLAVKALGSLMHKSQEKW  
LSVRDSEIWDMMQVGEDGILPALRSYSHLPSHLKQCFAFCAIFPKDDEMEKDILLIQLWMANGFIPSDGRKELEEKHEIFNELASRSF  
FQDIKLVEEAETYRGQHFYLGDRTRHELYYTTTCMHDMMHX